

XIII-th International Conference
Knowledge-Dialogue-Solution

June 18-24, 2007, Varna (Bulgaria)



P R O C E E D I N G S

Volume 1

ITHEA

SOFIA, 2007

Gladun V.P., Kr.K. Markov, A.F. Voloshin, Kr.M. Ivanova (editors)

Proceedings of the XIII-th International Conference “Knowledge-Dialogue-Solution” –

Varna, 2007

Volume 1

Sofia, Institute of Information Theories and Applications FOI ITHEA, Bulgaria– 2007

First Edition

The XIII-th International Conference “Knowledge-Dialogue-Solution” (KDS 2007) continues the series of annual international KDS events organized by Association of Developers and Users of Intelligent Systems (ADUIS).

The conference is traditionally devoted to discussion of current research and applications regarding three basic directions of intelligent systems development: knowledge processing, natural language interface, and decision making.

Edited by:

Institute of Information Theories and Applications FOI ITHEA, Bulgaria

Association of Developers and Users of Intelligent Systems, Ukraine

Printed in Bulgaria by Institute of Information Theories and Applications FOI ITHEA

Sofia-1090, P.O.Box 775, Bulgaria

e-mail: info@foibg.com

www.foibg.com

All Rights Reserved

© 2007 Viktor P. Gladun, Krassimir K. Markov, Alexander F. Voloshin, Krassimira M. Ivanova - Editors

© 2007 Krassimira Ivanova - Technical editor

© 2007 Institute of Information Theories and Applications FOI ITHEA, Bulgaria - Publisher

© 2007 Association of Developers and Users of Intelligent Systems, Ukraine - Co-edition

© 2007 For all authors in the issue

ISSN 1313-0087 (paperback) ISSN 1313-1176 (CD) ISSN 1313-1206 (online)

FISBN Volume 1: 978-954-16-2007-6 (paperback) 978-954-16-2008-3 (CD) 978-954-16-2009-0 (Online)

FISBN Volume 2: 978-954-16-2010-6 (paperback) 978-954-16-2011-3 (CD) 978-954-16-2012-0 (Online)

PREFACE

The scientific XIIIth International Conference "Knowledge-Dialogue-Solution" took place in June, 18-24, 2007 in Varna, Bulgaria. This two volumes include the papers presented at the conference. Reports contained in the Proceedings correspond to the scientific trends, which are reflected in the Conference name.

The Conference continues the series of international scientific meetings, which were initiated more than fifteen years ago. It is organized owing to initiative of ADUIS - Association of Developers and Users of Intelligent Systems (Ukraine), Institute of Information Theories and Applications FOI ITHEA, (Bulgaria), and IJ ITA - International Journal on Information Theories and Applications, which have long-term experience of collaboration.

Now we can affirm that the international conferences "Knowledge-Dialogue-Solution" in a great degree contributed to preservation and development of the scientific potential in the East Europe.

KDS-2007 is dedicated to:

- 60th Anniversary of the Institute of Mathematics and Informatics of Bulgarian Academy of Sciences;
- 15th Anniversary of the Association of Developers and Users of Intelligent Systems (Ukraine);
- 10th Anniversary of the Association for Development of the Information Society (Bulgaria).

The conference is traditionally devoted to discussion of current research and applications regarding three basic directions of intelligent systems development: knowledge processing, natural language interface, and decision making.

The basic approach, which characterizes presented investigations, consists in the preferential use of logical and linguistic models. This is one of the main approaches uniting investigations in Artificial Intelligence.

The organization of the papers in KDS-2007 is based on specialized sessions. They are:

- Neural Nets
- Mathematics of Computing
- Decision Tools and Techniques
- Decision Support
- Formal Models
- Expert Systems
- Ontologies
- Knowledge Acquisition
- Natural Language Processing
- Pattern Recognition
- Distributed Information Processing
- Philosophy and Methodology of Informatics

The official languages of the Conference are Russian and English.

The Conference is sponsored by FOI Bulgaria (www.foibg.com).

We appreciate the contribution of the members of the KDS 2007 Program Committee.

On behalf of all the conference participants we would like to express our sincere thanks to everybody who helped to make conference success and especially to Kr.Ivanova, I.Mitov, V.Velichko and L. Svyatogor.

Chairmen of the Program Committee: Victor Gladun, Alexey Voloshin, Krassimir Markov

CONFERENCE ORGANIZERS

National Academy of Sciences of Ukraine
 Association of Developers and Users of Intelligent Systems (Ukraine)
 International Journal "Information Theories and Applications"
 V.M.Glushkov Institute of Cybernetics of National Academy of Sciences of Ukraine
 Institute of Information Theories and Applications FOI ITHEA (Bulgaria)
 Institute of Mathematics and Informatics, BAS (Bulgaria)
 Institute of Mathematics of SD RAN (Russia)

PROGRAM COMMITTEE

Victor Gladun (Ukraine)
 Alexey Voloshin (Ukraine)
 Krassimir Markov (Bulgaria)

Boris Fedunov (Russia)	Genady Osipov (Russia)
Frank Brown (USA)	Alexander Palagin (Ukraine)
Vladimir Donskoy (Ukraine)	Vladimir Pasechnik (Ukraine)
Alexander Eremeev (Russia)	Zinoviy Rabinovich (Ukraine)
Natalia Filatova (Russia)	Alexander Reznik (Ukraine)
Constantine Gaidric (Moldova)	Vladimir Ryazanov (Russia)
Tatyana Gavrilova (Russia)	Galina Rybina (Russia)
Krassimira Ivanova (Bulgaria)	Vasil Sgurev (Bulgaria)
Vladimir Jotsov (Bulgaria)	Vladislav Shelepov (Ukraine)
Julia Kapitonova (Ukraine)	Anatoly Shevchenko (Ukraine)
Volodimir Donchenko (Ukraine)	Ekaterina Solovyova (Ukraine)
Vladimir Khoroshevsky (Russia)	Vadim Stefanuk (Russia)
Rumyana Kirkova (Bulgaria)	Tatyana Taran (Ukraine)
Nadezhda Kiselyova (Russia)	Valery Tarasov (Russia)
Alexander Kleshchev (Russia)	Adil Timofeev (Russia)
Oleg Kuznetsov (Russia)	Vadim Vagin (Russia)
Vladimir Lovitskii (GB)	Jury Valkman (Ukraine)
Vitaliy Lozovskiy (Ukraine)	Neonila Vashchenko (Ukraine)
Natalia Kussul	Vitaliy Velichko (Ukraine)
Nadezhda Mishchenko (Ukraine)	Stanislav Wrycza (Poland)
Iliia Mitov (Bulgaria)	Nikolay Zagoruiko (Russia)
Xenia Naidenova (Russia)	Jury Zaichenko (Ukraine)
Olga Nevzorova (Russia)	Arkady Zakrevskij (Belarus)

Official languages of the conference are Russian and English.

General sponsor of the KDS 2007 is **FOI BULGARIA** (www.foibg.com).

TABLE OF CONTENTS – VOLUME 1

<i>Preface</i>	3
<i>Table of Contents</i>	5
<i>Index of Authors</i>	11
F.1.1. Neural Nets	
Selfstructured Systems	
<i>Victor Gladun, Vitaly Velichko, Yuriy Ivaskiv</i>	13
Иерархическое логическое описание и нейросетевое распознавание сложных образов	
<i>Татьяна Косовская, Адиль Тимофеев</i>	22
Data Mining with Fuzzy ARTMAP Neural Networks: Prediction of Profiles of Potential Customers	
<i>Anatoly Nachev</i>	27
Almost Separable Data Aggregation by Layers of Formal Neurons	
<i>Leon Bobrowski</i>	34
Complex Neural Network Model of User Behavior in Distributed Systems	
<i>Andrii Shelestov, Serhiy Skakun, Olga Kussul</i>	42
Оценка биоразнообразия с использованием нейронных сетей	
<i>Андрей Шелестов, Екатерина Насуро, Сергей Скакун</i>	49
Forming of Learning Set for Neural Networks in Problems of Lossless Data Compression	
<i>Yuriy Ivaskiv, Victor Levchenko</i>	55
Система оптимизации на основе имитационного моделирования, генетического алгоритма и нейросетевых метамоделей	
<i>Павел Афонин</i>	60
Диагностирование на нейронных сетях в системе ГОМЕОПАТ	
<i>Лариса Катеринич, Александр Провотор</i>	64
Neural Network Approach Prediction of the Type of a Course of Multiple Sclerosis by the Clinical Characteristics of its Debut	
<i>Inna Panchenko, Tetyana Shatovska</i>	68
G. Mathematics of Computing	
Finding an Appropriate Partition on the Set of Arguments of a Partial Boolean Function to be Decomposed	
<i>Arkadij Zakrevskij</i>	71
Верификация логических описаний с функциональной неопределенностью	
<i>Людмила Черемисинова, Дмитрий Новиков</i>	78
Обобщённые варианты преобразования Хока: статистический и алгебраический аспекты	
<i>Владимир Донченко</i>	84
Алгебраический Jack Knife: кластеризация по гиперплоскостям	
<i>Николай Кириченко, Владимир Донченко</i>	89
Метаэвристический метод деформаций для решения задач комбинаторной оптимизации	
<i>Леонид Гуляницкий</i>	95
Один класс алгоритмов стохастического локального поиска	
<i>Леонид Гуляницкий, Александр Турчин</i>	102
Об эффективности функционалов эмпирического риска и скользящего экзамена как оценок вероятности ошибочной классификации	
<i>Виктор Неделько</i>	111

Прогнозирование разнотипного временного ряда методом адаптивного формирования пространства состояний в классе логических решающих функций <i>Светлана Неделько</i>	118
Detection of Logical-and-Probabilistic Correlation in Time Series <i>Tatyana Stupina</i>	123
Один подход к равновесиям в играх в условиях неопределенности <i>Сергей Мащенко</i>	129
Гиперслучайные явления: определения и описание <i>Игорь Горбань</i>	137
Анализ модели Леонтьева при нечётко заданных параметрах методом базисных матриц <i>Алексей Волошин, Владимир Кудин, Григорий Кудин</i>	147
Векторные комбинаторные задачи в пространстве сочетаний с дробно-линейными функциями критериев <i>Наталья Семенова, Людмила Колечкина, Алла Нагорная</i>	152
Аппроксимация экспериментальных данных кривыми Безье <i>Виталий Вишневецкий, Владимир Калмыков, Татьяна Романенко</i>	157

D.2.2. Decision Tools and Techniques

Application of Case-based Reasoning for Intelligent Decision Support Systems <i>Alexandr Ereemeev, Pavel Varshavsky</i>	163
Limit Behaviour of Dynamic Rule-based Systems <i>Gennady Osipov</i>	169
Многокритериальная оценка и оптимизация иерархических систем <i>Альберт Воронин</i>	174
Статистический анализ распределений полезностей <i>Виктор Семенов</i>	183
Variants of Encoding for Selection of Optimal Subset of Diagnostic Tests <i>Anna Yankovskaya, Yury Tsoy</i>	186
Аспекты нестабильности термометрических характеристик преобразователей температуры <i>Святослав Яцишин, Богдан Стадник, Юрий Лега, Ярослав Луцик, Вениамин Мельник</i>	192
Комбинированный метод решения задачи о рюкзаке <i>Татьяна Пермякова, Владимир Морозенко</i>	195
Информация, экономика, экология – компоненты социально-экономического потенциала территории (Инновационные характеристики региона и трехмерное представление кривой С. Кузнеця) <i>Анатолий Крисиллов, Виктор Крисиллов</i>	202
Основные аспекты математического моделирования управляемого диагностического процесса в рамках единого информационного пространства системы здравоохранения <i>Игорь Долгополов, Наталья Семенова</i>	210
Методы модификации оптимизационных моделей с целью увеличения диапазона выбора вариантов принятия управленческих решений <i>Николай Кузубов</i>	213
Проблемы формирования внутриличностного конфликта как фактора, влияющего на принятие решений <i>Елена Шинкаренко</i>	217
Культурная идентификация как фактор стабильности в двуязычном обществе <i>Ирина Горицына, Александр Глущенко</i>	225

H.4.2. Decision Support

Системы диагностирования в медицине, как персональный интеллектуальный инструментарий врача <i>Алексей Волошин, Максим Запорожец, Павел Мулеса</i>	233
Decision Support System in Ultrasound Investigations <i>Svetlana Cojocar, Constantin Gaidric</i>	241
Эволюционное моделирование процесса распространения пожара <i>Виталий Снитюк, Артем Быченко</i>	247
MultiDecision-2: A Multicriteria Decision Support System <i>Vassil Vassilev, Mariana Vassileva, Boris Staykov, Krassimira Genova, Filip Andonov, Plamena Chongova</i>	255
Проблемы прогнозирования экономических макропараметров <i>Алексей Волошин, Виктория Сатыр</i>	264
Интеллектуальные технологии в маркетинговом анализе <i>Галина Сетлак</i>	270
Анализ и сравнение результатов оптимизации инвестиционного портфеля при применении модели Марковица и нечетко-множественного метода <i>Юрий Зайченко, Малихех Есфандиярфард</i>	278
Модель интеллектуальной системы управления ненадежными элементами (людьми) <i>Тимофей Рябцев, Елена Антонова</i>	287
Подход к разработке интеллектуальной системы управления ненадежными элементами. Процедуры управления. <i>Роман Бенгер, Елена Антонова</i>	294
Математическое обеспечение программного комплекса "Персонал" при определении трудового участия членов коллектива <i>Юрий Бондарчук, Григорий Гнатиенко</i>	299
Designing and Evaluating of Crew Activity Algorithms on Initial Stage of Designing of Anthropocentric Object <i>Boris Fedunov, Denis Vidruk</i>	303

E.4 Formal Models

UML: история, спецификация, библиография <i>Дмитрий Буй, Елена Шишацкая</i>	309
Модель "сущность-связь": роли, сильные и слабые типы сущностей и типы связей <i>Дмитрий Буй, Людмила Сильвейструк</i>	316
Multialgebraic Structures Existence for Granular Computing <i>Alexander Kagramanyan, Vladimir Mashtalir, Vladislav Shlyakhov</i>	322
Prediction of Properties and States of Dynamic Objects Through Analogical Inference <i>Yurij Kuk, Helen Lavrikova</i>	331
Автоматический синтез функциональных схем <i>Наталья Филатова, Олег Ахремчик, Олег Куприянов</i>	338

About:

<i>60th Anniversary of Institute of Mathematics and Informatics, Bulgarian Academy of Science</i>	346
<i>15th Anniversary of Association of Developers and Users of Intellectualized Systems</i>	347
<i>10th Anniversary of Association of Developing of the Information Society</i>	348
<i>15th Volume of International Journal "Information Theories and Applications"</i>	349
<i>Second Volume of International Journal "Information Technologies and Knowledge"</i>	350

TABLE OF CONTENTS – VOLUME 2

<i>Preface</i>	353
<i>Table of Contents</i>	355
<i>Index of Authors</i>	361
I.2.1. Expert Systems	
The Experience of Development and Application Perspectives of Learning Integrated Expert Systems in the Educational Process <i>Galina Rybina, Victor Rybin</i>	363
Концепция компьютерного банка знаний по медицинской диагностике <i>Филипп Москаленко, Александр Клещёв, Мери Черняховская</i>	365
Information-Analytical System for Design of New Inorganic Compounds <i>Nadezhda Kiselyova, Andrey Stolyarenko, Vladimir Ryazanov, Vadim Podbel'skii</i>	373
A Model of Rule-based Logical Inference <i>Xenia Naidenova</i>	379
An Intelligent System for Investigations and Provision of Safety for Complex Constructions <i>Alexander Berman, Olga Nikolaychuk, Alexander Yurin, Alexander Pavlov</i>	389
Экспертная система контроля органолептических показателей качества мясной продукции <i>Вадим Зайцев</i>	396
I.2.4. Ontologies	
Multilevel Ontologies for Domains with Complicated Structures <i>Irene Artemieva</i>	403
Management of Information on Program Flow Analysis <i>Margarita Knyazeva, Dmitry Volkov</i>	411
Automatic Generation of Context-sensitive Help Using a User Interface Project <i>Valeriya Gribova</i>	417
In Search of a Vision: Ontological View on User Modelling Conferences' Scope <i>Tatiana Gavrilova, Seppo Puuronen</i>	422
Онтологический взгляд на теорию автоматов <i>Сергей Кривой, Людмила Матвеева, Елена Лукьянова, Ольга Седлецкая</i>	427
Ontology-driven Intrusion Detection Systems <i>Vladimir Jotsov</i>	436
Rule-Mining: подход к автоматизированному извлечению онтологий <i>Роман Гадиатулин, Светлана Чуприна</i>	445
Онтологический анализ Web-сервисов в интеллектуальных сетях <i>Анатолий Гладун, Юлия Рогушина, Виктор Штонда</i>	451
Инструментальная среда XG#: онтологический подход к разработке интегрированных оболочек экспертных систем <i>Михаил Никулин</i>	459
Концептуальные принципы реализации и структура инструментария контроля знаний на базе онтологий <i>Елена Нетаевская</i>	464

I.2.6. Knowledge Acquisition

Information Search Based on Analysis of Experts Statements <i>Gennadiy Lbov, Nikolai Dolozov, Pavel Maslov</i>	471
Interval Prediction Based on Experts' Statements <i>Gennadiy Lbov, Maxim Gerasimov</i>	474
Enhancing Information Retrieval by Using Evolution Strategies <i>Abdelmgeid Amin Aly</i>	478
Knowledge-based Robot Control <i>Agris Nikitenko</i>	487
LOM Manager: управление обучающими объектами в системе прототипирования обучающих курсов VITA II <i>Ольга Малиновская</i>	500
The New Software Package for Dynamic Hierarchical Clustering for Circles Types of Shapes <i>Tetyana Shatovska, Tetyana Safonova, Yuriy Tarasov</i>	507
The Metrics and Measure of Refutability on Formulas in the Theory T <i>Alexandr Vikent'ev</i>	513
Измеримые модели теории первого порядка в расстояниях на высказываниях и вероятностях на знаниях <i>Александр Викентьев</i>	518

I.2.7. Natural Language Processing

Информационная модель обработки естественно-языковых текстов <i>Александр Палагин, Виктор Гладун, Николай Петренко, Виталий Величко, Алексей Севрук, Андрей Михайлюк</i>	525
Knowledge-Based Approach to Document Analysis <i>Elena Sidorova, Yury Zagorulko, Irina Kononenko</i>	527
Automated Response to Query System <i>Vladimir Lovitskii, Michael Thrasher, David Traynor</i>	534
Analysis of Text Documents in Automatic Abstracting System <i>Stanislav Lipnitsky, Denis Nasuro</i>	544
Lexicon of Common Scientific Words and Expressions for Automatic Discourse Analysis of Scientific and Technical Texts <i>Elena Bolshakova</i>	551
Technology of Storage and Processing of Electronic Documents with Intellectual Search Properties <i>Yuri Kalafati, Konstantin Moiseyev, Sergey Starkov, Svetlana Shushkova</i>	558
Criteria of Loan Words Identification <i>Alla Zaboleeva-Zotova, Ilya Prokhorov</i>	564

I.5. Pattern Recognition

Критерии информативности и пригодности подмножества признаков <i>Ирина Борисова, Николай Загоруйко, Ольга Кутненко</i>	567
Conditions of Effectiveness of Pattern Recognition Problem Solution Using Logical Level Descriptions of Classes <i>Adil Timofeev, Tatiana Kosovskaya</i>	572
Bayesian Model of Recognition on a System of Events <i>Vladimir Berikov</i>	576
Процедуры локализации вектора весовых коэффициентов для нечетких моделей выбора <i>Елена Присяжнюк</i>	579

C.2.4. Distributed Information Processing

Разработка каталога метаданных системы GEO-Ukraine <i>Наталья Куссуль, Алина Рудакова, Алексей Кравченко</i>	585
Agent-based Anomalies Monitoring in Distributed Systems <i>Andrii Shelestov</i>	594
Автоматическое выявление ударных волн по измерениям спутника ACE <i>Андрей Шелестов, Ксения Житомирская, Николай Ильин, Игорь Кременецкий</i>	599
Safety Policy Problems of Cluster Supercomputers <i>Andrey Golovinskiy, Sergey Ryabchun, Anatoliy Yakuba</i>	606
Parametric Identification and Diagnosis of Integrated Navigation Systems in Bench Test Process <i>Ilya Prokoshev, Alexander Chernodarov</i>	611
Динамическое распределение объектов имитационной модели, основанное на знаниях <i>Александр Миков, Елена Замятина, Константин Осмехин</i>	618

I.2.0. Philosophy and Methodology of Informatics

Culture Aspects of Inforaction <i>Krassimir Markov, Stoyan Poryazov, Krassimira Ivanova, Iliia Mitov, Vera Markova</i>	625
Magic of Egregors <i>Vitaliy Lozovski</i>	634
Развивающиеся системы <i>Александр Резник</i>	650
О природе интеллекта <i>Александр Резник</i>	658
Two Fundamental Problems Connected with AI <i>Dimiter Dobrev</i>	667
Размышляющие компьютеры <i>Виталий Яценко</i>	673
Моделирование субъективного представления человека о действительности <i>Алексей Бычков, Михайл Меркурьев</i>	679
Предпосылки возникновения общей теории информации <i>Василий Луц</i>	687

About:

<i>In memoriam: Gennady Mikhailovich Bakan</i>	697
<i>10th Anniversary of Association of Developing of the Information Society</i>	698
<i>15th Anniversary of Association of Developers and Users of Intellectualized Systems</i>	699
<i>60th Anniversary of Institute of Mathematics and Informatics, Bulgarian Academy of Science</i>	700

INDEX OF AUTHORS

Pavel Afonin	60	Alexander Kagramanyan	322
Oleg Akhremchik	338	Yuriy Kalafati	558
Abdelmgeid Amin Aly	478	Vladimir Kalmykov	157
Filip Andonov	255	Larisa Katerynych	64
Elena Antonova	287, 294	Nickolay Kirichenko	89
Irene Artemieva	403	Nadezhda Kiselyova	373
Roman Bengier	294	Alexander Kleshchev	365
Vladimir Berikov	576	Margarita Knyazeva	411
Alexander Berman	389	Liudmila Kolechkina	152
Leon Bobrowski	34	Irina Kononenko	527
Elena Bolshakova	551	Tatyana Kosovskaya	22, 572
Yuriy Bondarchuck	299	Alexey Kravchenko	585
Irina Borisova	567	Igor Kremenetskiy	599
Dmitrii Buy	309, 316	Anatoliy Krissilov	202
Artyom Bychenko	247	Victor Krissilov	202
Alexey Bychkov	679	Sergey Krivoi	427
Liudmila Cheremisinova	78	Grigorii Kudin	147
Alexander Chernodarov	611	Vladimir Kudin	147
Mery Chernyahovskaya	365	Yuriy Kuk	331
Plamena Chongova	255	Oleg Kuprianov	338
Svetlana Chuprina	445	Natalia Kussul	585
Svetlana Cojocar	241	Olga Kussul	42
Dimitier Dobrev	667	Olga Kutnenko	567
Igor Dolgopolov	210	Nickolay Kuzubov	213
Nickolay Dolozov	471	Helen Lavrikova	331
Vladimir Donchenko	84, 89	Gennady Lbov	471, 474
Alexander Ereemeev	163	Yuriy Lega	192
Maliheh Esfandiartard	278	Victor Levchenko	55
Boris Fedunov	303	Stanislav Lipnitsky	544
Natalia Filatova	338	Vladimir Lovitskii	534
Roman Gadiatulin	445	Vitaliy Lozovski	634
Constantin Gaindric	241	Elena Lukyanova	427
Tatyana Gavrilova	422	Vasiliy Luts	687
Krassimira Genova	255	Yaroslav Lutsik	192
Maxim Gerasimov	474	Olga Malinovskaya	500
Anatoliy Gladun	451	Krassimir Markov	625
Victor Gladun	13, 525	Vera Markova	625
Alexander Glushchenko	225	Sergey Mashchenko	129
Grigorii Gnatienco	299	Vladimir Mashtalir	322
Andrey Golovinskiy	606	Pavel Maslov	471
Irina Gorytsina	225	Liudmila Matveyeva	427
Valeriya Gribova	417	Veniamin Melnik	192
Igor Horban	137	Mikhail Merkuriev	679
Leonid Hulyanitskiy	95, 102	Andrii Mikhailyuk	525
Nickolay Ilin	599	Alexander Mikov	618
Krassimira Ivanova	625	Iliya Mitov	625
Yuriy Ivaskiv	13, 55	Konstantin Moiseyev	558
Vladimir Jotsov	436	Vladimir Morozenko	195

Phillip	Moskalenko	365	Andrii	Shelestov	42, 49, 594, 599
Pavel	Mulesa	233	Elena	Shinkarenko	217
Anatoly	Nachev	27	Elena	Shishatskaya	309
Alla	Nagornaya	152	Vladislav	Shlyakhov	322
Xenia	Naidenova	379	Victor	Shtonda	451
Denis	Nasuro	544	Svetlana	Shushkova	558
Ekaterina	Nasuro	49	Elena	Sidorova	527
Svetlana	Nedelko	118	Liudmila	Silvestruk	316
Victor	Nedelko	111	Sergey	Skakun	42, 49
Elena	Netavskaya	464	Vitaliy	Snytyuk	247
Mikhail	Nickulin	459	Bogdan	Stadnik	192
Agris	Nikitenko	487	Sergey	Starkov	558
Olga	Nikolaychuk	389	Boris	Staykov	255
Dmitrii	Novikov	78	Andrey	Stolyarenko	373
Gennady	Osipov	169	Tatyana	Stupina	123
Konstantin	Osmehin	618	Yuriy	Tarasov	507
Alexander	Palagin	525	Michael	Thrasher	534
Inna	Panchenko	68	Adil	Timofeev	22, 572
Alexander	Pavlov	389	David	Traynor	534
Tatyana	Permyakova	195	Yuriy	Tsoy	186
Nickolay	Peternko	525	Alexander	Turchin	102
Vadim	Podbelskii	373	Pavel	Varshavskiy	163
Stoyan	Poryazov	625	Vassil	Vassilev	255
Ilya	Prokhorov	564	Mariana	Vassileva	255
Ilya	Prokoshev	611	Vitaliy	Velichko	13, 525
Alexander	Provotar	64	Denis	Vidruk	303
Elena	Prysiazhniuk	579	Alexander	Vikent'ev	513, 518
Seppo	Puuronen	422	Vitaliy	Vishnevskij	157
Alexander	Reznik	650, 658	Dmitry	Volkov	411
Yulia	Rogushina	451	Alexey	Voloshin	147, 233, 264
Tatyana	Romanenko	157	Albert	Voronin	174
Alina	Rudakova	585	Svyatoslav	Yacishin	192
Sergey	Ryabchun	606	Anatolij	Yakuba	606
Timofey	Ryabtsev	287	Anna	Yankovskaya	186
Vladimir	Ryazanov	373	Vitaliy	Yashchenko	673
Victor	Rybin	363	Alexander	Yurin	389
Galina	Rybina	363	Alla	Zaboleeva-Zotova	564
Tetyana	Safonova	507	Nickolay	Zagoruiko	567
Victoria	Satyr	264	Yuriy	Zagorulko	527
Olga	Sedletsкая	427	Yuriy	Zaichenko	278
Victor	Semenov	183	Vadim	Zaitsev	396
Natalia	Semenova	152, 210	Arkadij	Zakrevskij	71
Halina	Setlak	270	Elena	Zamyatina	618
Alexey	Sevruk	525	Maxim	Zaporozhets	233
Tetyana	Shatovska	68, 507	Xenia	Zhitomirskaya	599

F.1.1. Neural Nets

SELFSTRUCTURIZED SYSTEMS

Victor Gladun, Vitaly Velichko, Yurii Ivaskiv

Abstract: *The problems of constructing the selfstructured systems of memory of intelligence information processing tools, allowing formation of associative links in the memory, hierarchical organization and classification, generating concepts in the process of the information input, are discussed. The principles and methods for realization of selfstructured systems on basis of hierarchic network structures of some special class – growing pyramidal network are studied. The algorithms for building, learning and recognition on basis of such type network structures are proposed. The examples of practical application are demonstrated.*

Keywords: *knowledge discovery, classification, prediction, growing pyramidal networks, concept formation.*

ACM Classification Keywords: *I.2.4 Knowledge Representation Formalisms and Methods - Semantic networks, F.1.1 Models of Computation - Self-modifying machines (e.g., neural networks)*

Introduction

The task of constructing the selfstructured systems is considered in context with intellectualization of the information processing tools. Selfstructurization provides a possibility of changing the structure of data, stored in memory, in the process of the tools functioning as a result of interaction between the received and already stored information. Systems in which the perception of new information is accompanied by simultaneous structurization of the information stored in memory, we call hereinafter selfstructured.

Development of principles and methods of constructing the selfstructured systems in many respects defines a possibility of intellectualization of the information processing tools. Adaptability to the task, being solved, has to do with changing the structure of data. As a result, the possibility of searching in the memory focused on storage of complex data of the large volume occurs, which allows increasing productivity of the used tools, raise accuracy and reliability of received results.

Main processes of structurization of the perceived information consist in formation of semantic and syntactic links among objects by separation of crossings of their attributive representations, as well as the generalized logic attributive models of classes of objects - concepts. As a result of these processes realization, the semantic and syntactic similarity of the perceived information with the stored information is established. Detected associations are fixed as structural changes in memory.

Following basic requirements to data structures in the intellectual systems are set for the decision of such tasks as regularity discovery, classification, forecasting, diagnostics [1]:

The structure of the data should be the multipleparameter model, reflecting significant properties of researched object. It should provide a possibility to account for the simultaneous influence on researched factor of various combinations of known properties of the researched object.

The model of the researched object should minimize scanning of large-scale data: along with growth of data size, the time of performing of the choice operations grows. It interferes with application of some analysis methods. The model also should be applicable for verification and interpretation.

It should be noted, that in solving tasks of diagnostics and forecasting the models characterized by higher level of generalization of models of classes of objects have advantage. The logic expressions describing such models turn out easier if the complexity is evaluated by number of variables. Simplification of logic expressions results in simple structure of memory and, therefore, simplifies the process of structurization.

In knowledge representation in intelligent systems, those network structures have advantages, which have some information units in vertices, and arches describing links among them. In similar systems, the elements of knowledge representation are combined in the hierarchical structure, realizing such functions, as formation of links among attributive presentations of researched object by allocation of their crossings, hierarchical ordering, classification, concepts formation. In selfstructurized systems, such functions should be performed in the process of the information perceiving.

Condition of an element formation of network structure, for example, unit or link between units, is some relation between determined structural elements of a network. The relations determining formation of structure elements of selfstructurized systems we call structurized.

There are two basic ways of objects representation in the information processing systems: by name (condensed) or by sets of attribute values (expanded). The memory structures in selfstructurized systems and the appropriate network structures should provide bidirectional conversion between such representations.

Building of selfstructurized systems is proposed to be realized on basis of network with hierarchical structures, named as growing pyramidal networks (GPN) [5].

The theory as well as practical application of GPN is expounded in a number of publications [3-6]. GPN realization has following stages:

- to construct the structure of a network for some initial set of objects, assigned by attributive descriptions,
- to train the structure, with a purpose to allocate its elements, allowing to classify all objects of the initial set,
- to recognize belonging to some class of objects of certain object, which is not belonging to initial set of objects.

The mechanisms, providing conversion between converged representation of objects and representation as a set of attributes values in human neurosystem, are discussed in the article [2]. The present work illustrates recent versions of algorithms for building and training GPN, as well as examples of their application.

Building of GPN

A *growing pyramidal network* is an acyclic oriented graph having no vertexes with a single incoming arc. Examples of the pyramidal networks are shown in Figs.1,2,3. Vertexes having no incoming arcs are referred to as *receptors*. Other vertexes are called *conceptors*. The subgraph of the pyramidal network that contains vertex *a* and all the vertexes from which there are paths to vertex *a* is called the *pyramid* of vertex *a*. The set of vertexes contained in the pyramid of vertex *a* is referred to as the *subset* of vertex *a*. The set of vertexes reachable by paths from vertex *a* is called the *superset* of vertex *a*. The set of vertex, having paths from vertex *a*, is referred to its *superset*.

In *subset* and *superset* of the vertex, *0-subset* and *0-superset* are allocated, consisting of those vertexes, which are connected to it directly. When the network is building, the input information is represented by sets of attributes values describing some objects (materials, states of the equipment, a situation, illness etc.). Receptors correspond to values of attributes. In various tasks, they can be represented by names of properties, relations, states, actions, objects or classes of objects. Conceptors correspond to descriptions of objects in general and to crossings of descriptions and represent GPN vertexes.

Initially the network consists only of receptors. Conceptors are formed as a result of algorithm of construction of a network. After input of object attribute description, corresponding receptors switch to a *state of excitation*. The process of excitation propagates through the network. A conceptor switches into the state of excitation if all

vertices of its 0-subset are excited. Receptors and conceptors retain their state of excitation during all operations of network building.

Let F_a be the subset of excited vertices of the 0-subset of vertex a ; G be the set of excited vertices in the network that do not have other excited vertices in their supersets. New vertices are added to the network by the following two rules:

Rule A1. If vertex a , that is a conceptor, is not excited and the power of set F_a exceeds 1, then the arcs joining vertices of set F_a with the vertex a are liquidated and a new conceptor is added to the network which is joined with vertices of set F_a by incoming arcs and with the vertex a by an outgoing arc.

The new vertex is in the state of excitation. Rule A1 is illustrated in Fig.1 (a,b). According to the Rule A1, the condition for adding a new vertex to the network is a situation, when certain network vertex is not completely excited (at least two vertices of 0-subset are excited). Fig 1.a shows a fragment of network in some initial state. Receptors 4,5 switch to a state of excitation, the network switches to state 1.b, a new vertex appears – a new conceptor. Receptors 2,3 switch to a state of excitation additionally. The network switches to state 1.c.

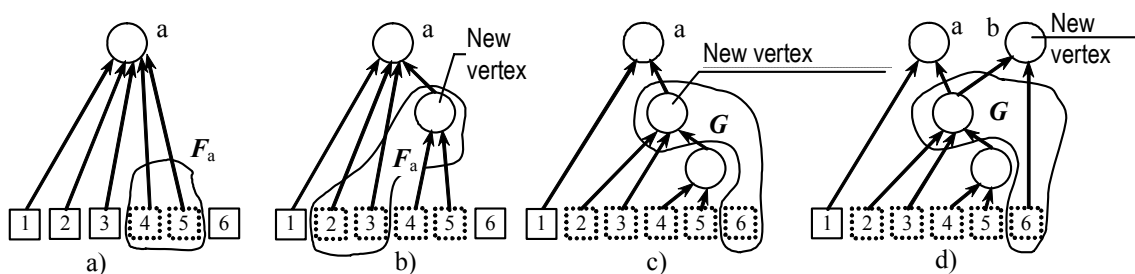


Fig. 1.

New vertices are inserted in 0-subset of vertices, which are not completely excited. New vertices correspond to intersection of object descriptions, represented by incoming arches. Once new vertices have been introduced into all network sections where the condition of rule A1 is satisfied, rule A2 is applied to the obtained network fragment, concluding the object pyramid building.

Rule A2. If the power of set G exceeds 1 element, a new conceptor is added to the network, which is joined with all vertices of set G by incoming arcs.

The new vertex is in the state of excitation. Rule A2 is illustrated in Fig.1c,d. Network Fig1.d was obtained after the excitation of receptors 2-6.

In applying the Rule A1 the main cross-linking relation is a relation of intersection of receptor set, excited by input of the object description and other sets of receptors included into pyramid, recently formed by conceptors. Rule A2 concludes the building of pyramid, which represents complete description of the introduced object.

Pyramidal networks are convenient for execution of various operations of associative search.

For example, it is possible to select all the objects that contain a given combination of attribute values by tracing the paths that outgo from the network vertex corresponding to this combination. To select all the objects whose descriptions intersect with the description of a given object it is necessary to trace the paths that outgo from vertices of its pyramid. Rules A1, A2 establishes associative proximity between objects having common combinations of attribute values.

Hierarchical organization is an important property of pyramidal networks. This provides a natural way for reflecting the structure of complex objects and generic-species interconnections.

Conceptors of the network correspond to combinations of attribute values that define separate objects and conjunctive classes of objects. By introducing the excited vertices into the object pyramid, the object is referred to classes, which descriptions are represented by these vertices. Thus, during the network building the conjunctive classes of objects are formed, the classification of objects is performed without a teacher. Classifying properties of pyramid network are vital for modeling environments and situations.

Conversion from converged representation of objects (conceptors) to expanded (sets of receptors) is performed by scanning pyramids in top-down and down-top directions.

Training GPN

Training GPN consists in formation of the structures representing concepts, on a basis of attributive descriptions of the objects incorporated into classes with known properties.

Concept is an element of knowledge system, representing generalized logic attributive model of a class of objects, by which processes of recognition of objects are realized. The set of objects generalized in concept is its *volume*.

Consider a task of inductive formation of concepts for not intersected sets of objects V_1, V_2, \dots, V_n , each set represents some class of objects with known properties. Let L - be a set of objects used as training set. All the objects of set L are represented by sets of attribute values. Relations $L \cap V_i \neq \emptyset$ and $V_i \not\subset L (i=1,2,\dots,n)$ are set. Each object from set L corresponds to one set V_i . It is necessary to generate n concepts by analysis L . The amount of these concepts must be sufficient for correct recognition of belongings of anyone $l \in L$ to one of sets V_i .

Each concept, generated on the basis of training set, is approximation to real concept, the proximity of concepts depends on representativeness of training set, i.e. on the detalization of peculiarities of the concept volume.

In forming the concept corresponding to set V_i , the objects of training set included in V_i , are considered as examples of set V_i , and the objects, not included in V_i , - as counterexamples of set V_i .

The combinations of attributes allocated in building of a pyramidal network, representing descriptions of objects of training set, are used as "a building material", a basis of further logic structure of concept.

Let L be the pyramidal network representing all of training set objects. For formation of concepts A_1, A_2, \dots, A_n , corresponding to sets V_1, V_2, \dots, V_n , pyramids of all objects of training set are scanned in order. The vertices of scanned pyramid during its scanning are considered excited. Special vertices in network are identified in order to recognize objects from the concept volume. They are referred to as *check vertices* of a certain concept. In selecting the check vertexes, two characteristics of network vertexes are used: $\{m_1, m_2, \dots, m_n\}$, where $m_i (i=1,2,\dots,n)$ is a number of objects of volume of concept A_i , which pyramids include the given vertex; and k , which is the number of receptors in the pyramid of this vertex. For receptors $k=1$. While scanning, the pyramid is transformed by the following rules

Rule B1. If in the pyramid of an object from concept volume A_i , the vertex, having the largest k among all the vertices with the largest m_i , is not a check vertex of concept A_i , then it is marked as a check vertex of the concept A_i .

The rule allows existence several vertexes among the excited vertexes with identical m_i , exceeding m_i of other excited vertexes. If in group of the vertexes having largest m_i , values k of all vertexes are equal, any of vertexes can be marked as check vertex of concept A_i .

The rule B1 is illustrated in Fig. 2. In a situation demonstrated by Fig. 2, in excitation in pyramid of vertex 2 vertex 6 is selected as check vertex as having the largest k among vertices with the largest m_i (6, 13, 14). Values m_i are shown inside symbols of vertexes.

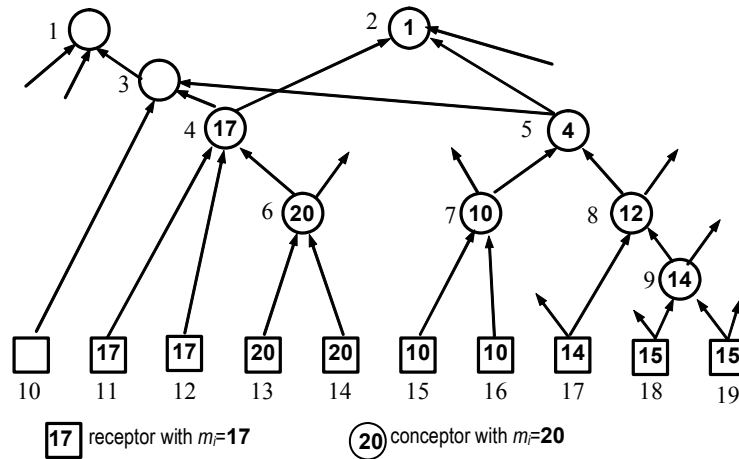


Fig. 2.

Rule B2. If the pyramid of an object from concept volume A_i contains check vertices of other concepts whose supersets do not contain excited check vertices of concept A_i , then in each of these supersets the vertex, having the largest k among all excited vertices with the largest m_i , is marked as a check vertex of concept A_i .

According to this rule the excitation of the pyramid of vertex 2 (Fig.3.a) on the condition, that it represents an objects from concept volume A_i , results in choosing vertex 5 as the check vertex of concept A_i (Fig. 3.b).

By check vertexes we select the most typical (having the largest m_i) combinations of attribute values, belonging to objects from concept volume. For example, selecting the vertex 8 (Fig 3a.) as a check vertex means selection of combination of value attributes, corresponding to receptors 17,18,19.

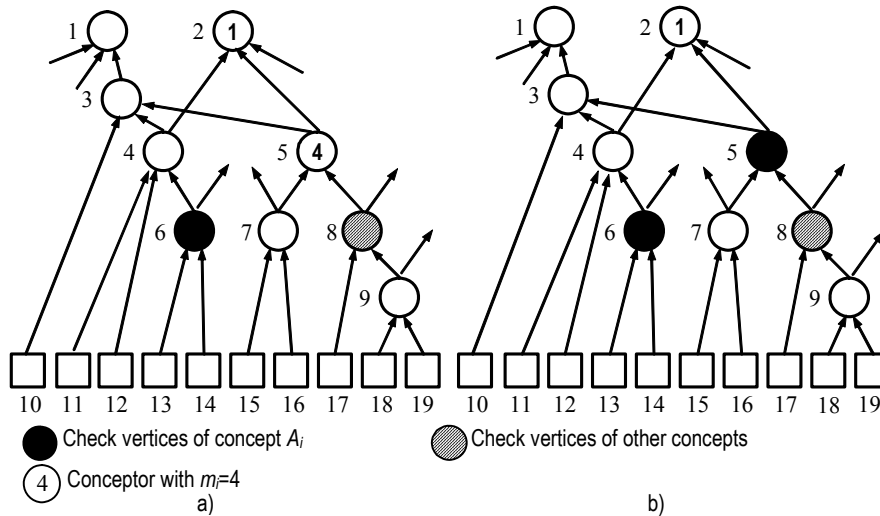


Fig. 3.

If at least one new check vertex appears while scanning objects of the training set, i.e. conditions of Rules B1 or B2 have been performed once at least, the training set is rescanned. The algorithm stops if during the scanning of the training set no new check vertex appears.

Recognition on basis of GPN

The task of recognition is based on the following rule.

Certain object belongs to the concept volume A_i if its pyramid has check vertexes A_i and does not contain check vertices of any other concept not having excited check vertices of concept A_i concept in their supersets. If this condition does not hold for any of the concepts, the object is referred to as unrecognized.

The execution time of the above algorithm is always finite. If the volumes of the formed concepts $V_1, V_2, \dots, V_i, \dots, V_n$ do not intersect, then after execution the algorithm the recognition rule completely divides the training set into subsets $L_i = V_i \cap L (i = 1 \dots n)$

The formed concepts are represented in the network as ensembles of check vertexes.

There is an algorithm[] of composing the logic descriptions of concepts, formed in the network as a result of the training process, described above. The formed logical expression contains logical relations, represented by allocation of check vertexes, describing the concepts in the network, defining different classes of objects.

The analytical tasks, such as diagnostics or prognosis, can be reduced to the task of classification, i.e. to belonging the research object to a class of objects, with a property characteristic or a set of properties significant for diagnostics of prognosis.

GPN Application

The following example illustrates the result of concepts formation on the basis of the analysis of a fragment of training set shown in the table. The table has descriptions of ceramic materials of two classes with the following attributes: M - material, T - fineness of powder, C - mix proportion, PP – powder manufacturing method, GP - conditions of obtaining the sample at hot pressing, NoGP - conditions of obtaining the sample without hot pressing, DU - special conditions of manufacturing of a sample, Por - porosity, Z - granularity.

Letters and figures in sections specify values of the appropriate attributes.

Table 1. Training set.

Object	Class	M	T	C	PP	GP	NoGP	DU	Por	Z
97	1	Al	2		SYN	2		2GP		
96	1	Al	2		SYN	2		1GP		
92	1	Al	2		SYN	2		2GP	1	
227	1	TiB	11	TiO-C	SYN		9		3	2
228	1	TiB	11	TiO-C	SYN		9		3	2
229	1	TiB	11	TiO-C	SYN		9		3	2
233	1	SiC	11	TiO-C	SYN		9		3	2
234	1	SiC	11	SiO-C	SYN		9		3	2
235	1	SiC	11	SiO-C	SYN		9		3	2
237	1	SiC	11	SiO-C	SYN		9		3	2
239	1	ZrB	11	ZrO-C	SYN		9		3	2
240	1	ZrB	11	ZrO-C	SYN		9		3	2
241	1	ZrB	11	ZrO-C	SYN		9		3	2
242	1	ZrB	11	ZrO-C	SYN		9		3	2
154	1	TiB	7	TiO-C	KRB	3			3	4
156	1	TiB	7	TiO-C	KRB	3			3	4
163	1	1AlO	1	AlO	SYN	1			4	
158	2	TiB	8	TiO-C	KRB	3			3	6
160	2	1AlO	1	AlO	SYN	1			1	
159	2	BC	1		SYN	1			1	
308	2	ZrB	11	ZrO-C	SYN		9			2

Fig. 4 demonstrates the appropriate pyramidal network with the formed concepts. Check vertexes PP_SYN, Por_3, 239, 163 characterize class 1, check vertexes 158, 308 and \$7 characterize class 2.

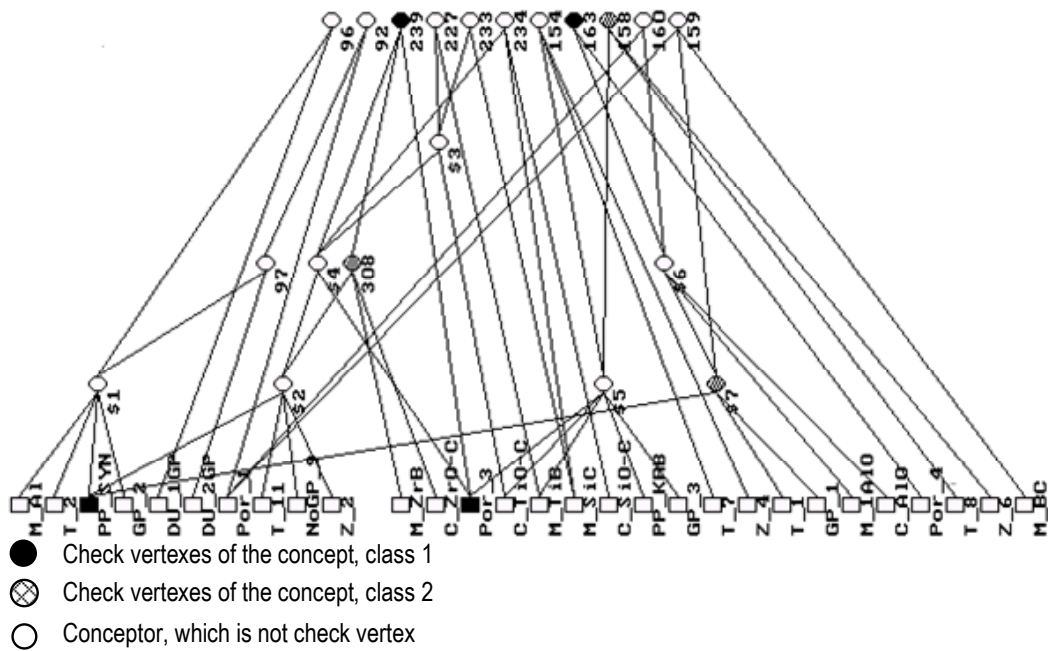


Fig 4.

The class 1 is described by the following logical expression, where \vee , \wedge , \neg - logical operations of a disjunction, conjunction and negation:

$$\begin{aligned}
 &PP_SYN \wedge \neg\{T_1 \wedge GP_1\} \wedge \neg\{M_ZrB \wedge C_ZrO-C \wedge T_11 \wedge NoGP_9 \wedge Z_2\} \vee \\
 &Por_3 \wedge \neg\{T_8 \wedge Z_6 \wedge M_TiB \wedge C_TiO-C \wedge PP_KRB \wedge GP_3\} \quad \vee \\
 &M_ZrB \wedge C_ZrO-C \wedge T_11 \wedge PP_SYN \wedge NoGP_9 \wedge Por_3 \wedge Z_2 \quad \vee \\
 &M_1AIO \wedge T_1 \wedge C_AIO \wedge PP_SYN \wedge GP_1 \wedge Por_4.
 \end{aligned}$$

The logic expressions, defining various classes of objects, are united in *cluster databases* (CDB). CDB contain the information on groups of objects (clusters), specific to the area of study. On basis CDB problems of classification, diagnostics and forecasting are solved. After the concept for some class of objects has been formed, problems of forecasting and diagnostics are reduced to a problem of classification. Classification of new objects is performed by comparing the attribute descriptions with the concept, defining a class of predictable or diagnosing objects. Objects can be classified by evaluating the value of the logical expressions that represent corresponding concepts. The variables, corresponding to the attribute values of the recognized object, set 1, other variable set 0. If the entire expression takes the value 1, that means that the object is included into volume of concept.

The next geometric interpretation of the concept formation algorithm can be proposed.

Every network vertex, having k receptors in its subset, corresponds to $(s-k)$ -dimensional plane in s -dimensional attribute space. The plane contains all the points corresponding to objects whose perceiving results in exiting of this vertex. $(s-k)$ -dimensional planes corresponding to check vertices of concept A_i are referred to as *zones* of concept A_i .

The following statements are true for growing pyramidal networks.

Statement 1. The zone of any network vertex is totally included in zones of its subset vertices and totally includes all zones of its superset vertices.

Statement 2. The point corresponding to an object in the attribute space is located inside an intersection of zones of those check vertices, which are exited when the object is perceived.

Point a corresponding to the object in the attribute space is directly included in the zone Z of concept A_i if there is no other zones of this concept which include point a and totally are included in zone Z .

The geometric interpretation of the above-described rules for concept formation algorithm is as follows.

Rule B1. For every object of concept volume A_i ($s-k$)-dimensional plane of the exited vertex having the highest k among all the vertices with the highest m_i becomes the zone of concept A_i .

Rule B2. If the point, corresponding to an object of concept volume A_i in the attribute space, is directly included in zones of the other concepts, then a zone of concept A_i is created inside each of those zones.

The algorithm of concept formation stops, when during regular examination of the training set, points corresponding to objects from any class are not directly included in zones of the other concepts. When learning is finished, an object corresponds to concept volume A_i if the appropriate point in the attribute space is directly included in at least one zone of concept A_i and is not included in any zone of the other concepts.

Zones of concept A_i , directly inclusive points of objects, corresponding to objects from its volume, as well as points, corresponding to objects from different concepts, are referred to as *boundary zones* of concept A_i .

Statement 3. According to Rule B2 new zones can be created only directly inside boundary zones.

Formation of new zones inside boundary zones results in division of boundary zones.

Construction of approximating region for concept A_i consists of two processes: rough covering with concept A_i zones the distribution domain of training set objects corresponding to concept A_i (Rule B1); and division of arising boundary zones (Rule B2).

On the basis of geometrical interpretation, algorithm convergence can have the following explanation.

For each concept the total covering by zones of allocation area of the training set objects, which are included in its volume, results in scanning of all objects, i.e. during single scanning of training set. The boundary zones include points of objects of training set, for which conditions of Rule B2 work. Therefore in every scanning of training set the division of all boundary zones, formed by previous scanning, occurs.

Process of division of boundary zones proceeds, as long as boundary zones exist, and can result in allocation of separate points of attribute space as zones. As number of the points corresponding to objects of training set is finite in each boundary zone, the process of division of boundary zones is finite too.

Absence of boundary zones after the termination of process of division means, that each of concepts in attribute space has area containing all points, corresponding objects of training set which are included in concept volume, and not including any point corresponding to other objects of training set. Thus, after the termination of division of boundary zones total division of training set into subsets $H_i = V_i \cap H (i = 1, 2, \dots, n)$ occurs. As a result algorithm operation for each of the formed concepts, the area is composed of zones of attribute space. This area contains all points of objects of the appropriate class and does not contain any point corresponding to objects of other classes. This area approximates allocation area of objects of the corresponding class. As the approximating area consists of linear elementary areas (hyperplanes), its limiting surface is piecewise-linear. Therefore, the algorithm performs the piecewise-linear division of objects, which correspond to different concepts.

The described method provides decisions of analytical problems of classification, diagnostics and forecasting on the basis of logic models of objects classes. The model displays dependences of an investigated class on combinations of values of attributes, i.e. allows taking account for combined influence of several attributes.

An important distinction of a method of concepts formation in growing pyramidal networks is the possibility to introduce in concepts the so-called excluding attributes which do not correspond to objects of a researched class. As a result, the formed concepts have more compact logic structure, which allows increasing the accuracy of diagnosis or forecasting. In logic expression the excluding attributes are presented by variables with negation.

All search operations in growing pyramidal network are limited to rather small fragment of a network, which includes an object pyramid and vertices directly linked to it. As a result, we have a possibility solve practical analytical problems based on large-scale data.

In a pyramidal network the information is stored by its representation in structure of network. Rules A1-A2, B1-B2 define the rules of memory organization while new information perception. The information of objects and classes of objects is presented by ensembles of vertices (pyramids), allocated in all network. Incoming of the new information causes redistribution of links among vertices of network, i.e. modifying of its structure.

The advantages of growing pyramidal networks become obvious in implementation, which allows parallel distribution of signals in network. The important property of a network as means of information storage is that the possibility of parallel distribution of signals is combined with parallel reception of signals to receptors.

Despite of the certain similarity of the processes proceeding in GPN and neural networks there some distinctions in operating. Main distinction of GPN is that its structure is formed depending on the input data automatically. The adaptation of network structure to the structure of data results in optimization of the information representation. In addition, in contrast to neural networks, the adaptation does not require the introduction of aprioristic redundancy of a network, and training process does not depend on the predetermined configuration of a network. The weakness of neural networks comparing with GPN is that the allocated generalized knowledge cannot be explicitly represented as rules or logic expression. It complicates their understanding by person.

Various set-theoretic descriptions of GPN are given in [4,5]. The [5] considers the so-called β -pyramidal networks (β -PN) modification of GPN for the ranked data. β -PN are useful for data presentation in problems of management, taking and planning decisions (for example, in planning the actions of robots), and also in semantic analysis and synthesis of natural-language texts. In [3-5] the algorithm of formation of concepts in GPN for nondetermined learning process, i.e. for a case when crossing volumes of different concepts occurs, is considered.

The program complex used for experimentation and solving the applied tasks using GPN[8], includes systems CONFOR, realizing processes of building and training GPN, and DISCRET by which the attributes given in numerical scales, are transformed in nominal scales. Discretization of attributes is performed on numerical scales by analysis of distributions of training set objects belonging to different classes.

Typical application fields for GPN are as follows: forecasting of new chemical compounds and materials with the predefined properties[7-9], forecasting in genetics, geology, the solar activity forecasting, medical and technical diagnostics, the robot planning, forecasting of failures of complex units etc. As an example we offer tasks of inorganic compounds forecasting with predefined properties. The tables containing attribute descriptions of binary, ternary and quaternary systems of chemical elements, forming or not forming the chemical compounds were used as training set. Training sets for binary, ternary and quaternary systems included 1333, 4278 and 4963 descriptions, and test set - 692, 2156 and 2536 descriptions. Each chemical element was described by the set of 87 attribute values. Descriptions of binary, ternary and quaternary systems had 174, 261 and 348 attributes. The recognition furnished the 99% accuracy result.

Conclusion

The growing pyramidal network is the network memory self-adapting to the structure of incoming information. In selfstructured systems the structure of data adapts to the task (classes of objects are allocated and defined) which results in optimization of the solution. In contrast to neural networks, the adaptation does not require the introduction of aprioristic redundancy of a network. In GPN various combinations of the assigned initial properties are formed, which increase the accuracy of analytical tasks solving. Selfstructured systems allow not only to locate the dependences providing the diagnosis or the forecasting but also to create their logic descriptions.

The researches, operating on complex large-scale data, have shown high efficiency in applying the growing pyramidal networks for solving the analytical tasks. Such properties as simplicity of modification, combining the input of information with classification, generalization and allocation of essential attributes, high associativity, all make the growing pyramidal networks an indispensable component of intellectual systems.

Bibliography

- Pospelov D.A. Logic-linguistic models in control systems.-Moscow: Energoizdat.-1981.
- Voronkov G.V., Rabinovich Z.L. Natural environment of memory and thinking: modelling representation. Proceedings of international conference. "Knowledge - Dialogue-Solution"-2001.-SPb.-2001.
- Gladun V.P. Partnership with computer. Man-Computer Purposeful Systems.-Kiev: Port-Royal. - 2000.
- V.P.Gladun. Processes of New Knowledge Formation. Sofia: SD Pedagog, 1994, 192 p.
- V.P.Gladun. Planning of Solutions. Kiev: Naukova Dumka. 1987. 168 p.
- Gladun V.P. and Vashchenko N.D. Analytical processes in pyramidal networks // Intern. Journal on Information Theories and Applications. FOI-COMMERCE, Sofia.-2000.-Vol.7, - №3.
- Kiselyova N., Gladun V., Vashchenko N. Computational Materials Design Using Artificial Intelligence Methods. Journal of Alloys and Compounds. 279 (1998), pp. 8-13.
- www.aduis.com.ua <<http://www.aduis.com.ua>>
- Kiseleva N.N. [editor V.S.Zemskov] Computer designing of inorganic compounds: use of databases and methods of artificial intelligence; Institute metallurgy and science of material named for A.A.Bajkov.-of M.: Nauka.-2005.

Authors' Information

Victor Gladun, Vitaly Velichko – V.M.Glushkov Institute of cybernetics of NAS of Ukraine, Prospekt akad. Glushkova 40, 03680 Kiev, Ukraine; e-mail: glad@aduis.kiev.ua

Yurii Ivaskiv – National Aviation University, Prospekt Kosmonavta Komarova 1, 03058, Kiev, Ukraine.

ИЕРАРХИЧЕСКОЕ ЛОГИЧЕСКОЕ ОПИСАНИЕ И НЕЙРОСЕТЕВОЕ РАСПОЗНАВАНИЕ СЛОЖНЫХ ОБРАЗОВ

Татьяна Косовская, Адиль Тимофеев

Аннотация: Авторы предложили ранее иерархический способ задания описаний классов при решении задач распознавания образов. В данной работе предлагается развитие и использование таких иерархических описаний для параллельного представления сложных образов на базе много-ядерных компьютеров или нейронных сетей.

Ключевые слова: сложные образы, логическое описание классов, нейронная сеть.

Введение

В работах [1–4] авторами был предложен логико-аксиоматический подход к решению ряда задач распознавания образов. В работах [1, 2] в рамках этого подхода был предложен иерархический способ задания описаний классов, предназначенный в первую очередь для уменьшения сложности решения рассмотренных задач. В настоящей работе предлагается использовать такие иерархические описания для классов (ОК) для их параллельного представления на базе нейронных сетей. Нейросетевое представление логических описаний классов и решающих правил обеспечивает массовый параллелизм и высокое быстродействие при распознавании сложных образов.

Для построения логических ОК было предложено использовать следующие методы и технологии:

- словесное (лингвистическое) описание классов, взятое из специальной учебной, научной или справочной литературы [2–4];

- непосредственное описание обучающей выборки в терминах заданных предикатов [4];
- логико-частотный метод [3, 4];
- тестовые таблицы [5];
- диофантовые нейронные сети [6–8];
- мульти-агентные технологии принятия решений коллективом нейронных сетей [9].

1. Постановка задач логического распознавания сложных образов

Пусть на частях (фрагментах) x , входящих в объекты произвольной природы ω из множества Ω , задан набор предикатов, характеризующих свойства и отношения между элементами, фрагмента x объекта $\omega \in \Omega$.

Пусть задано разбиение множества Ω на M (непересекающихся или, возможно, пересекающихся) классов (образов) вида

$$\Omega = \bigcup_{k=1}^M \Omega_k.$$

Описанием $S(\omega)$ объекта ω называется набор атомарных формул вида $p_i(x)$ или $\neg p_i(x)$, истинных для x , выписанных для всех возможных частей (фрагментов) x объекта ω .

Описанием класса (ОК) Ω_k называется такая логическая формула $A_k(\omega)$, что

1. $A_k(\omega)$ содержит в качестве атомарных только формулы вида $p_i(x)$, где $x \subseteq \omega$;
2. $A_k(\omega)$ не содержит кванторов;
3. если истинна формула $A_k(\omega)$, то $\omega \in \Omega_k$.

ОК всегда может быть записано в виде логической формулы

$$\bigvee_{j=1}^{J_k} \&_{i \in I_k^j} \&_{x \subseteq \omega_{i,j}} p_i^{\alpha_{ijx}}(x), \quad (1)$$

где J_k – натуральное число, $I_k^j \subseteq \{1, \dots, n\}$, $\omega_{i,j} \subseteq \omega$, α_{ijx} – логические константы. Выражение p^α используется как сокращение для записи p или $\neg p$ в зависимости от того, является ли α константой И (1) или Л (0) соответственно.

С помощью построенных логических ОК предлагается решать следующие задачи распознавания простых и сложных образов:

1. Задача идентификации. Определить, принадлежит ли объект ω или его часть x классу Ω или его часть классу Ω_k .

Эта задача в работах [3, 4] сведена к доказательству выводимости формулы

$$\exists y (y \subseteq \omega \& A_k(y)) \quad (2)$$

из описания распознаваемого объекта $S(\omega)$.

2. Задача классификации. Найти все такие номера классов k , что $\omega \in \Omega_k$.

Эта задача в работах [3, 4] сведена к доказательству выводимости формулы

$$\bigvee_{j=1}^M A_k(\omega) \quad (3)$$

из описания распознаваемого объекта $S(\omega)$ с указанием всех таких номеров k , для которых соответствующий дизъюнктивный член истинен на ω .

3. Задача анализа сложного объекта. Найти и классифицировать все части x объекта ω , для которых $x \in \Omega$.

Эта задача в работах [3, 4] сведена к доказательству выводимости формулы

$$\bigvee_{j=1}^M \exists x(x \subseteq \omega \ \& \ A_k(y)) \quad (4)$$

из описания распознаваемого сложного объекта $S(\omega)$ с указанием (локализацией) и идентификацией всех частей сложного объекта ω , поддающихся классификации, т.е. определить принадлежность к тому или иному классу каждого выделенного фрагмента.

Число шагов работы алгоритма, строящего вывод указанных формул, особенно в случае решения задачи анализа сложного объекта, может быть достаточно велико. Для сокращения числа шагов решения задач распознавания в [2] было предложено более экономное иерархическое описание классов, допускающее распараллеливание при реализации на много-ядерных компьютерах и нейронных сетях.

2. Иерархическое логическое описание классов

Рассматриваются объекты, структура которых позволяет выделить более простые их составляющие, т.е. фрагменты, и дать описание объекта в терминах свойств этих составляющих (фрагментов) и отношений между ними. В частности, это можно сделать, выделяя "часто" встречающиеся подформулы формул $A_k(\omega)$ "небольшой сложности". При этом формируется система эквивалентностей вида

$$p_i^l(x) \leftrightarrow P_i^l(x), \quad (5)$$

где p_i^l – предикаты l -го уровня, P_i^l – подформулы формул $A_k(\omega)$.

Более точно, будем говорить, что система исходных предикатов p_1, \dots, p_n определяет свойства и отношения первого уровня и иногда писать p_i^1 вместо p_i .

Зафиксируем целые положительные числа r и N . Они будут характеризовать такие нечеткие понятия, как "небольшая сложность" подформулы (количество переменных в подформуле меньше r) и "часто" (количество вхождений, с точностью до имен переменных, данной подформулы в уже имеющуюся больше N).

Выделим все часто встречающиеся подформулы небольшой сложности формул $A_k(\omega)$ и обозначим их $P_i^2(x)$ (где x – список переменных, входящих в подформулу).

Обозначим предикаты, задаваемые этими подформулами посредством $p_i^2 (i = 1, \dots, n_2)$ и будем называть их составными предикатами второго уровня. Эти предикаты определяются соотношениями

$$p_i^2(x) \leftrightarrow P_i^2(x) \quad (6)$$

Обозначим формулы, полученные из $A_k(\omega)$ путем замены всех вхождений формул вида $P_i^2(x)$ на атомарные формулы $p_i^2(x)$ (при $x \subseteq \omega$) через $A_k^2(\omega)$. Такие формулы можно рассматривать как логические ОК в терминах предикатов первого и второго уровней.

Процедуру выделения часто встречающихся (типичных) подформул небольшой сложности можно повторить с формулами $A_k^2(\omega)$.

Пусть имеются составные предикаты 1-го, 2-го, ..., l -1-го уровней.

Выделим все часто встречающиеся (типичные) подформулы небольшой сложности формул $A_k^{l-1}(\omega)$ и обозначим их $P_i^2(x)$. Здесь x – список переменных, входящих в подформулу).

Обозначим предикаты, задаваемые этими подформулами посредством $p_i^l (i = 1, \dots, n_l)$ и будем называть их составными (сложными) предикатами l -го уровня.

Эти предикаты определяются соотношениями

$$p_i^l(x) \leftrightarrow P_i^l(x) \quad (7)$$

Формулы, полученные из $A_k^{l-1}(\omega)$ с помощью замены всех вхождений формул вида $P_i^l(x)$ на атомарные формулы $p_i^l(x)$, обозначим через $A_k^l(\omega)$. Такие формулы можно рассматривать как логические описания классов в терминах предикатов 1-го, 2-го, ..., l -го уровней.

Прекратить создание составных (сложных) предикатов очередного уровня можно в любой момент, но не позднее текущей ситуации, когда либо длины всех формул $A_k^l(\omega)$ меньше r , либо среди этих формул не найдется N подформул одинакового вида. В этом заключается правило остановки предлагаемого конечно-сходящегося алгоритма иерархического ОК.

В результате построения составных (сложных) предикатов и многоуровневого логического ОК исходная система ОК может быть записана с помощью равносильной ей многоуровневой логической системы ОК вида

$$\begin{array}{ccc} & A_k^L(\omega), & \\ p_1^2(\omega) & \Leftrightarrow & P_1^2(\omega), \\ & \vdots & \\ p_{n_2}^2(\omega) & \Leftrightarrow & P_{n_2}^2(\omega), \\ & \vdots & \\ p_i^l(\omega) & \Leftrightarrow & P_i^l(\omega), \\ & \vdots & \\ p_{n_L}^L(\omega) & \Leftrightarrow & P_{n_L}^L(\omega). \end{array} \quad (8)$$

Можно показать, что при достаточно удачном выборе параметров r и N время работы алгоритмов, решающих различные задачи распознавания уменьшится [1].

3. Нейросетевое представление иерархического описания классов

При построении нейронных сетей для распараллеливания представлений логических ОК и распознавания сложных образов вычисление значений предикатов того или иного уровня может производиться соответствующим нейронным слоем или ядром много-ядерного компьютера. Так, например, при использовании только предикатов первого (исходного) и второго (синтезированного) уровней архитектура многослойной нейронной сети будет выглядеть так, как это представлено на рис. 1.

Таким образом, в первом слое (сенсорные нейроны) вычисляются все возможные значения исходных признаков распознаваемого объекта. Во втором слое вычисляются все возможные значения признаков второго уровня. В третьем слое проверяются формулы, задающие логические ОК.

В процессе распознавания образов сложный объект или его фрагмент будут отнесены к тому классу с

номером k , для которого оказалась истинной формула $A_k^2(\omega)$.

Следует отметить, что процесс логического распознавания сложных образов на нейронной сети можно организовать таким образом, чтобы запоминать те фрагменты исходного объекта ω , для которых оказалась верна хотя бы одна из формул, определяющих принадлежность к классу. Тем самым может быть автоматически решена задача полного логического анализа сложного объекта.

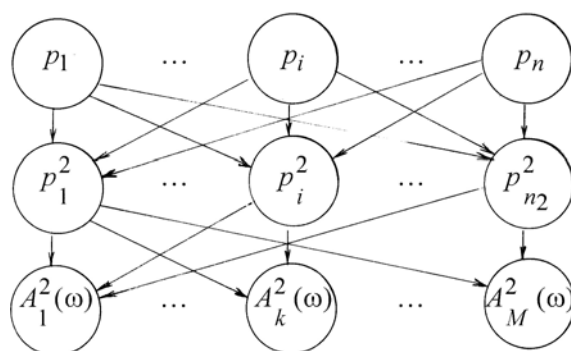


Рис.1. Архитектура многослойной нейронной сети

Заключение

Таким образом, для сложных объектов (сложные изображения и сцены, смешанные сигналы и т.п.) предложены методы иерархического описания классов (образов) и оценки их эффективности в задачах логического анализа и распознавания образов. Описана информационная технология нейросетевого или много-ядерного представления иерархических описаний классов (образов), обеспечивающая массовый параллелизм при обработке данных и значительное ускорение процессов принятия решений в процессе распознавания образов.

Результаты получены при поддержке грантов РФФИ № 05-01-08044-офи-а и № 06-08-01612-а.

Список литературы

- [1]. Косовская Т.М. Оценки сложности решения задач распознавания при иерархическом описании классов – Тез. докл. на XI Междунар. конф. по проблемам теоретической кибернетики. Ульяновск, 1996.
- [2]. Косовская Т.М. Многоуровневые описания классов для принятия решений в задачах распознавания образов – Тр. III Междунар. конф. "Дискретные модели в теории управляющих систем". М., Диалог — МГУ, 1998.
- [3]. Тимофеев А.В. Роботы и искусственный интеллект. – М., "Наука", 1978. 191 с.
- [4]. Косовская Т.М., Тимофеев А.В. Об одном новом подходе к формированию логических решающих правил – Вестник ЛГУ, 1985, №8. С. 22–29. No 1, 160–176.
- [5]. Яблонский С.В. Тест – Математическая энциклопедия, т.5. М., 1985. С. 342–346. Publication. John Wiley & Sons, Inc. 2000. 491p.
- [6]. Тимофеев А.В., Каляев А.В. Методы обучения и минимизации сложности когнитивных нейромодулей супермакро-нейрокомпьютеров с программируемой архитектурой. – Доклады Академии наук, 1994, т. 273, № 2, с. 180–183 (на русском и английском языках).
- [7]. Тимофеев А.В. Методы синтеза диофантовых нейросетей минимальной сложности. – Доклады Академии наук, 1995, т. 301, № 3, с. 1106–1109 (на русском и английском языках).
- [8]. Timofeev A.V., Semyonov A.V. Genetic Algorithms of Database Control and Knowledge Base Synthesis and Their Applications. – International Journal of Information Theories & Applications, Sofia, 1996, v.4, N1, pp. 17–22.
- [9]. Тимофеев А.В., Шеожев А. М., Шибзухов З. М. Мульти-агентные диофантовые нейронные сети в задачах распознавания и диагностики. – Нейрокомпьютеры: разработка и применение, 2005, № 10–11, с. 69–74.

Информация об авторах

Косовская Татьяна Матвеевна – канд. физико-математических наук, доцент, докторант Санкт-Петербургского института информатики и автоматизации Российской академии наук (СПИИРАН); e-mail: kosov@NK1022.spb.edu

Тимофеев Адиль Васильевич – докт. техн. наук, проф., Засл. деятель науки РФ, зав. лаб. информационных технологий в управлении и робототехнике СПИИРАН, зав. базовой кафедрой нейроинформатики и робототехники ГУАП; e-mail: tav@iias.spb.su

DATA MINING WITH FUZZY ARTMAP NEURAL NETWORKS: PREDICTION OF PROFILES OF POTENTIAL CUSTOMERS

Anatoly Nachev

Abstract: The task how to predict profiles of potential customers for a product is important for a direct mailing company. A good prediction allows the company to detect potential customers and to avoid unnecessary and unwanted mailing. This paper describes a non-linear methodology to predict profiles of potential customers using Fuzzy ARTMAP neural networks. The paper discusses advantages of the proposed approach over similar techniques.

Keywords: Fuzzy ARTMAP, neural networks, data mining

ACM Classification Keywords: F.1.1 Models of Computation - neural networks, H.2.8 Database Applications - data mining

Introduction

The task how to predict the profiles of potential customers for a product, given information about the clients and a test sample of customers possessing the particular product is a well-known data mining problem from the world of direct marketing. Direct mailings to a company's potential customers, or 'junk mail' to many can be a very effective way for to market a product or service. However, much of this junk mail is really of no interest to the majority of people that receive it.

A lot of techniques, both linear and non-linear, have been used to predict those customers who are likely to respond or purchase a product, for example statistical methods such as linear regression, decision trees, MPL neural networks (NNs), etc.

The prediction task discussed in this paper, in other words the underlying problem, is to find a subset of customers with a probability of having a caravan insurance policy above some boundary probability. Those customers can be targeted by mailing promotional materials. The boundary depends on the cost and benefits such as of the costs of mailing and benefits of selling insurance policies. The dataset used for experiments is based on real world business data [Van Der Putten, 2000]. It is a block of very detailed survey information on the people, some of whom bought and plan to buy a caravan insurance policy. The people were asked to answer 85 questions, each of which can be regarded as one feature in the classification. The block of data consists of 3 parts. The first is training data, which contains a number of survey responses, some of which come from caravan policy holders. The second part is testing data, and it contains answers from potential caravan insurance policy buyers. The last part is the true data that shows who of those potential buyers actually bought the policy at last. The maximum number of policy owners that could be found is 238. If a random selection is applied, average results provide 42 policy owners, or a hit rate (percentage of real policy buyers out of all predictions made) of 6%.

A wide variety of methodological approaches were used to solve this prediction task. Methods include: standard statistics [Van Der Putten, 2000], backpropagation MLP neural networks [Brierly, 2000], [Crocoll, 2000], [Shtovba et al., 2000], self-organizing maps (SOMs) [Vesanto et al., 2000], genetic programming, C4.5, CART, and other decision tree induction algorithms, fuzzy clustering and rule discovery, support vector machines (SVMs), logistic regression, boosting and bagging, all described in [Van Der Putten, 2000]. The best predictive technique reported in [Elkan, 2001] and [Van Der Putten, 2000] is the Naive Bayesian learning. It has been tested on 800 predictions and gives a hit rate about 15.2%. Predictors based on the backpropagation MLP networks show accuracy rate about 71% and hit rate about 13% as reported in [Brierly, 2000], [Candocia, 2004], [Crocoll, 2000], and [Van Der Putten, 2000].

This paper proposes a non-linear approach based on Fuzzy ARTMAP neural networks to solve the prediction task outlined above.

Section 1 outlines the prediction task and variety of approaches to solve it.

Section 2 discusses the main characteristics of a predictor based on the Fuzzy ARTMAP model and outlines how it functions.

Section 3 describes the preprocessing steps needed to prepare an input dataset for a Fuzzy ARTMAP network.

Section 4 describes experiments conducted to solve the prediction task by a Fuzzy ARTMAP simulator and discusses experimental results.

Predictors Based on ART Neural Networks

Adaptive Resonance Theory (ART) began with an analysis of human cognitive information processing [Grossberg, 1976]. Fundamental computational design goals have therefore always included memory stability with fast or slow learning in an open and evolving input environment. As a real-time model of dynamic processes, an ART network is characterized by a system of ordinary differential equations, which are approximated by an algorithm for implementation purposes [Grossberg, 1980].

ART is a family of neural networks for fast learning, pattern recognition, and prediction, including both unsupervised: ART1, ART2, ART2-A, ART3, Fuzzy ART, Distributed ART; and supervised: ARTMAP, Fuzzy ARTMAP, ART-EMAP, ARTMAP-FTR, Distributed ARTMAP, and Default ARTMAP systems.

ARTMAP architectures are neural networks that develop stable recognition codes in real time in response to arbitrary sequences of input patterns. They were designed to solve the stability-plasticity dilemma that every intelligent machine learning system has to face: how to keep learning from new events without forgetting previously learned information. ARTMAP networks were designed to accept binary input patterns [Carpenter et al., 1991]. ARTMAP networks consist of two ART1 networks, ARTa and ARTb, bridged via an inter-ART module, as shown on Figure 1. An ART module has three layers: the input layer (F0), the comparison layer (F1), and the recognition layer (F2) with m , m and n neurons, respectively. The neurons, or nodes, in the F2 layer represent input categories. The F1 and F2 layers interact with each other through weighted bottom-up and top-down connections, which are modified when the network learns. There are additional gain control signals in the network that regulate its operation.

Fuzzy ARTMAP, introduced in [Carpenter et al., 1992], is a natural extension to ARTMAP. Fuzzy ARTMAP is completely equivalent to ARTMAP, when the input domain is the Hamming cube $\{0,1\}$. It is capable of forming associative maps between clusters of its input and output domains in a supervised manner.

Each module features its own set of parameters, whose values can be assigned independently. ARTa is clustering the input feature space and ARTb the output feature space. The inter-ART's role is to establish the correct association between input and output categories (cluster associations)

The Fuzzy ARTMAP networks have been found useful in pattern recognition, because classification may be viewed as a many-to-one mapping task that entails clustering of the input space and then association of the produced clusters with a limited number of class labels (output clusters that encode a single class label).

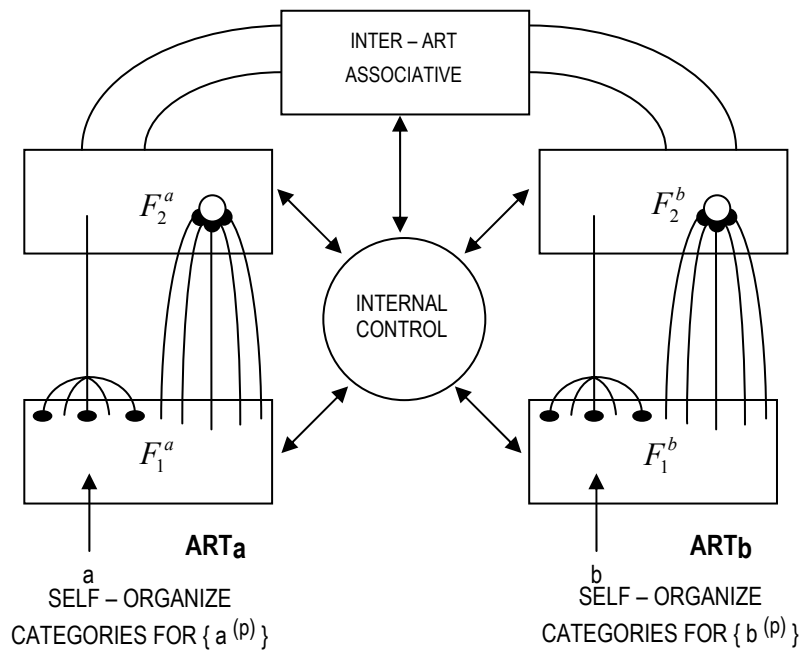


Figure 1. Components of an ARTMAP system.

Many applications of supervised learning systems such as Fuzzy ARTMAP are classification problems, where the trained system tries to predict a correct category given a test set input vector.

From another hand, when multi-layer perceptron (MLP) NNs have been used for classification problems, they employ slow off-line learning to avoid catastrophic forgetting in an open input environment, which limits adaptation for each input and so requires multiple presentations (epochs) of the training set. With fast learning, MLP memories suffer catastrophic forgetting.

Features of a fast-learn system, such as its ability to encode significant rare cases and to learn quickly in the field, may be essential for the given application domain.

Data Pre-processing

The dataset used for simulations is owned and supplied by the data mining company Sentient Machine Research [Van Der Putten, 2000]. It is subdivided into two parts: a train dataset of 5822 customer records and an evaluation dataset of 4000 customer records

Each record consists of 86 attributes containing socio-demographic data represented by attributes (numbers 1-43) and product ownership attributes (numbers 44-86). The socio-demographic data is derived from zip codes of addresses. All customers living in areas with the same zip code have the same socio-demographic attributes. Attribute 86, "CARAVAN: Number of mobile home policies", is the target variable that shows if a customer hold an insurance policy or not.

The evaluation dataset is used for validation of the prediction model. It has the same format as the training data set, but only the target attribute 85 is missing. Targets for the evaluation set have been provided by a separate file.

The prediction task can be solved involving a subset of selected attributes, or features, and their selection is critical for a successful prediction. [Van Der Putten, 2000] reports a variety of selection techniques that rank similarly importance and sensitivity of the attributes in the light of the prediction task. For the purposes of the

experiments reported here all important attributes have been taken on board. Many simulations were conducted to explore how these attributes influence the predictiveness of the Fuzzy ARTMAP model both individually and in groups. Results show that the highest predictive rate can be achieved by a set of the following features: (numbers correspond to the original dataset notations):

$S = \{1, 5, 12, 16, 18, 25, 30, 32, 34, 37, 42, 43, 44, 47, 59, 61, 65, 68, 80, 82, 85\}$. See full feature description in Table 1.

No	Feature Name and Description	No	Feature Name and Description
1	MOSTYPE Customer Subtype	43	MKOOKPLA Purchasing power class
5	MOSHOOFD Customer main type	44	PWAPART Contribution private third party insurance
12	MRELOV Other relation	47	PPERSAUT Contribution car policies
16	MOPLHOOG High level education	59	PBRAND Contribution fire policies
18	MOPLLAAG Lower level education	65	AWAPART Number of private third party insurance 1 - 12
25	MSKA Social class A	68	APERSAUT Number of car policies
30	MHHUUR Rented house	80	ABRAND Number of fire policies
34	MAUTO No car	82	APLEZIER Number of boat policies
37	MINKM30 Income < 30.000	85	ABYSTAND Number of social security
42	MINKGEM Average income		

Table 1. Selected features from train and evaluation datasets.

The feature set can be generally interpreted as identifying customers who are:

1. Car owners with high contribution to car policy purchases. Those who do not have a car are unlikely to own a caravan, as they generally require to be towed. Car owners can be readily identified as those having existing car insurance policies.

The amount spent on policies is also important. People who spend more on car insurance are most likely to be caravan policy buyers, and the more they spend, the more likely a buyer they are.

2. People having fire policy with high level of contribution. This may indicate that the fire insurance is for a caravan. The level of the fire insurance cover that is most likely to be indicative of a caravan policy is level 4.

3. People having a high level of purchasing power. People with high purchasing power are not necessarily enthusiastic about insuring their property, but they do have quite enough wealth to own a caravan, even if using it were not their prime hobby. Typical customers have high, or at least medium, education, status, social class, and income levels.

Experiments

A number of experiments were conducted using a simulator of the Fuzzy ARTMAP model. All experiments explored how this model solves the prediction task paying attention to the following critical factors:

- Sensitiveness to the order in which features and input patters are submitted. This is due to the fact that some ART models commit LTM nodes differently in different orders.
- Best tuning of the network and using optimal values of the network parameters.
- How network parameters affect the train time, test time, and memory consumption.

To maximize use of the datasets and to avoid bias in the selection of the training and test sets, a cross-validation technique was applied. Cross-validation created N copies of the classifier and tested each on 1/N of the

evaluation dataset, after training it on 1/N-th of the training set. In other words, each classifier makes predictions for its 1/N-th of the data, yielding predictions for the whole set. Cross-validation was applied using N=5.

The first group of experiments showed that the Fuzzy ARTMAP is not sensitive to the order in which features are ordered. The experiments also explored how order of input patterns influences the predictiveness by submitting various randomly generated sequences of input patterns. Results reveal that variations in results are slightly small and can be ignored.

A series of simulations were conducted to explore how network parameters affect the predictiveness of the model. Experiments show that best results can be achieved by the default values of most of the network parameters, namely: $\rho_{test} = 0$, $\alpha = 0.01$, $\varepsilon = -0.001$, and $p = 1.0$. Default value of the parameter $\beta = 1.0$ however does not ensure best performance. Results show that all values in the interval $0.95 \leq \beta \leq 0.97$ provide a better performance with a maximum of correct predictions at $\beta = 0.968$. Confusion matrix is shown in Table 1.

		Predicted		
		No	Yes	
Actual	No	3625	137	3762
	Yes	211	30	238
		3852	167	4000

Table 1. Confusion matrix of prediction made by Fuzzy ARTMAP with feature set S and parameters $\beta = 0.968$; $\rho = 0.94$

The vigilance parameter ρ (Rho-bar) affects the Fuzzy ARTMAP performance by tuning the details and granularity of the clusters, thus changing accuracy of predictions and hit rate. The parameter was set to various values between $0.915 \leq \rho \leq 0.955$ with step of increment 0.005. Figure 2 shows network performance and the best accuracy rate 91.4% obtained at $\rho = 0.94$. Figures 3 and 4 reveal number of total positive predictions and correct positive predictions respectively. Figure 5 shows how vigilance affects the hit rate. In most of the parameter values the model provides a hit rate above 15% with best rate nearly 18% (17.96%) at $\rho = 0.94$. This result exceeds the best hit rate of 15% reported in [Elkan, 2001] and [Van Der Putten, 2000], but direct comparison of results would not be accurate, as they are based on different boundaries of the prediction task. For boundaries where the scale of a direct mailing is comparable with the scale adopted by the experiments reported here, a Fuzzy ARTMAP predictor outperforms the other techniques in terms of hit rate.

Training time is another advantage of the Fuzzy ARTMAP model that should be pointed out. With this prediction task the Fuzzy ARTMAP neural networks train for about 5 seconds in contrast to the MPL networks that require about 35 minutes as reported in [Shtovba et al., 2000]. From another hand the MPL networks outperforms Fuzzy ARTMAP in the test time, but both models respond for less that a second, which them makes equally effective and working in a real time.

All simulations show that the LTM memory used by the Fuzzy ARTMAP model requires about 4.9 KB RAM, which makes predictors based on this model efficient with large scale prediction tasks.

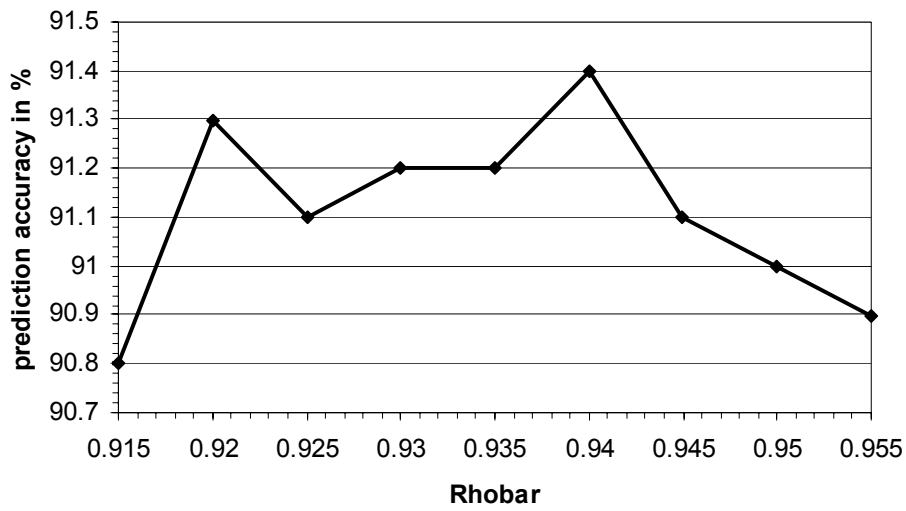


Figure 2. Prediction accuracy (both positive and negative) of Fuzzy ARTMAP neural network.

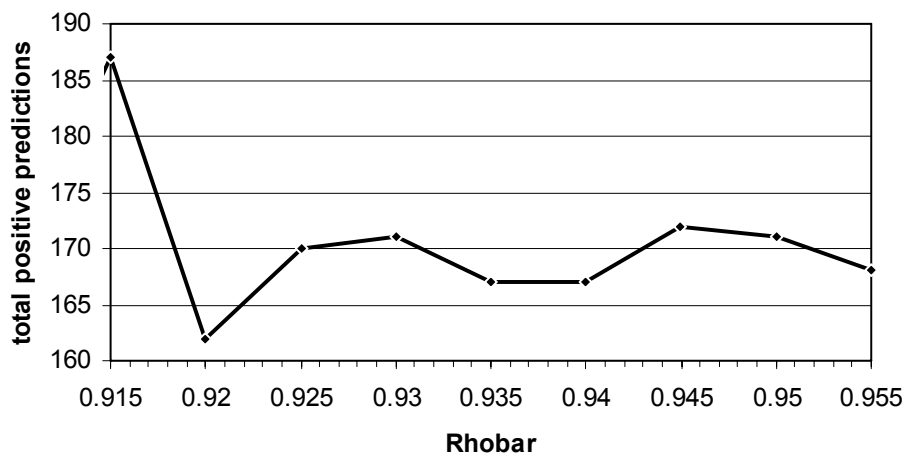


Figure 3. Total positive predictions (both correct and wrong) by Fuzzy ARTMAP neural network.

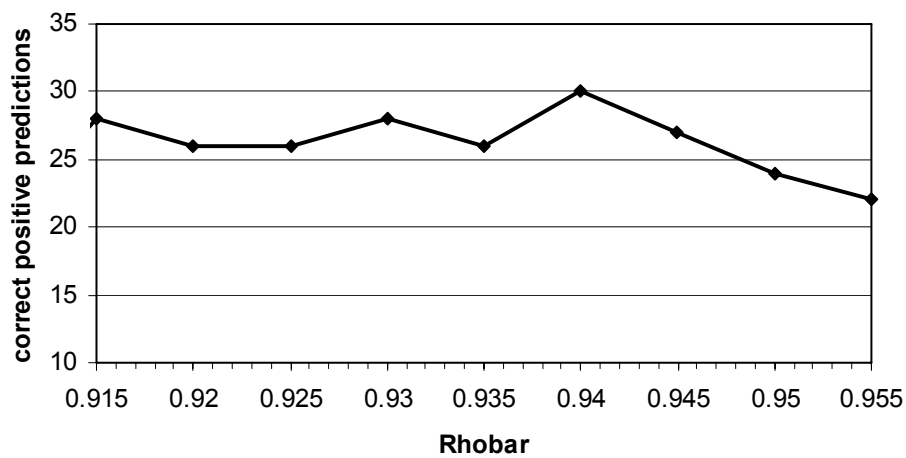


Figure 4. Correct positive predictions by Fuzzy ARTMAP neural network.

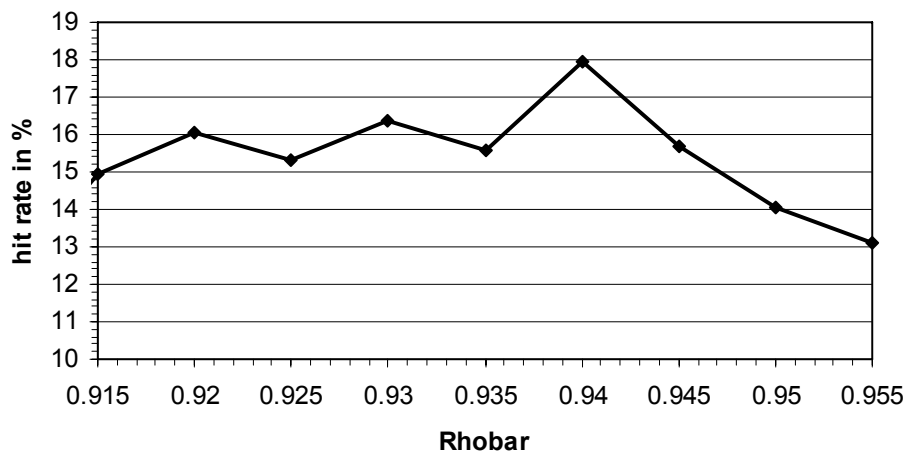


Figure 5. Prediction hit rate by Fuzzy ARTMAP neural network

Conclusion

This paper proposes a non-linear approach, based on Fuzzy ARTMAP neural networks, for solving a prediction task to identify potential buyers of insurance policies. Solution requires an initial processing of the data set to prepare input for a Fuzzy ARTMAP simulator. All conducted experiments lead to the following conclusions:

- Fuzzy ARTMAP model outperforms other predictive techniques, including similar non-linear approaches based on MLP neural networks.
- The model provides stable predictive abilities regardless of order of features submitted and order of input patterns.
- Predictors based on Fuzzy ARTMAP neural networks have a very short training period, in contrast to the MPL neural networks, and small resource consumption, which makes them applicable for large scale prediction tasks.

All conclusions above feature the Fuzzy ARTMAP model as an effective data mining tool for problem areas similar to the prediction task discussed here.

Bibliography

- [Brierly, 2000] P. Brierly, Characteristics of caravan insurance policy owners, Available at <http://www.liacs.nl/~putten/library/cc2000/brierl~1.pdf>, 2000
- [Candocia, 2004] F. Candocia, EEL 6825 Pattern Recognition, Available at http://www.cise.ufl.edu/~bfeng/eel6825/6825_pattern.htm, 2004
- [Carpenter et al., 1991] Carpenter, G.A., Grossberg, S., & Reynolds, J.H. (1991). ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks*, 4, 565–588.
- [Carpenter et al., 1992], Carpenter, G.A., Grossberg, S., Markuzon, N., Reynolds, J.H., & Rosen, D.B., Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps, *IEEE Transactions on Neural Networks*, 3, 698-713
- [Crocoll, 2000] W. Crocoll, Artificial Neural Network Portion of Coil Study, Available at <http://www.liacs.nl/~putten/library/cc2000/crocol~1.pdf>, 2000
- [Elkan, 2001] C. Elkan, Magical Thinking in Data Mining: Lessons From CoLL Challenge 2000. In *Proceedings of the Seventh International Conference on Knowledge Discovery and Data Mining (KDD'01)*, pp. 426-431.

-
- [Grossberg, 1976] Grossberg, S. Adaptive pattern classification and universal recoding. II: Feedback, expectation, olfaction, and illusions. *Biological Cybernetics*, 23, 187–202, 1976.
- [Grossberg, 1980] Grossberg S. (1980). How does a brain build a cognitive code? *Psychological Review*, 87, 1–51., 1980
- [Shtovba et al., 2000] S. Shtovba, Y. Mashnitskiy, The Backpropagation Multilayer Feedforward Neural Network Based Competition Task Solution, Available at <http://www.liacs.nl/~putten/library/cc2000/shtob~1.pdf> , 2000
- [Van Der Putten, 2000] P. Van Der Putten and M. van Someren (eds). CoIL Challenge 2000: The Insurance Company Case. Published by Sentient Machine Research, Amsterdam. Also a Leiden Institute of Advanced Computer Science Technical Report 2000-09. June 22, 2000.
- [Vesanto et al., 2000] J. Vesanto, J. Sinkonen, Submission for the CoIL Chalange 2000, Available at <http://www.liacs.nl/~putten/library/cc2000/vesant~1.pdf>, 2000
-

Author's Information

Anatoli Nachev – Information Systems, Dept. of A&F, NUI Galway, Ireland; e-mail: anatoli.nachev@nuiqalway.ie

ALMOST SEPARABLE DATA AGGREGATION BY LAYERS OF FORMAL NEURONS¹

Leon Bobrowski

Abstract: Information extraction or knowledge discovery from large data sets should be linked to data aggregation process. Data aggregation process can result in a new data representation with decreased number of objects of a given set. A deterministic approach to separable data aggregation means a lesser number of objects without mixing of objects from different categories. A statistical approach is less restrictive and allows for almost separable data aggregation with a low level of mixing of objects from different categories. Layers of formal neurons can be designed for the purpose of data aggregation both in the case of deterministic and statistical approach. The proposed designing method is based on minimization of the of the convex and piecewise linear (CPL) criterion functions.

Key words: data aggregation, layers of formal neurons, separability principles

1. Introduction

Data exploration or data mining tools should allow for efficient discovering of regularities (*patterns*) in large data sets. Data models can be designed on the basis of such patterns. Data exploration tools can be based on variety of methods of multivariate data analysis or pattern recognition [1], [2], [3]. In these approaches, each object or event is typically represented as a feature vector or as a point in a multidimensional feature space. Feature vectors are often divided by experts into categories in such a way that each vector belongs to no more than one category (*class*). In this way the reference (*learning*) set can be generated for each category. For example, teams of medical experts want to obtain such representative learning set for each important disease. Such a set should contain a large number of multidimensional vectors representing particular patients linked to this disease.

In the presented paper, data aggregation term means a reduction of numbers of different feature vectors in learning sets resulting from nonlinear transformation of these vectors. Such transformations may cause merging of a large number of different feature vectors into the same transformed vector. In the case of separable data

¹ This work was partially supported by the grants KBN 3T11F011 30, by the grant W/II/1/2007 from the Białystok University of Technology and by the grant 16/St/2007 from the Institute of Biocybernetics and Biomedical Engineering PAS.

aggregation, only some feature vectors belonging to the same category are merged. In the case of almost separable data aggregation, a low fraction of feature vectors from different categories can be merged.

Data aggregation can be performed by a layer of formal neurons [4]. The feature vectors are transformed by a layer of formal neurons into vectors with binary components. The dipolar and the ranked strategies of designing separable layers of formal neurons were proposed earlier [5]. The possibility of applying the dipolar and ranked strategies to designing almost separable layers is analyzed in the presented paper.

2. Separable learning sets

Let us assume that m objects O_j ($j = 1, \dots, m$) are represented as the so-called feature vectors $\mathbf{x}_j = [x_{j1}, \dots, x_{jn}]^T$, or as points in the n -dimensional feature space $F[n]$ ($\mathbf{x}_j \in F[n]$). Components (*features*) x_i of the feature vector \mathbf{x} represent numerical results of different measurements on a given object O ($x_i \in \{0, 1\}$ or $x_i \in \mathbb{R}$).

We assume that the feature vector $\mathbf{x}_j(k)$ ($j = 1, \dots, m$) has been labelled in accordance with the object $O_j(k)$ category (*class*) ω_k ($k = 1, \dots, K$). The learning set C_k contains m_k feature vectors $\mathbf{x}_j(k)$ assigned to the k -th category ω_k

$$C_k = \{\mathbf{x}_j(k) \mid j \in I_k\} \quad (1)$$

where I_k is the set of indices j of the feature vectors $\mathbf{x}_j(k)$ assigned to the class ω_k .

Definition 1: The learning sets C_k (1) are *separable* in the feature space $F[n]$, if they are disjoint in this space ($C_k \cap C_{k'} = \emptyset$, if $k \neq k'$). It means that the feature vectors $\mathbf{x}_j(k)$ and $\mathbf{x}_{j'}(k')$ belonging to different learning sets C_k and $C_{k'}$ cannot be equal:

$$(k \neq k') \Rightarrow (\forall j \in I_k) \text{ and } (\forall j' \in I_{k'}) \quad \mathbf{x}_j(k) \neq \mathbf{x}_{j'}(k') \quad (2)$$

We are also considering the separation of the sets C_k (1) by the hyperplanes $H(\mathbf{w}_k, \theta_k)$ in the feature space $F[n]$:

$$H(\mathbf{w}_k, \theta_k) = \{\mathbf{x} \mid \mathbf{w}_k^T \mathbf{x} = \theta_k\}. \quad (3)$$

where $\mathbf{w}_k = [w_{k1}, \dots, w_{kn}]^T \in \mathbb{R}^n$ is the weight vector, $\theta_k \in \mathbb{R}^1$ is the threshold, and $(\mathbf{w}_k)^T \mathbf{x}$ is the inner product.

Definition 2: The feature vector \mathbf{x}_j is situated on the *positive side* of the hyperplane $H(\mathbf{w}_k, \theta_k)$ (3) if and only if $(\mathbf{w}_k)^T \mathbf{x}_j > \theta_k$. Similarly, vector \mathbf{x}_j is situated on the *negative side* of $H(\mathbf{w}_k, \theta_k)$ if and only if $(\mathbf{w}_k)^T \mathbf{x}_j < \theta_k$.

Definition 3: The learning sets (1) are *linearly separable* in the n -dimensional feature space $F[n]$ if each of the sets C_k can be fully separated from the sum of the remaining sets C_i by some hyperplane $H(\mathbf{w}_k, \theta_k)$ (3):

$$(\exists k \in \{1, \dots, K\}) (\exists \mathbf{w}_k, \theta_k) (\forall \mathbf{x}_j(k) \in C_k) \quad \mathbf{w}_k^T \mathbf{x}_j(k) > \theta_k \quad (4)$$

and $(\forall \mathbf{x}_j(i) \in C_i, i \neq k) \quad \mathbf{w}_k^T \mathbf{x}_j(i) < \theta_k$

In accordance with the relation (4), all the vectors $\mathbf{x}_j(k)$ from the set C_k are situated on the positive side of the hyperplane $H(\mathbf{w}_k, \theta_k)$ (3) and all vectors $\mathbf{x}_j(i)$ from the remaining sets C_i are situated on the negative side of this hyperplane.

Linear independence of the feature vectors $\mathbf{x}_j(k)$ is a sufficient condition for linear separability of the learning sets C_k (1) [5]:

Remark 1: If the feature vectors $\mathbf{x}_j(k)$ constituting the learning sets C_k (1) are linearly independent in given feature space $F[n]$, then the sets C_k (1) are linearly separable (4) in this space.

3. Separable layers of formal neurons

The formal neuron $NF(\mathbf{w}, \theta)$ can be defined by the threshold activation function $r_i(\mathbf{w}, \theta; \mathbf{x})$

$$r = r(\mathbf{w}_k, \theta_k; \mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{w}_k^T \mathbf{x} \geq \theta_k \\ 0 & \text{if } \mathbf{w}_k^T \mathbf{x} < \theta_k \end{cases} \quad (5)$$

where $\mathbf{w} = [w_1, \dots, w_n]^T \in \mathbb{R}^n$ is the weight vector, $\theta \in \mathbb{R}^1$ is the threshold, and r is the output.

The layer of L formal neurons $NF(\mathbf{w}_k, \theta_k)$ transforms feature vectors \mathbf{x} into output vectors $\mathbf{r} = [r_1, \dots, r_L]^T$ with L binary components $r_i \in \{0, 1\}$:

$$\mathbf{r} = \mathbf{r}(\mathbf{W}; \mathbf{x}) = [r_1(\mathbf{w}_1, \theta_1; \mathbf{x}), \dots, r_L(\mathbf{w}_L, \theta_L; \mathbf{x})]^T \quad (6)$$

where $\mathbf{W} = [\mathbf{w}_1^T, \theta_1, \dots, \mathbf{w}_L^T, \theta_L]^T$ is the vector of the layer parameters.

The relation (6) determines the transformed vectors $\mathbf{r}_j(k)$ with binary components $r_i(\mathbf{w}_i, \theta_i; \mathbf{x})$.

$$(\forall k \in \{1, \dots, K\}) \quad (\forall \mathbf{x}_j(k) \in C_k) \quad \mathbf{r}_j(k) = \mathbf{r}(\mathbf{W}; \mathbf{x}_j(k)) \quad (7)$$

The transformed learning sets C'_k (1) constitute of the vectors $\mathbf{r}_j(k)$:

$$C'_k = \{\mathbf{r}_j(k)\} \quad (j \in I_k) \quad (8)$$

We are examining the properties of the transformation (7) which assure the separability ($(k \neq k') \Rightarrow \mathbf{x}_i(k) \neq \mathbf{x}_i(k')$) (2) of the transformed sets C'_k (8). Such a property can be based on the concept of the mixed dipoles separation [5].

Definition 3: A pair of feature vectors $(\mathbf{x}_j(k), \mathbf{x}_j(k'))$ creates a *mixed dipole* if and only if these vectors belong to different classes ω_k ($k \neq k'$). Similarly, a pair of vectors from the same class ω_k create the *clear dipole* $(\mathbf{x}_j(k), \mathbf{x}_j(k))$.

Definition 4: The formal neuron $NF(\mathbf{w}_k, \theta_k)$ (5) separates the dipole $(\mathbf{x}_j(k), \mathbf{x}_j(k'))$ if only one vector $\mathbf{x}_j(k)$ or $\mathbf{x}_j(k')$ is situated on the positive side of the hyperplane $H(\mathbf{w}_k, \theta_k)$ (3).

Definition 5: The layer of formal neurons $NF(\mathbf{w}_k, \theta_k)$ (5) is *separable* in respect to the learning sets C_k (1) if and only if the transformed sets C'_k (8) are separable (2) and each feature vector $\mathbf{x}_j(k)$ (1) is situated on the positive side of at least one of the hyperplane $H(\mathbf{w}_k, \theta_k)$ (3).

Lemma 1: The necessary condition for separability (*Def. 5*) of the layer of formal neurons $NF(\mathbf{w}_k, \theta_k)$ (5) is the separation (*Def. 4*) of each mixed dipole $(\mathbf{x}_j(k), \mathbf{x}_j(k'))$ by at least one neuron $NF(\mathbf{w}_k, \theta_k)$ (9) of this layer. [5].

The formal neuron $NF(\mathbf{w}_k, \theta_k)$ (5) is *activated* by the vector $\mathbf{x}_j(k)$ ($r_i(\mathbf{w}_k, \theta_k; \mathbf{x}_j(k)) = 1$) if and only if this vector is situated on the positive side of the hyperplane $H(\mathbf{w}_k, \theta_k)$ (3). In accordance with *Definition 5*, the separable layer should assure that each feature vector $\mathbf{x}_j(k)$ (1) activates at least one neuron $NF(\mathbf{w}_k, \theta_k)$ (5). The separable layer can be designed in a multistage procedure, when at one stage a successive neuron $NF(\mathbf{w}_k, \theta_k)$ (5) is added to the layer. In order to increase the generality of the designed neural layers, the following postulate has been introduced [6]:

Postulate of dipolar designing: The separating hyperplane $H(\mathbf{w}_k, \theta_k)$ (3) should divide the highest possible number of mixed and undivided yet dipoles $(\mathbf{x}_j(k), \mathbf{x}_j'(k'))$ and at the same time the lowest possible number of the clear dipoles $(\mathbf{x}_j(k), \mathbf{x}_j'(k))$ should be divided.

The linear separability (4) of the sets C'_l (11) could also be achieved in the layer of formal neurons $NF(\mathbf{w}_k, \theta_k)$ (5):

Definition 6: The layer of formal neurons $NF(\mathbf{w}_k, \theta_k)$ (5) is *linearly separable* in respect to the learning sets C_k (1) if and only if the transformed sets C'_k (8) are separated (4) by the hyperplanes $H(\mathbf{w}_k, \theta_k)$ (3) and each feature vector $\mathbf{x}_j(k)$ activates at least one of these neurons.

The linear separability (4) of the sets C'_l (11) can be achieved by applying a multistage designing process consistent with the following postulate []:

Postulate of ranked designing: The hyperplane $H(\mathbf{w}_k, \theta_k)$ (3) designed during the l -th stage should separate $((\mathbf{w}_k)^T \mathbf{x}_j(k) > \theta_k)$ as many as possible feature vectors $\mathbf{x}_j(k)$ from one set $C_k[l]$ under the condition that no vector $\mathbf{x}_j(k')$ from the remaining sets $C_k[l']$ ($k' \neq k$) is separated.

The symbol $C_k[l]$ in the above postulate means that the learning set C_k (1) has been reduced as a result of neglecting feature vectors $\mathbf{x}_j(k)$ which have been separated by the hyperplanes $H(\mathbf{w}_k, \theta_k)$ (3) during previous $l - 1$ steps. In the deterministic approach, the designing procedure is stopped during the L -th step if all the sets $C_k[L]$ become empty.

Lemma 2: The layer of L formal neurons $NF(\mathbf{w}_k, \theta_k)$ (5) designed in accordance with the above ranked postulate results in linearly separable (4) transformed sets C'_k (8) [5].

4. Convex and piecewise linear criterion functions (CPL)

The procedure of designing a separable layer can be based on a sequence of minimisation of the convex and piecewise linear (CPL) criterion functions $\Psi_k(\mathbf{w}, \theta)$ [3], [4]. The perceptron criterion function belongs to the CPL family. Let us define the function $\Psi_k(\mathbf{w}, \theta)$ by using the positive G_k^+ and the negative G_k^- sets of the feature vectors $\mathbf{x}_j = [x_{j1}, \dots, x_{jn}]^T$ (1):

$$G_k^+ = \{\mathbf{x}_j\} \quad (j \in J_k^+) \quad \text{and} \quad G_k^- = \{\mathbf{x}_j\} \quad (j \in J_k^-) \quad (9)$$

Each element \mathbf{x}_j of the set G_k^+ defines the positive penalty function $\varphi_j^+(\mathbf{w}, \theta)$

$$(\forall \mathbf{x}_j \in G_k^+) \quad \varphi_j^+(\mathbf{w}, \theta) = \begin{cases} 1 - \mathbf{w}^T \mathbf{x}_j + \theta & \text{if } \mathbf{w}^T \mathbf{x}_j - \theta \leq 1 \\ 0 & \text{if } \mathbf{w}^T \mathbf{x}_j - \theta > 1 \end{cases} \quad (10)$$

Similarly, each element \mathbf{x}_j of the set G_k^- defines the negative penalty function $\varphi_j^-(\mathbf{w}, \theta)$

$$(\forall \mathbf{x}_j \in G_k^-) \quad \varphi_j^-(\mathbf{w}, \theta) = \begin{cases} 1 + \mathbf{w}^T \mathbf{x}_j - \theta & \text{if } \mathbf{w}^T \mathbf{x}_j - \theta \geq -1 \\ 0 & \text{if } \mathbf{w}^T \mathbf{x}_j - \theta < -1 \end{cases} \quad (11)$$

The penalty function $\varphi_j^+(\mathbf{w}, \theta)$ is aimed at positioning the vector \mathbf{x}_j from the set G_k^+ ($\mathbf{x}_j \in G_k^+$) on the positive side of the hyperplane $H(\mathbf{w}_k, \theta_k)$ (3). Similarly, the function $\varphi_j^-(\mathbf{w}, \theta)$ should set the vector \mathbf{x}_j from the set G_k^- ($\mathbf{x}_j \in G_k^-$) on the negative side of this hyperplane.

The criterion function $\Psi_k(\mathbf{w}, \theta)$ is the positively weighted sum of the penalty functions $\varphi_j^+(\mathbf{w}, \theta)$ and $\varphi_j^-(\mathbf{w}, \theta)$

$$\Psi_k(\mathbf{w}, \theta) = \sum_{j \in J_k^+} \alpha_j \varphi_j^+(\mathbf{w}, \theta) + \sum_{j \in J_k^-} \alpha_j \varphi_j^-(\mathbf{w}, \theta) \quad (12)$$

where α_j ($\alpha_j > 0$) are the positive parameters (*prices*).

The criterion function $\Psi_k(\mathbf{w}, \theta)$ belongs to the family of the convex and piecewise linear (*CPL*) criterion functions. Minimization of the function $\Psi_k(\mathbf{w}, \theta)$ allows to find optimal parameters $(\mathbf{w}_k^*, \theta_k^*)$:

$$\Psi_k^* = \Psi_k(\mathbf{w}_k^*, \theta_k^*) = \min \Psi_k(\mathbf{w}, \theta) \geq 0 \quad (13)$$

The basis exchange algorithms which are similar to the linear programming allow to find the minimum of the criterion function $\Psi_k(\mathbf{w}, \theta)$ efficiently, even in the case of large, multidimensional data sets G_k^+ and G_k^- (29) [5].

The parameters $(\mathbf{w}_k^*, \theta_k^*)$ constituting the minimum of the function $\Psi_k(\mathbf{w}, \theta)$ (12) define the k -th neuron $NF(\mathbf{w}_k, \theta_k)$ (5) of the layer and the separating hyperplane $H(\mathbf{w}_k, \theta_k)$ (3). The criterion functions $\Psi_k(\mathbf{w}, \theta)$ can be specified both for the dipolar and for the ranked designing postulates. The specification of the criterion function $\Psi_k(\mathbf{w}, \theta)$ (12) is performed through the choice of adequate sets G_k^+ and G_k^- (9) and the prices α_j related to particular vectors \mathbf{x}_j from these sets.

It has been proved that the minimal value Ψ_k^* (13) of the criterion function $\Psi_k(\mathbf{w}, \theta)$ (12) is equal to zero ($\Psi_k^* = 0$) if and only if the positive G_k^+ and the negative G_k^- sets (9) are linearly separable (4). In this case, all elements \mathbf{x}_j of the set G_k^+ (9) are located on the positive side of the hyperplane $H(\mathbf{w}_k^*, \theta_k^*)$ (3) and all elements \mathbf{x}_j of the set G_k^- are located on the negative side:

$$\begin{aligned} (\forall \mathbf{x}_j \in G_k^+) \quad (\mathbf{w}_k^*)^T \mathbf{x}_j > \theta_k^* \\ \text{and} \quad (\forall \mathbf{x}_j \in G_k^-) \quad (\mathbf{w}_k^*)^T \mathbf{x}_j < \theta_k^* \end{aligned} \quad (14)$$

If the sets G_k^+ and G_k^- (9) are not linearly separable (4), then $\Psi_k^* > 0$ and the inequalities (14) are fulfilled only partly, not by all, but by a majority of elements \mathbf{x}_j of the sets (9).

Minimization of the function $\Psi_k(\mathbf{w}, \theta)$ (12) allows one to find the optimal parameters $(\mathbf{w}_k^*, \theta_k^*)$ defining such hyperplane $H(\mathbf{w}_k^*, \theta_k^*)$ (3) which relatively well separates two sets G_k^+ and G_k^- (9). The parameters $(\mathbf{w}_k^*, \theta_k^*)$ can be also used in the definition of the l -th element $NF(\mathbf{w}_k^*, \theta_k^*)$ (5) of a neural layer.

5. Almost separable layers

A layer of L formal neurons $NF(\mathbf{w}_k, \theta_k)$ (6) can transform a large number of feature vectors $\mathbf{x}_j(k)$ ($k = 1, \dots, K$) (1) into the same output vector $\mathbf{r}_l = [r_{l1}, \dots, r_{lL}]^T$ (6) with L binary components $r_l \in \{0, 1\}$, where different indexes l and l' mean different vectors \mathbf{r}_l and $\mathbf{r}_{l'}$:

$$(l \neq l') \Rightarrow (\mathbf{r}_l \neq \mathbf{r}_{l'}) \quad (15)$$

The set of such feature vectors $\mathbf{x}_j(k)$ (1) which are transformed into the same output vector \mathbf{r}_l is called the l -th *activation field* SI of the layer of formal neurons:

$$SI = \{ \mathbf{x}_j(k): [r_{l1}(\mathbf{w}_1, \theta_1; \mathbf{x}_j(k)), \dots, r_{lL}(\mathbf{w}_L, \theta_L; \mathbf{x}_j(k))]^T = \mathbf{r}_l \} \quad (16)$$

Definition 7: The set SI (16) is the *clear activation field* if all feature vectors $\mathbf{x}_j(k)$ (1) from this set ($\mathbf{x}_j(k) \in SI$) belong to the same learning set C_k (1). Similarly, the set SI is the *mixed activation field* if it contains feature vectors $\mathbf{x}_j(k)$ from different sets C_k .

Lemma 3: The layer of L formal neurons $NF(\mathbf{w}_k, \theta_k)$ (5) is separable in respect to the learning sets C_k (Def. 5) if and only if all the activation fields SI (16) of this layer are clear.

It results from the above *Lemma* that the layer with all the clear activation fields SI (16) is separable and preserves the learning sets C_k (1) separability during data aggregation. Even one mixed field SI (16) results in a nonseparability of the layer. The concept of mixed activation fields SI (16) is useful in the analysis of nearly separable layers.

Definition 8: The class ω_k is *dominant* in the active the set SI (16) if and only if the most of the feature vectors $\mathbf{x}_j(k)$ (1) from this set are assigned to the class ω_k .

The activation field $SI(k)$ (16) and the output vector $\mathbf{r}(k)$ is assigned to the dominant class ω_k .

Definition 9: The set SI (16) is the ε -clear (almost clear) activation field if the fraction f of the feature vectors $\mathbf{x}_j(k')$ (1) from nondominant classes $\omega_{k'}$ in this set ($\mathbf{x}_j(k') \in SI(k)$) is less than ε .

$$f(k) = ml' / (ml(k) + ml') < \varepsilon \quad (17)$$

where $ml(k)$ is the number of elements $\mathbf{x}_j(k)$ of the set $SI(k)$ (16) belonging to the dominant class ω_k and ml' is the number of elements $\mathbf{x}_j(k')$ of the set $SI(k)$ belonging to non-dominant classes $\omega_{k'}$ ($k' \neq k$).

All feature vectors $\mathbf{x}_j(k)$ from the l -th activation field SI are *aggregated* by the layer of formal neurons into one output vector \mathbf{r} . In other words, the vector \mathbf{r} *generalizes* all feature vectors $\mathbf{x}_j(k)$ from the field SI (26). It can be expected that the layer of formal neurons with a large and ε -clear activation fields SI (26) could have a great *generalization power*. Such layer could be also used as a classifier with the following decision rule:

$$\text{if } (\mathbf{x}_0 \in SI(k)) \text{ then } \mathbf{x}_0 \in \omega_k \quad (18)$$

The quality of a layer of formal neurons $NF(\mathbf{w}_k, \theta_k)$ (5) can be evaluated by the *error rate* of the decision rule (18). The error rate is often estimated through an *apparent error* er [1]:

$$er = me / m \quad (19)$$

where me is the number of such feature vector $\mathbf{x}_j(k)$ from the sets C_k (1) which are wrongly allocated by the decision rule (18).

Lemma 4: The error rate er (19) of the decision rule (18) is equal to zero if and only if the layer of formal neurons is separable (Def. 5).

The error rate evaluation (19) is positively biased (*optimistic bias*) [1]. The unbiased error rate er evaluations can be based on the technique of cross-validation [3].

6. Designing almost separable layers of formal neurons

The separable layer of formal neurons with the decision rule (18) assures the correct classification of all the feature vectors $\mathbf{x}_j(k)$ from the learning sets C_k (1) and the apparent error rate er (19) is equal to zero ($er = 0$). As it results from previous considerations, if the learning sets C_k (1) are separable (2), the separable layers (Def. 5) can be designed in accordance with the dipolar strategy, and in accordance with the ranked strategy. Unfortunately, a separable layer can cause the *overfitting problem* [3]. This problem is manifested in such a way that despite the fact that the apparent error rate er (19) of the rule (18) evaluated on the vectors $\mathbf{x}_j(k)$ from learning sets C_k (1) is equal to zero, the error rate of this rule on new feature vectors \mathbf{x} that does not belong to the sets C_k (1) is too large.

A layer of formal neurons with the apparent error rate er (19) greater than zero ($er > 0$) could have a larger *discriminative power* than a separable layer ($er = 0$). We shall take into considerations almost separable (ε -separable) layers of formal neurons.

Definition 10: The layer of formal neurons is ε -separable (*almost separable*) if and only if the apparent error rate er (19) is less than ε ($er < \varepsilon$).

Lemma 5: If all the activation fields $Sl(k)$ (16) of the layer of formal neurons are of the ε -clear type (*Def. 9*), then this layer is ε -separable

Proof: The apparent error rate er (19) can be represented in the below manner:

$$er = me / m = (m1' + \dots + mM') / m \quad (20)$$

where mi' (17) is the number of elements $\mathbf{x}_j(k')$ of the activation field $Sl(k)$ (16) belonging to non-dominant classes $\omega_{k'}$ ($k' \neq k$), and M is the number of activation fields. Thus

$$\begin{aligned} er = m1' / (m1(k(1)) + m1') (m1(k(1)) + m1') / m + \dots \dots \\ \dots \dots m1' / (m1(k(M)) + mM') (mM(k(M)) + mM') / m < \varepsilon \end{aligned} \quad (21)$$

and the thesis is proved. \square

The *Lemma 5* gives indications for designing almost separable layers of formal neurons $NF(\mathbf{w}_k, \theta_k)$ (6). Designing process can be based on a generation of such activation fields $Sl(k)$ (16) which are of the ε -clear type (*Def. 9*). As a consequence, the *Ranked designing postulate* can be modified in the below manner:

Postulate of almost ranked designing: The hyperplane $H(\mathbf{w}_k, \theta_k)$ (3) designed during the l -th stage should separate $((\mathbf{w}_k)^T \mathbf{x}_j(k) > \theta_k)$ as many as possible feature vectors $\mathbf{x}_j(k)$ from one set $C_k[l]$ under the condition that the fraction fl (17) of the separated vectors $\mathbf{x}_j(k')$ (1) from other sets $C_{k'}[l]$ ($k' \neq k$) is less than ε .

One can see that all the activation fields $Sl(k)$ (16) of the layer of formal neurons designed in accordance with the above postulate are of the ε -clear type (*Def. 9*).

The *Postulate of dipolar designing* can be also modified in a similar manner, for example by neglecting such mixed dipoles $(\mathbf{x}_j(k), \mathbf{x}_j(k'))$ that the feature vectors $\mathbf{x}_j(k)$ and $\mathbf{x}_j(k')$ belong to the activation field $Sl(k)$ (16) of the ε -clear type (*Def. 9*).

The data aggregation process can be based on a layer of formal neurons $NF(\mathbf{w}_k, \theta_k)$ (6). Let us define the *aggregation coefficient* η_a of such layer in the following manner

$$\eta_a = (m - m(\mathbf{rl})) / (m - K) \quad (22)$$

where, m is the number of the feature vectors $\mathbf{x}_j(k)$ in the sets C_k (1), $m(\mathbf{rl})$ is the number of different output vectors \mathbf{rl} (15) from a separable layer, and K is the number of the classes ω_k or the learning sets $C_k(1)$.

The minimal number $m(\mathbf{rl})$ of the output vectors \mathbf{rl} (15) from a separable layer is equal to K ($m(\mathbf{rl}) = K$). The aggregation coefficient η_a (22) takes the maximal value equal to one ($\eta_a = 1$) in this ideal situation. The aggregation coefficient η_a (22) of a layer of formal neurons $NF(\mathbf{w}_i, \theta_i)$ (5) can take the maximal value $\eta_a = 1$ if and only if the learning sets C_k (1) are linearly separable. The maximal value of the number $m(\mathbf{rl})$ is equal to m . There is no aggregation in this case and the aggregation coefficient η_a (29) takes the minimal value equal to 0 ($\eta_a = 0$). As a result:

$$0 \leq \eta_a \leq 1 \quad (23)$$

It can be noted that a solution of the *Optimization problem* leads to the maximisation of the aggregation coefficient η_a (22).

Optimization problem: To design such ε -separable (Def. 10) layer of formal neurons $NF(\mathbf{w}_k, \theta_k)$ (6) which has a minimal number M of activation fields $Sl(k)$ (16) or different output vectors \mathbf{rl} (15).

The minimal number M of the activation fields Sl (16) can not be less than the number K of the classes ω_k ($M \geq K$).

7. Concluding remarks

Separable layer of formal neurons $NF(\mathbf{w}_k, \theta_k)$ (5) can be induced from the learning sets C_k (1) only when these sets are separable (2) in a given feature space. The dipolar strategy allows for preserving the separability of the sets C_k (1) during their transformation by the induced layer of formal neurons. The ranked strategy also allows to achieve the linear separability (4) of the transformed sets C'_k (8).

Separable layers of formal neurons with the decision rule (18) secure correct classification of all the feature vectors $\mathbf{x}_j(k)$ from the learning sets C_k (1) and the apparent error rate er (19) is equal to zero ($er = 0$) in this case. Unfortunately, such property of the designed layers is often linked with the occurrence of overfitting [3]. Despite the fact that apparent error rate er (19) evaluated on the basis of elements $\mathbf{x}_j(k)$ of the learning sets C_k (1) is equal to zero, the decision rule (18) can be burdened with a too large error rate on new feature vectors \mathbf{x} ($\mathbf{x} \notin C_k$). In this case, the generalization power of the designed network is too low.

Almost separable layer of formal neurons $NF(\mathbf{w}_k, \theta_k)$ (5) allow apparent error rate er (19) greater than zero. Designing almost separable layers should allow for achieving a larger generalization power and level of data aggregation in comparison to strictly separable layers. There are still many problems to resolve in the search for efficient strategy for designing almost separable layer.

Bibliography

- [1]. Johnson R. A., Wichern D. W.: Applied Multivariate Statistical Analysis, Prentice-Hall, Inc., Englewood Cliffs, New York, 1991
- [2]. Duda O. R. and Hart P. E., Stork D. G.: Pattern Classification, J. Wiley, New York, 2001,
- [3]. Fukunaga K., Introduction to Statistical Pattern Recognition, Academic Press 1990,
- [4]. Rosenblatt F.: Principles of neurodynamics, Spartan Books, Washington 1962
- [5]. Bobrowski L.: Data mining based on convex and piecewise linear (CPL) criterion functions, (in Polish), Technical University Bialystok, 2005,
- [6]. Bobrowski L.: "Piecewise-Linear Classifiers, Formal Neurons and separability of the Learning Sets", Proceedings of ICPR'96, pp. 224-228, (13th International Conference on Pattern Recognition", August 25-29, 1996, Vienna)
- [7]. Bobrowski L.: "Design of piecewise linear classifiers from formal neurons by some basis exchange technique" Pattern Recognition, 24(9), pp. 863-870, 1991
- [8]. Bobrowski L., "The ranked neuronal networks", Biocybernetics and Biomedical Engineering, Vol. 12, No. 1-4, pp. 61-75, 1992

Author's Information

Leon Bobrowski – leon@ii.pb.bialystok.pl

Faculty of Computer Science, Bialystok Technical University

Institute of Biocybernetics and Biomedical Engineering, PAS, Warsaw, Poland

COMPLEX NEURAL NETWORK MODEL OF USER BEHAVIOR IN DISTRIBUTED SYSTEMS

Andrii Shelestov, Serhiy Skakun, Olga Kussul

Abstract: We present a complex neural network model of user behavior in distributed systems. The model reflects both dynamical and statistical features of user behavior and consists of three components: on-line and off-line models and change detection module. On-line model reflects dynamical features by predicting user actions on the basis of previous ones. Off-line model is based on the analysis of statistical parameters of user behavior. In both cases neural networks are used to reveal uncharacteristic activity of users. Change detection module is intended for trends analysis in user behavior. The efficiency of complex model is verified on real data of users of Space Research Institute of NASU-NSAU.

Keywords: distributed systems, user behavior model, neural networks.

Introduction

At present the solution of complex large-scale problems arising in the areas of Earth observations from space [Shelestov, *et al.*, 2006], [Fusco, 2006], [Fusco, *et al.*, 2003], high-energy physics [Holtman, 2001], bioinformatics [Peltier, *et al.*, 2002], astronomy [Annis, *et al.*, 2002] is impossible without use of distributed computer systems. (e.g. Grid systems). Many tasks, such as computing and data resources sharing, distributed data processing, data storage, archiving and transfer are relied on them. One of the important challenges in the development of heterogeneous distributed infrastructure is the security provision. For this purpose many problems must be solved such as user authentication, authorization, rights delegation, etc. This can be done by using, for example, Globus Grid Security Infrastructure (GSI) [Foster, *et al.*, 1998] which is an extension of the Public Key Infrastructure [IETF], [Adams and Lloyd, 2002]. On the other hand, there are many monitoring tools intended for resources state and jobs monitoring (e.g. GridICE (<http://gridice.forge.cnaf.infn.it/>), MOGAS (<http://ntu-cg.ntu.edu.sg/pragma/index.jsp>)), but they do not provide monitoring of users' activities in order to detect anomalies and potential intrusions. Though, many sources report that the majority (80%) of information security incidents is perpetrated by insiders [Tulloch, 2003]. Hence, the problem of monitoring and detection of malicious user activity in distributed computer system is an important issue.

Related Works

Nowadays different methods and approaches are applied for the analysis of user activity. They are mostly based on the analysis and exposure of regularities and common actions in user behavior for automation, prediction, anomaly detection, etc. Data that are used for model construction possess individual features that define user behavior. For data processing different approaches can be applied such as history-matching methods and machine-learning methods.

In general, creation of user behavior model involves the following steps: data collection and data pre-processing, when useful information about user activity is collected from log-files; data processing, when feature extraction is done to represent data, and dimension reduction methods are used to reduce the size of the data; application of different techniques to obtain interesting characteristics of user behavior; interpretation of the results.

Among the existing approaches to user activity analysis we may consider so called Personal Security Programs that are used by commercial companies to monitor the activity of their employees. The results of such monitoring can be used to reveal malicious users in the case of information leakage, or to find out whether users use computers for their personal purposes. For example, such programs as PC Spy (www.softdd.com/pcspy/index.htm), Inlook Express (www.jungle-monkey.com), Paparazzi (www.industar.net) allow to capture and save screen images (screenshots) showing exactly what was being viewed by users. All

screens can be captured, including Web pages, chat windows, email windows, and anything else shown on the monitor. However, these programs have some disadvantages; among them are high volume of stored information and manual configuration of snapshots frequency. That is, if the frequency is low it would be rather difficult to find out something abnormal in user activity. Otherwise, a lot of the data should be stored.

Another example refers to Intrusion Detection Systems (IDS), particularly anomaly detection in computer systems. Usually, a model of normal user behavior is firstly created, so during monitoring any abnormal activity can be regarded as potential intrusion. Different approaches are applied to the development of anomaly detection systems: statistical methods [Javitz and Valdes, 1991], expert systems [Dowell and Ramstedt, 1990], finite automata [Kussul and Sokolov, 2003], neural networks [Cannady and Mahaffey, 1998], [Reznik, *et al.*, 1999], [Skakun, *et al.*, 2005], agent-based systems [Skakun, *et al.*, 2005], [Gorodetski, *et al.*, 2001], rule-based networks, genetic algorithms, etc.

It is worth mentioning that existing approaches do not provide adequate description of user behavior. There exist methods that exhibit only dynamical features of user behavior, and do not consider statistical properties, and vice versa. This paper describes a complex model of user behavior in distributed systems. The model consists of three components: on-line model, off-line model and change detection module. The use of on-line and off-line models allows the reflection of both dynamical and statistical features of user's activity. In order to provide adaptive and robust approach for the analysis and generalization of data obtained from user activity neural networks are applied. The proposed approach is verified on real data gathered during the work of users on the resources of Space Research Institute of NASU-NSAU.

Complex Model of Users Behavior in Distributed Systems

For adequate description of user behavior in distributed systems we propose complex model that consists of the following components:

- on-line model that describes user's activity during its work by predicting his actions;
- off-line model that is based on the analysis of statistical data acquired during user's work;
- change detection module intended for detection of trends in user's activity.

The proposed structure of complex model is depicted in Fig. 1.

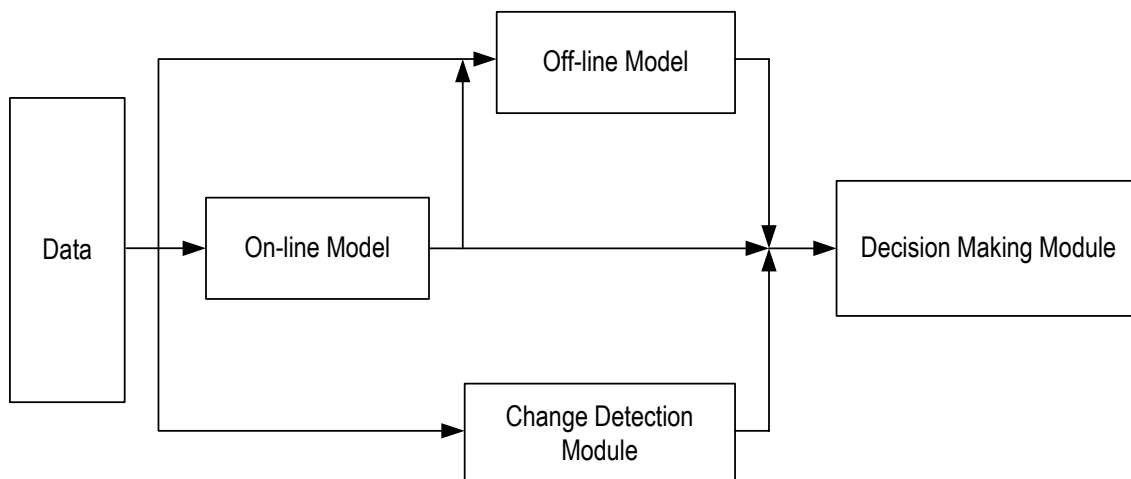


Fig. 1. Structure of complex neural network model of user behavior

On-line model describes dynamical features of user behavior by predicting user actions based on previous ones. For prediction of user actions neural network is used. The use of neural network is motivated by the fact that user behavior represents a complex non-linear process and by the need to reveal regularities in it. As a neural network paradigm, we use feed-forward neural network trained by means of back-propagation algorithm [Haykin, 1999].

Therefore, for each user a neural network is built, and trained in such way to predict user actions. The result of neural network work after completion of $i-1$ actions by user during session s_i is given by the following equation:

$$\widehat{c}_i^{s_i} = F(\mathbf{X}_i), \quad \mathbf{X}_i = (c_{i-1}^{s_i}, c_{i-2}^{s_i}, \dots, c_{i-m}^{s_i}),$$

where F — non-linear transformation performed by neural network; $\mathbf{x}_i, \widehat{c}_i^{s_i}$ — neural network input and output respectively; $c_i^{s_i}$ — the number of i -th user action during session s_i ; m — number of previous actions used to predict the next one.

The decision about user behavior (normal/abnormal) is based on the number of correctly predicted user actions by neural network (i.e. when $\widehat{c}_i^{s_i}$ is equal to $c_i^{s_i}$). Construction of on-line model must also take into account the possibility that user behavior will be changing in the course of time. In order to provide adaptation of model to these changes (i.e. retraining neural network) change detection module is used.

On contrast, *off-line model* is based on the use of statistical (integral) parameters obtained during user behavior. The following set of characteristics about user behavior was taken:

$$\{n_{s_i}, o_{s_i}, h_{s_i}, d_{s_i}, s_{s_i}\}, \quad (1)$$

where n_{s_i} — number of actions performed by user; o_{s_i} — results of on-line model use, i.e. the number of correctly predicted user actions; h_{s_i} — user login host; d_{s_i} — user session duration time; s_{s_i} — the time of user session start.

This set is used as input feature to neural network for detection of normal/abnormal user activity. As in the case of on-line model, for each user feed-forward neural network is trained with back-propagation algorithm in order to distinguish normal and abnormal user behavior. The expected output of neural network during training is binary, i.e. 1 corresponds to normal behavior and 0 corresponds to anomaly. The neural network output is defined as follows:

$$\Delta_{s_i} = F(x_{s_i}), \quad \mathbf{x}_{s_i} = (n_{s_i}, o_{s_i}, h_{s_i}, d_{s_i}, s_{s_i}),$$

where F — non-linear transformation performed by neural network; $\mathbf{x}_{s_i}, \Delta_{s_i}$ — neural network input and output respectively; $n_{s_i}, o_{s_i}, h_{s_i}, d_{s_i}, s_{s_i}$ are defined by (1).

If an input to neural network is an independent sample (that was not used during training process), the corresponding output Δ_{s_i} will lie in the range $[0; 1]$, and provide probability of user normal activity (higher values correspond to normal user behavior).

The user behavior does not represent stationary process, and it will be changing in the course of time (as a rule during 2-3 months). This can be caused by different reasons, e.g. due to software version changes, new tasks accomplishment. That is why, complex model is required to include *change detection module* that will detect trends in user behavior.

Let A be the alphabet of user actions, i.e. the set of all actions performed by user during set of sessions $\{s_t\}_{t=1}^T$.

We assume that user behavior has not been changed during this time of work. Let N be the number of actions in alphabet A , and each action has a number from 1 to N . The number $N+1$ will be reserved to new actions that were not performed during sessions $\{s_t\}_{t=1}^T$. Let s_t ($t > T$) be the current session of user work with the following actions performed $\mathbf{c}^{s_t} = (c_1^{s_t}, c_2^{s_t}, \dots, c_{N_{s_t}}^{s_t})$. Then $c_i^{s_t} = N+1$, if $c_i^{s_t} \notin \{1, \dots, N\}$. In order to detect changes in user behavior after session s_t we construct vector $\mathbf{g}(s_t)$ with the following components:

$g_j(s_t) = \begin{cases} 1, & \text{if exists such } k = \overline{1, N_{s_t}}, \text{ that } c_k^{s_t} = j, j \in A \\ 0, & \text{otherwise} \end{cases}$. That is, if an action was performed

during session s_t , then corresponding component of vector $\mathbf{g}(s_t)$ is equal to 1, otherwise is equal to 0. Then the obtained vector $\mathbf{g}(s_t)$ is compared in pairs with vectors obtained during previous sessions $s_{t-1}, s_{t-2}, \dots, s_{t-l}$. As a measure of comparison Hamming distance is applied:

$$\aleph(\mathbf{g}(s_{t'}), \mathbf{g}(s_{t''})) = \sum_{j=1}^N \chi(g_j(s_{t'}), g_j(s_{t''})),$$

where $\chi(g_j(s_{t'}), g_j(s_{t''})) = \begin{cases} 1, & \text{if } g_j(s_{t'}) \neq g_j(s_{t''}) \\ 0, & \text{otherwise} \end{cases}$. That is, \aleph corresponds to number of components

of two vectors that are different. As a result of comparisons we will obtain l values, following which we average

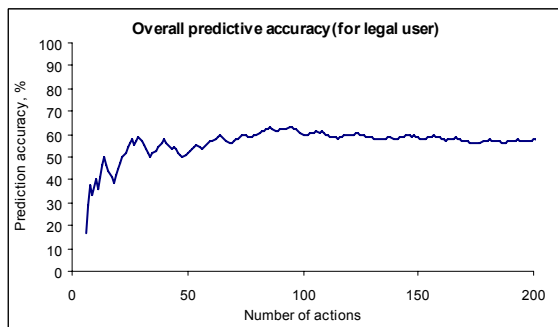
and normalize on N : $H_t = \frac{1}{N} \left(\frac{1}{l} \sum_{k=1}^l \aleph(\mathbf{g}(s_t), \mathbf{g}(s_{t-k})) \right)$. If user behavior has not changed, then vector $\mathbf{g}(s_t)$

would not differ considerably from vectors for previous sessions. Hence, H_t would be below some threshold H : $H_t < H$. And vice versa, if anomaly occurred, vector $\mathbf{g}(s_t)$ would differ considerably from vectors for previous sessions and H_t would be under some threshold H^* : $H_t > H^*$. If $H_t \in (H^*; H)$, then a natural changes in user behavior took place.

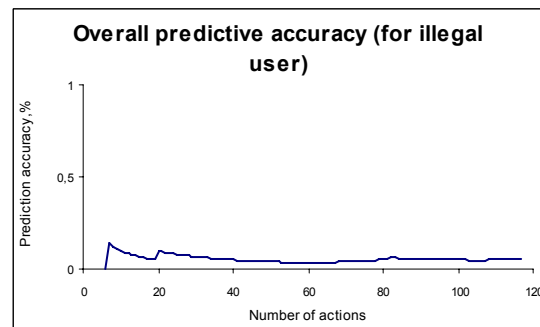
Description of Experiments

Different experiments were run to demonstrate the efficiency of both on-line and off-line models and change detection module. For this purpose data needed for neural network training were acquired during a real work of users on the resources of Space Research Institute of NASU-NSAU.

Experiments for on-line model. For on-line model log files were transformed into format suitable for neural network. That is, for each user an alphabet of user actions was created, and each action was assigned an identifier (decimal number). For neural network input a binary coding was applied (7 bits per action). Feed-forward neural network was used to predict user actions based on 5 previous ones (the value of 5 previous actions was derived using autocorrelation function for sequence of user actions). Thus, the dimension of input data space was 35. In turn, for output data decimal coding was applied, and the dimension of output data space was 1. As to neural network architecture, we used neural network with 3 layers: input layer with 35 neurons, hidden layer with 35 neurons, and output layer with 1 neuron.



(a)



(b)

Fig. 2. Predictive accuracy for: (a) legal user; (b) illegal user

Then all data were randomly mixed and divided into train and test sets (70% for training and 30% for testing). Results of neural network work on test data showed that overall predictive accuracy (that is, the number of correctly predicted actions divided by total number) for different users varied from 33% to 59% (an example of overall predictive accuracy variations within number of actions is depicted in Fig. 2,a). To demonstrate that neural network was able to distinguish one user from another we run so called cross experiments. It was done in two ways. First one consisted in the following: data obtained during the work of one user (name him illegal user) were put to neural network trained for another (legal user). In such a case, overall predictive accuracy of neural network hardly exceeded 5% (it is shown on Fig. 2,b an example where overall predictive accuracy was 0,05%). Such experiment modeled the situation when illegal user logged on and begun to work under account of another user.

The second method of cross experiments was done by inserting data of illegal user into data of legal one. This experiment modeled the situation when intruder begun to work under account of another user already logged on. In such a case, we used short-time predictive accuracy to measure the number of correctly predicted actions (this measure takes into account only last actions performed by user, for example, twenty last actions). Variations of short-time predictive accuracy for both legal and illegal user are depicted in Fig. 3. It is evident that short-time predictive accuracy for illegal user is considerably less then for legal one.

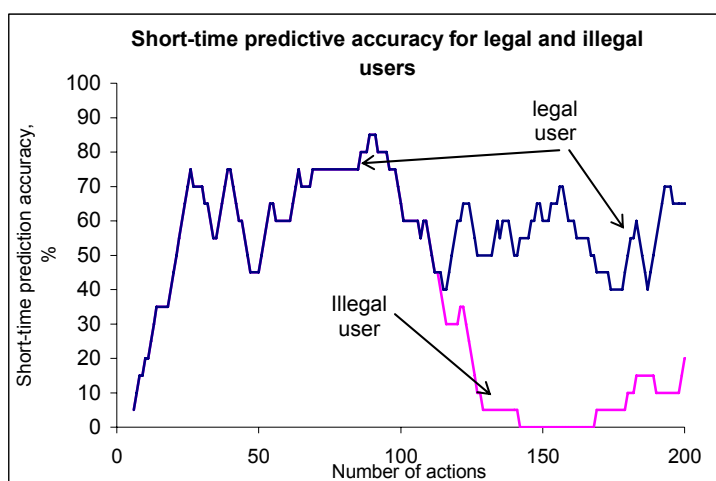


Fig. 3. Short-time predictive accuracy

Experiments for off-line model. Considering off-line model, all needed statistical data were obtained from log files. Then they were encoded, divided into training and test sets, and input to neural network. Results of neural network work on test data gave 80% accuracy of correct user behavior identification. That is, experiments showed that off-line model was able to distinguish normal and abnormal (anomalous) user behavior.

Experiments for change detection module. In order to verify change detection module first H_t was estimated for those user sessions when the use behavior was normal. Derived values of H_t did not differ considerably and lied in the range $(0; H^*)$. (For each user different values of threshold H^* were obtained and varied from 0,1 to 0,17). For anomaly behaviour modelling cross-experiments were run. That is, vector $\mathbf{g}(s_i)$ was calculated for sessions of another user and compared with vectors obtained for a given user. Obtained values of H_t considerably increased (average in 2,5 times).

In order to model natural changes in user behavior (that do not correspond to anomaly) for each user components of vector $\mathbf{g}(s_i)$ were randomly changed (with probability 0,25). In this case the value H_t increased in 2 times, and then decreased to ordinary level.

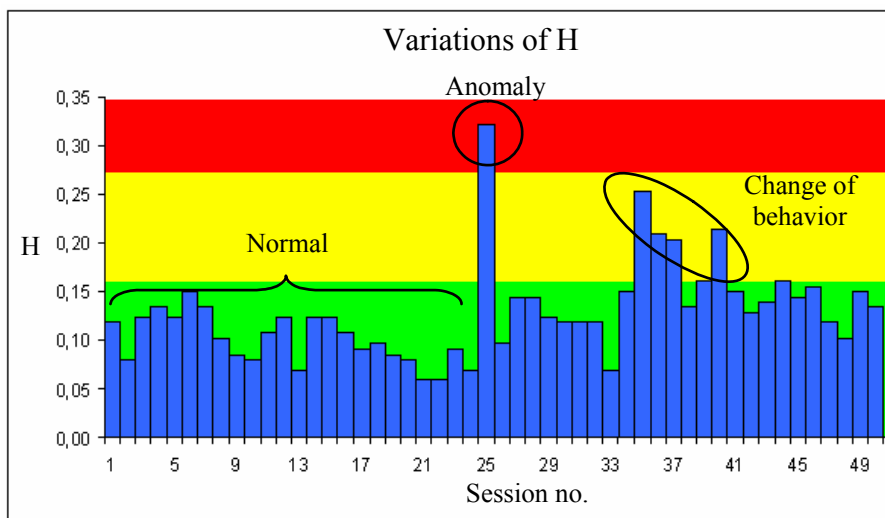


Fig. 4. Variations of value H_t depending on user behavior

The variations of value H_t for typical user depending on user behaviour are depicted in Fig. 4. Sessions #1-24 correspond to normal behavior, and H_t lie in the range (0; 0,15). When data of another user were inserted (session #25) H_t increased up to 0,32. When we modelled natural change of user behaviour (sessions #35-50) H_t increased up to 0,25, and then decreased to ordinary level.

Therefore, experimental results showed the possibility of proposed complex neural network model to distinguish with confidence normal and abnormal (anomalous) user behavior.

Conclusions

In this paper we proposed a complex model of user behaviour in distributed system. In order to adequately describe different features of user behaviour the model consists of three components: on-line model considers dynamics of user behaviour by predicting user actions, off-line model is based on the analysis of statistical parameters, and change detection module that is intended for detection of trends in user's activity. In contrast to existing methods the proposed model enables complex analysis of user behaviour both during its work (in real-time) and after user's work completion (in off-line mode). The use of neural network provides intelligent approach to analysis and generalization of data acquired during user activity. In order to demonstrate efficiency of complex model different experiments were run on real data obtained during user work on resources of Space Research Institute of NASU-NSAU. The results of experiments showed applicability of the proposed approach.

Acknowledgement

This research is supported by the Science and Technology Center in Ukraine (STCU) and the National Academy of Sciences of Ukraine (NASU) within project "GRID technologies for environmental monitoring using satellite data", no. 3872.; and INTAS-CNES-NSAU project "Data Fusion Grid Infrastructure", Ref. No 06-100024-9154.

Bibliography

- [Adams and Lloyd, 2002] Adams C., Lloyd S. "Understanding PKI: Concepts, Standards, and Deployment Considerations", 2nd ed. Addison-Wesley, 2002.
- [Annis, *et al.*, 2002] Annis J., Zhao Y., et al., "Applying Chimera Virtual Data Concepts to Cluster Finding in the Sloan Sky Survey," Technical Report GriPhyN-2002-05, 2002.
- [Cannady and Mahaffey, 1998] Cannady J., Mahaffey J. "The Application of Artificial Neural Networks to Misuse Detection: Initial Results", In Proc. of the 1998 National Information Systems Security Conf. (NISSC'98), Arlington, VA, 1998.

-
- [Dowell and Ramstedt, 1990] Dowell C., Ramstedt P. "The ComputerWatch data reduction tool", In Proc. 13th National Computer Security Conf., 1990, pp. 99–108.
- [Foster, *et al.*, 1998] Foster I., Kesselman C., Tsudik G., Tuecke S. "A Security Architecture for Computational Grids", In ACM Conf. on Computers and Security, 1998, pp. 83-91.
- [Fusco, 2006] Fusco L. "Earth Science GRID on Demand", Presented on CEOS WGISS-21 GRID Task Team meeting, Budapest, Hungary, May 2006.
- [Fusco, *et al.*, 2003] Fusco L., Goncalves P., Linford J., Fulcoli M., Terracina A., D'Acunzo G. "Putting Earth-Observation on the Grid", ESA Bulletin, 2003, 114, pp. 86-91.
- [Gorodetski, *et al.*, 2001] Gorodetski V., Karsaev O., Khabalov A., Kotenko I., Popyack L., Skormin V. "Agent-based model of Computer Network Security System: A Case Study", In Proc. of the Int. Workshop 'Mathematical Methods, Models and Architectures for Computer Network Security', Lecture Notes in Computer Science, 2052, Springer Verlag, 2001, pp. 39-50.
- [Haykin, 1999] Haykin S. "Neural Networks: a comprehensive foundation", Upper Saddle River, New Jersey: Prentice Hall, 1999, 842 p.
- [Holtman, 2001] Holtman K., "CMS Requirements for the Grid", In Proc. of the Int. Conf. on Computing in High Energy and Nuclear Physics (CHEP2001), 2001.
- [IETF] IETF, Public-Key Infrastructure (pkix) Charter.
- [Javitz and Valdes, 1991] Javitz H., Valdes A. "The SRI IDES statistical anomaly detector", In: Proc. IEEE Symp. on Research in Security and Privacy, 1991, pp. 316–326.
- [Kussul and Sokolov, 2003] Kussul N., Sokolov A. "Adaptive anomaly detection of user behaviour using Markov chains with variable order", J. of Automation and Control, Vol. 4, pp. 83-88. (in Russian)
- [Peltier, *et al.*, 2002] Peltier S.T., *et al.* "The Telescience Portal for Advanced Tomography Applications", J. of Parallel and Distributed Computing: Computational Grid, 2002, 63(5), pp. 539-550.
- [Reznik, *et al.*, 1999] Reznik A., Kussul N., Sokolov A. "Identification of user activity using neural networks", J. of Cybernetics and Computer Science, 1999, No. 123, pp. 70–79. (in Russian)
- [Ryan, *et al.*, 1998] Ryan J., Lin M-J., Miiikulainen R. "Intrusion Detection with Neural Networks", Advances in Neural Information Processing Systems, Cambridge, MA: MIT Press, 1998, pp. 943–949.
- [Shelestov, *et al.*, 2006] Shelestov A.Yu., Kussul N.N., Skakun S.V. "Grid Technologies in Monitoring Systems Based on Satellite Data", J. of Automation and Information Science, 2006, Vol. 38, Issue 3, pp. 69-80.
- [Skakun, *et al.*, 2005] Skakun S.V., Kussul N.N., Lobunets A.G. "Implementation of the Neural Network Model of Users of Computer Systems on the Basis of Agent Technology", J. of Automation and Information Sciences, 2005, Vol. 37, Issue 4, pp. 11-18.
- [Tulloch, 2003] Tulloch M. "Microsoft Encyclopedia of Security", Redmond, Washington: Microsoft Press, 2003, 414 p.

Authors' Information

Andrii Yu. Shelestov – PhD, Senior Researcher, Department of Space Information Technologies and Systems, Space Research Institute of NASU-NSAU, Glushkov Ave 40, Kyiv-187, 03650 Ukraine, e-mail: inform@ikd.kiev.ua.

Serhiy V. Skakun – PhD, Research Assistant, Department of Space Information Technologies and Systems, Space Research Institute of NASU-NSAU, Glushkov Ave 40, Kyiv-187, 03650 Ukraine, e-mail: inform@ikd.kiev.ua.

Olga M. Kussul – BSc, Physics and Technology Institute, National Technical University "KPI", Peremoga Ave 37, Kyiv-056, 03056 Ukraine, e-mail: olgakussul@gmail.com.

ОЦЕНКА БИОРАЗНООБРАЗИЯ С ИСПОЛЬЗОВАНИЕМ НЕЙРОННЫХ СЕТЕЙ

Андрей Шелестов, Екатерина Насуро, Сергей Скакун

Резюме: Рассматривается роль обработки данных дистанционного зондирования Земли в задаче оценки биоразнообразия. Предлагается модификация метода оценки биоразнообразия за счет применения нейросетевой классификации подстилающей поверхности.

Ключевые слова: биоразнообразие, данные ДЗЗ, нейронные сети.

ACM Classification Keywords: F.1.1 Models of Computation (F.4.1) - Self-modifying machines (e.g., neural networks)

Введение

Основными признаками здоровья, благополучия и жизнеспособности биосферы Земли является количество и биоразнообразие экосистем, составляющих биосферу. Экосистемы обеспечивают ресурсы, необходимые для поддержания жизни на Земле и влияют на жизнедеятельность и процветание всех биологических видов, включая и человека [1].

Управление изменениями в биоразнообразии и понимание процессов, происходящих в экосистемах Земли возможно при наличии оперативной, достоверной и систематизированной информации. Научное изучение экосистем, своевременное информирование мировой общественности, консолидация усилий по защите окружающей среды позволит правильно интерпретировать знания об изменениях, происходящих в окружающей среде в масштабах всего земного шара, изучить роли экосистем и их взаимодействие.

"Система систем" GEOSS

Решением проблем глобализации и объединения усилий по сохранению здоровья биосферы Земли может стать "система систем" глобального наблюдения земли GEOSS. Она ориентирована на объединение существующих и будущих систем наблюдения Земли, развивающихся во многих странах мира. Глобальность системы позволяет привлечь наибольшее количество стран и наблюдать процессы, происходящие в масштабах, значительно больших, чем масштаб отдельно взятой страны. С другой стороны такая система позволит объединить усилия на более низком уровне, привлекая к решению задач функционирующие региональные системы [2].

Разработка высококачественной, современной и всесторонней глобальной системы мониторинга биоразнообразия, обеспечит данными все заинтересованные стороны. Среди которых: участники многосторонних соглашений по окружающей среде, правительства разных стран, органы по управлению природными ресурсами, научные деятели, структуры по экологии, сельскому хозяйству, здравоохранению, чрезвычайным ситуациям и гидрометео службы.

Среди ближайших приоритетов в создании GEOSS называют создание условий и координирование непрерывного наблюдения на всех уровнях, начиная от глобального мониторинга поверхности земли, до локальных измерений и обеспечение к полученным данным свободного доступа для любого пользователя системы. Применять данные дистанционного наблюдения Земли для описания, картографирования и мониторинга глобальных защищенных зон, включая зоны Всемирного наследия, природные области, области большого культурного, геологического и археологического значения. Использовать данные наблюдения Земли и другие пространственные данные для обеспечения оконтуривания и корректировки границ защищенных зон.

Роль спутников ДЗЗ в программах по сохранению биоразнообразия

Необходимость сохранить хрупкие экосистемы и их биоразнообразие усложнена их распределением по всему земному шару, отдаленностью и национальными границами. В таких случаях важную роль играют спутники дистанционного зондирования Земли. Такие данные предоставляют различную информацию, необходимую для принятия природоохранных решений. Проекты по дистанционному зондированию Земли используют возможности мультиспектральных оптических и радарных приборов спутников для анализа изменений поверхности земли, определения массы растительности, распознавания типов растительности и обнаружения и картографирования отдельных биологических видов. Эти данные полезны для составления графиков, диаграмм, таблиц основных факторов окружающей среды. Анализ такой информации позволит сформировать элементы и параметры биоразнообразия: развитие растений, их среду обитания и климат.

Биоразнообразие

Биоразнообразие – это разнообразие видов жизни на Земле. По мнению авторов [2] биоразнообразие имеет структуру, состоящую из трех основных уровней организации и трех различных аспектов разнообразия. Уровни биоразнообразия делятся на генетический, видовой уровень и уровень экосистем. Аспектами разнообразия называются композиция, структура и функции (см. рис.1).

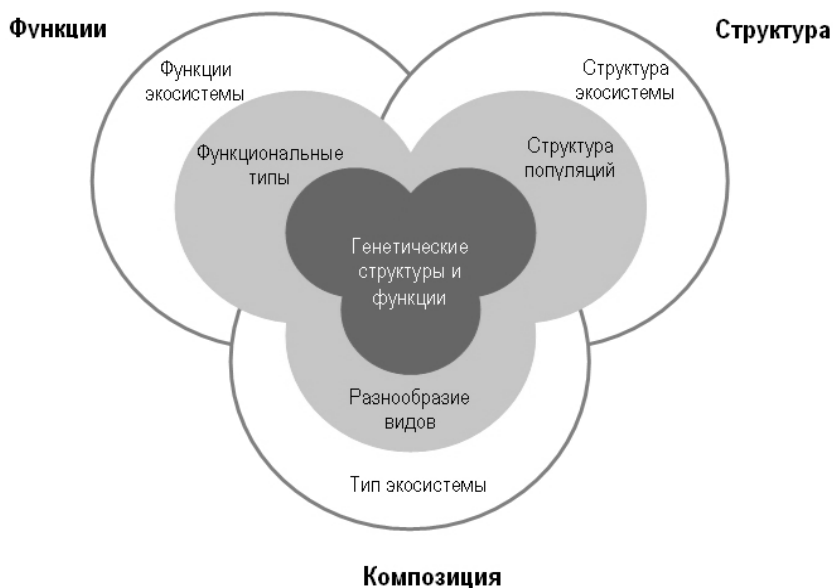


Рисунок 1. Концептуальная диаграмма, отображающая структурные уровни биоразнообразия

Биоразнообразие необходимо для сохранения жизни на земле в целом. Объединение различных сервисов охраны окружающей среды позволит осознать важность биоразнообразия, степень и последствия его истощения, построить глобальную, локальные и национальные системы, принимать правильные решения по сохранению биоразнообразия и успешно использовать результаты охранных мер.

Методика вычисления биоразнообразия

Успешное сбережение биоразнообразия невозможно без разработки четкой научной концепции его оценивания и мониторинга. Перспективы изучения и картографирования биоразнообразия тесно связаны с использованием материалов ДЗЗ. Важным заданием во время реализации этого подхода является определение показателей биоразнообразия, которые могут быть получены с использованием данных ДЗЗ [3].

Биоразнообразие часто ассоциируется с количеством видов, как в конкретной группе, так и по всему ландшафту в целом. Видовое разнообразие состоит из двух компонентов. Во-первых, видовое богатство, т.е. общее число существующих видов. Во-вторых, равномерность распределения видов, которая базируется на определении положения вида в структуре доминирования на основе мер значимости [4].

При оценивании видового разнообразия широко используется индекс Шеннона, который учитывает не только богатство видов, но и равномерность распределения этих видов по любому из признаков [5]

$$B = \sum_{i=1}^N P_i \log_2 P_i \quad (1)$$

где N – количество видов, P_i – значимость вида i .

Использование данных дистанционного зондирования Земли и современных цифровых технологий обработки аэрокосмических изображений предоставляет ряд преимуществ в процессе расчета индекса видового разнообразия по Шеннону. Например, позволяет значительно расширить территорию исследований, отображать реальную информацию о пространственном распределении групп растений и предоставить возможность постоянного обновления этих данных. Кроме этого, представление данных в цифровом виде позволит обеспечить их эффективную обработку и визуализацию.

Для построения методики оценки биоразнообразия используется экосистемный подход, предложенный А.Н.Киселевым [6]. Его главный принцип состоит в том, что оценка биоразнообразия может опираться на любые интегральные характеристики, которые обеспечивают "представительство" всех биоконпонентов. Основной идеей экосистемного подхода является следующее: элементы живой и неживой природы находятся в тесной связи и взаимодействии друг-с другом, то есть создают эко- и геосистемы различных уровней и длительности. В упрощенном виде принцип взаимодействия элементов этих систем выглядит следующим образом: условия обитания влияют на процессы заселения территории видами растительности, что в свою очередь обеспечивает условия для развития определенных видов животных.

Для связи биоразнообразия и ряда влияющих на него факторов используется модель, основанная на нечеткой логике [7]. Она позволяет описать классические динамические связи с неопределенными или слабоопределенными параметрами и статистические зависимости с неизвестными законами распределения, а так же формализовать экспертные представления про характер процессов в исследуемых экосистемах.

Если известны количественные значения $F_i, i = 1..n$ факторов, которые влияют на оценку биоразнообразия B , где n – общее количество факторов, для каждого из m классов ландшафта земного покрытия можно определить потенциально возможное биоразнообразие $B_0^{(j)}, j = 1..n$, которое при использовании индекса Шеннона (1) определяется количеством характерных биологических видов данного класса $N_0^{(j)}$ по формуле:

$$B_0^{(j)} = \log_2 N_0^{(j)} \quad (2)$$

Базой для оценки биоразнообразия поверхности служит классификация покрытий ландшафта. Ее можно получить по данным ДЗЗ и наземным измерениям. Именно по классификации выполняется оценка потенциального биоразнообразия для каждого класса земной поверхности с использованием формулы (2). Если возможности космической съемки позволяют выполнить видовую классификацию растительности, то есть возможность получить непосредственную оценку видимого биоразнообразия территории исследования, установить зависимость между видимым и скрытым биоразнообразием и вычислить полное биоразнообразие для каждого участка ландшафта.

Факторы, которые влияют на биоразнообразие и могут быть определены по материалам ДЗЗ, приведены в таблице 1. Сейчас в мире существует большое количество центров, которые предоставляют не только аэрокосмические снимки, но и информационные продукты более высоких уровней. Примером может

служить международный научный продукт ДЗЗ EOS, в рамках которого предоставляется достаточно широкий спектр информационных продуктов, которые могут использоваться при оценке биоразнообразия.

Таблица 1. Факторы, которые влияют на биоразнообразиие

Фактор	Показатель	Диапазон значений	Оптимальное значение
Рельеф местности	Среднесуточное облучение солнечным излучением	0 – 300 Вт/м ²	максимум
Вода	Влажность почв	0 – 100 %	максимум
Тепло	Дневная и ночная температура	253 – 323 К	293 К
Осадки	Среднегодовое количество осадков	0 – 200 мм	максимум
Растительность	Вегетационный индекс	-1 .. +1	максимум
Жизненная активность	Относительное количество усвоенного растительностью излучения	-1 .. +1	максимум
Жизненная эффективность	Чистая продуктивность растительности	0 – 600 кг карбона/м ² в год	максимум

В настоящее время существует множество программ, направленных на создание продуктов по классификации земной поверхности, которые отличаются системами классификации, покрытием, используемыми данными, технологиями построения и методами обработки. Среди них следует выделить европейский проект Corine [8] и американский проект National Land Cover Data (NLCD) [9], которые в качестве основы используют данные спутника Landsat, а также MODIS Land Cover and Land-Cover Change [10], который основан на данных прибора MODIS спутника Terra и использует глобальную систему классификации программы International Geosphere-Biosphere Program (IGBP) [11]. Стоит отметить, что для территории Украины доступны лишь данные по глобальной классификации земной поверхности с разрешением не менее 1 км. Кроме того, самые свежие данные классификации поверхности для Украины датируются 2000 годом. Поэтому актуальной является задача создания более эффективных методов классификации покрытий ландшафта.

Классификация данных дистанционного зондирования Земли

Применение нейронных сетей для классификации данных ДЗЗ

В настоящее время при выборе подхода к классификации земного покрова по данным дистанционного зондирования земли все чаще отдается предпочтение нейросетевым методам. Нейронные сети обеспечивают адаптивный и устойчивый подход к анализу и обобщению информации, который не требует априорных знаний о статистическом распределении данных. Существующее разнообразие различных нейросетевых архитектур, а также методов обучения нейронных сетей, позволяют выбрать наиболее оптимальную для поставленной задачи конфигурацию, найти компромисс между скоростью, размерностью сети и эффективностью работы. В то же время такое разнообразие возможностей приводит к необходимости проводить большое количество экспериментов для получения достоверных и наиболее подходящих к задаче вариантов решения.

Наши исследования, проведенные в работах [12, 13, 14] раскрывают результаты применения различных нейросетевых архитектур, методов обучения нейронных сетей первого и второго порядка, использования сетей на основе радиальных базисных функций и сетей на основе теории адаптивного резонанса для обработки данных спутника Landsat. Полученные результаты показывают, что для разных классов наибольшее значение верно классифицированных образов достигалось для разных методов, что объясняется сложностью топологии исходных данных.

Для нейронных сетей с обучением с учителем, необходимо определить набор эталонных данных. Эталоном могут быть результаты замеров на местности, приведенные к нужному виду. Кроме этого, данные о спектральных свойствах объектов, можно получить в лаборатории или из библиотек спектральных кривых. В таком случае данные необходимо приводить к виду, информации, получаемой из космоса, в зависимости от числа каналов прибора и диапазонов измерения, а так же с учетом функций отклика прибора [15]. Кроме перечисленных методов доступны данные международных программ, например, проект Corine по классификации наземных объектов [8].

Сравнение результатов работы различных нейросетевых парадигм

Для сравнительного анализа работы различных нейронных сетей с обучением с учителем было использовано изображение, полученное мультиспектральным прибором ETM+ (Enhanced Thematic Mapper Plus) спутника Landsat-7 [16]. Прибор предоставляет данные с пространственным разрешением 30м. Для эксперимента были взяты каналы 1-5 и 7. Размерность входных векторов составляла 6 элементов. На выходе был представлен вектор из 8 элементов – по количеству определяемых классов. Для определения эталонных данных использовалась информация проекта Corine по классификации земной поверхности. На тестовом снимке были определены следующие 8 классов (в скобках указан код классификации согласно проекту Corine): лиственный лес (311), хвойный лес (312), смешанный лес (313), пашня (211), пастбище (231), вода (51), искусственные материалы (1), открытая местность с разреженной растительностью (33) [12, 13].

В процессе работы были рассмотрены и протестированы следующие нейросетевые парадигмы: многослойный персептрон, обученный методом обратного распространения ошибки, EDBD, квази-Ньютоновским методом и методом сопряженных градиентов; нейронная сеть на основе радиальных базисных функций; нейронная сеть, основанная на теории адаптивного резонанса.

Выполненные эксперименты были направлены на определение оптимальной архитектуры нейронной сети, предназначенной для классификации мультиспектральных спутниковых снимков; анализ методов обучения нейронных сетей и их влияние на результаты работы сети; анализ разделимости данных и применимости нейронных сетей для решения задач такого класса. Сравнение результатов работы различных нейросетевых архитектур приведено в таблице 2.

Таблица 2. Сравнение результатов работы различных нейросетевых архитектур

Классы		Архитектура сети					
		Обратного распространения ошибки	EDBD	Квази-Ньютоновский	Сопряженных градиентов	Радиальная базисная функция	Нейронные сети ARTMAP
лиственный лес	1	99,6%	98,9%	100,0%	99,9%	99,9%	99,3%
хвойный лес	2	94,6%	85,7%	98,3%	96,3%	88,8%	80,2%
смешанный лес	3	53,4%	67,4%	0,5%	35,2%	57,7%	68,7%
пашня	4	91,2%	89,6%	87,2%	92,7%	97,1%	82,0%
пастбище	5	90,5%	91,4%	0,0%	0,0%	77,8%	75,8%
вода	6	97,8%	97,7%	0,0%	0,0%	92,6%	96,2%
искусственные материалы	7	81,4%	83,6%	58,8%	67,2%	70,8%	77,8%
открытая местность	8	89,6%	96,5%	0,0%	0,0%	79,4%	97,3%
Всего		87,9%	88,9%	58,5%	64,35%	85,6%	83,7%

Анализируя результаты проведенных экспериментов видно, что для разных классов наилучший результат достигался разными методами. Это можно объяснить сложностью входных данных и сложными свойствами разделимости. В среднем по всей выборке данных наилучшие результаты дает многослойный перцептрон, обученный методом EDBD. Кроме того, обучение этим методом занимает значительно меньше времени, по сравнению с другими методами [14].

Выводы и направления дальнейшей деятельности

В статье предложен метод оценки биоразнообразия на основе использования результатов нейросетевой классификации земного покрова, который позволяет выполнить видовую классификацию участка ландшафта по данным дистанционного зондирования.

Сложное разбиение поверхности земли на классы и подклассы влечет за собой значительное усложнение архитектуры данных. При увеличении количества классов усложняется архитектура нейронной сети, что приводит к замедлению процесса обучения и снижению точности классификации. Наиболее оптимальной с точки зрения разбиения данных на классы является нейронная сеть, которая разбивает предложенные образцы на два класса. Для улучшения классификации спутниковых снимков целесообразно проанализировать применение модульных нейронных сетей со сложной архитектурой. Разные модули такой сети, используя различные методы и архитектуры, проводят вначале грубое, а потом более точное разбиение данных на классы.

В докладе будет рассмотрена разработка структуры модульной нейронной сети, каждый модуль которой проводит разбиение на два класса, а результаты такой классификации являются входными данными для следующей нейронной сети, которая проведет более точную классификацию. Это даст возможность оценить преимущества такого подхода и применимость его для решения задач классификации спутниковых данных.

Список литературы

- [1] Earth Observation Handbook. Committee on Earth Observation Satellites. www.eohandbook.com
- [2] Global Earth Observation System of Systems GEOSS. Group on Earth Observation. ESA Publication Divisions, Netherlands, February 2005.
- [3] Багатоспектральні методи дистанційного зондування Землі в задачах природокористування / За ред. В.І. Лялько та М.О. Попова.- Київ: Наукова думка, 2006.- 360 с.
- [4] Одум Ю. Экология.- Т.2.- Пер. с англ.– М.: Мир, 1986.– 376 с.
- [5] Протасов А.А. Биоразнообразие и его оценка: Концептуальная диверсикоология.- К.: Институт гидробиологии, 2002.- 108 с.
- [6] Киселев А.Н. Оценка и картографирование биологического разнообразия (на примере Приморья) // Геоботаническое картографирование, 2000.- Т.3.- С.3–15.
- [7] Козлова А.А., Станкевич С.А. Оценка и картографирование биоразнообразия Северо-Причерноморского региона Украины на основании многоспектральных космоснимков и геоинформационных технологий // Сборник тезисов Шестой украинской конференции по космическим исследованиям.- Евпатория: ИКИ НАН Украины, 2006.- С.117.
- [8] NASA Landsat 7, <http://landsat.gsfc.nasa.gov>
- [9] National Land Cover Data (NLCD) [<http://www.epa.gov/mrlc/nlcd.html>]
- [10] Strahler Alan, et al. MODIS Land Cover and Land-Cover Change, Algorithm Theoretical Basis Document (ATBD), Version 5.0, 1999. – 72 p.
- [11] Belward A.S., Estes J.E., Kline K.D. The IGBP-DIS Global 1-km Land-Cover Data Set DISCover: A Project Overview// Photogram. Eng. Remote Sens. — 1999. — 65. — P. 1013-1020.
- [12] N. Kussul, S. Skakun, O. Kussul. Comparative Analysis of Neural Networks and Statistical Approaches to Remote Sensing Image Classification// International Scientific Journal of "Computing". – 2006. – Vol. 5, Issue 2. – P 93-99.

- [13] Куссуль Н.Н., Скакун С.В., Куссуль О.М. Анализ нейросетевых и статистических подходов к классификации изображений спутника Landsat-7 // Интелектуальні системи прийняття рішень і прикладні аспекти інформаційних технологій: Матеріали науково-практичної конференції. Том 1. – Херсон: Видавництво Херсонського морського інституту, 2006. – С. 117-120.
- [14] Скакун С.В., Насуро Е.В., Лавренюк А.Н., Куссуль О.М. Анализ применимости нейронных сетей для классификации спутниковых данных // Проблемы управления и информатики. — 2007. — №2. — С. 105-117.
- [15] Лавренюк А.Н., Яровая Е.В., Скакун С.В. Нейросетевая классификация спектральных характеристик наземных объектов // Кибернетика и вычислительная техника, выпуск 148, 2005. – С. 21-28.
- [16] Landsat-7 Science Data User's Handbook. – http://ftpwww.gsfc.nasa.gov/las/handbook/handbook_toc.html

Информация об авторах

Шелестов Андрей Юрьевич - Институт космических исследований НАН и НКА Украины, кандидат технических наук, старший научный сотрудник, докторант; проспект Академика Глушкова 40, 03680 Киев, Украина; e-mail: inform@ikd.kiev.ua.

Насуро Екатерина Валериевна - Национальный авиационный университет, аспирантка; проспект Комарова 1, 03058 Киев, Украина; e-mail: mail_aspir@ukr.net.

Скакун Сергей Васильевич - Институт космических исследований НАН и НКА Украины, кандидат технических наук, научный сотрудник; проспект Академика Глушкова 40, 03680 Киев, Украина; e-mail: inform@ikd.kiev.ua.

FORMING OF LEARNING SET FOR NEURAL NETWORKS IN PROBLEMS OF LOSLESS DATA COMPRESSION

Yuriy Ivaskiv, Victor Levchenko

Abstract: questions of forming of learning sets for artificial neural networks in problems of losless data compression are considered. Methods of construction and use of learning sets are studied. The way of forming of learning set during training an artificial neural network on the data stream is offered.

Keywords: neural network; modelling of data sources; learning set.

ACM Classification Keywords: C.1.3 Other Architecture Styles – Neural Nets. D.4.8 Performance – Modelling and predictions. E.4 Coding and Information Theory – Data compaction and compression.

Introduction

In problems of data compression wide application was found with artificial neural networks (ANN). The neural methods of losless data compression have got wide spread occurrence [Jiang, 1999]. Applicability of lossy compression methods is limited to a class of multimedia data, and neural realization of such methods is specific to each concrete type of data that limits an opportunity of their use [Cottrell, 1987; Verma, 1999]. Methods of losless data compression differ that at their use, as a rule, special restrictions on structure and type of data are not imposed that expands opportunities of their practical application for the decision of various classes of problems.

The most of modern methods of losless data compression are entropic, that is such in which the length of a code necessary for representation of an element of data, contacts probability of occurrence of an element in a stream of compressed data [Ватолин, 2002]. Division of a problem of compression into two problems is standard: *coding* and *statistical modelling* of a data source [Rissanen, 1981]. Now various methods and algorithms of coding

redundancy of which practically does not differ from theoretically minimal possible are developed and applied [Кричевский, 1989]. However, a problem of increase of efficiency of statistical modelling is insufficiently studied and an opportunity of increase of compression ratio in particular is connected with its decision.

The existing statistical models used for data compression, differ complexity and heterogeneity of structure, and also that they are sensitive to type of modelled data that complicates their practical realization and application [Bell, 1989; Witten, 1989; Ker-Chang Chang, 1993]. At the same time the models which are based on ANN, differ simplicity and uniformity of structure, an opportunity of use of the same algorithms of adjustment and functioning in various models, and also some other useful properties [Хайкин, 2006]. Therefore, prospects of development of statistical models are connected with use as modelling system at lossless compression of the artificial neural networks.

Target of Settings

The problem of statistical modelling of data sources during realization of lossless compression consists in definition of probabilities of occurrence of elements of compressed data on an input of coding system. One of methods of reception of estimations of these probabilities is forecasting occurrence of elements of data in an input stream. At use ANN such forecasting is carried out similarly to forecasting of time series [Ежов, 1998]. Distinction consists basically in principle of forming of learning set for ANN. At forecasting time series there are accessible all values that enables training of ANN with the given level of an error. At statistical modelling data come in an input of modelling system in the consecutive portions that demands permanent recustomizing of model during its functioning with the purpose of adaptation to changes of data [Schmidhuber, 1996]. In this connection there is a necessity of development and research of special methods and algorithms of formation of learning set for neural models which would consider features of its functioning on the given data stream.

Main Results

Description of neural model.

For carrying out of researches the two-layer back propagation neural network with adjustable quantity of neurons in an input layer (if the inputs of ANN are considered as the first layer such ANN is considered to be three-layer) was used. As one of possible variants of realization of transferring function the hyperbolic tangent – $\tanh x$ – was used. Other parameters ANN had following values:

- the maximal absolute error of training per one example: 0,001;
- a steepness of transfer function: 1,0 – for an input layer, 0,1 – for a target layer;
- factor of learning speed: 0,1 – for all layers.

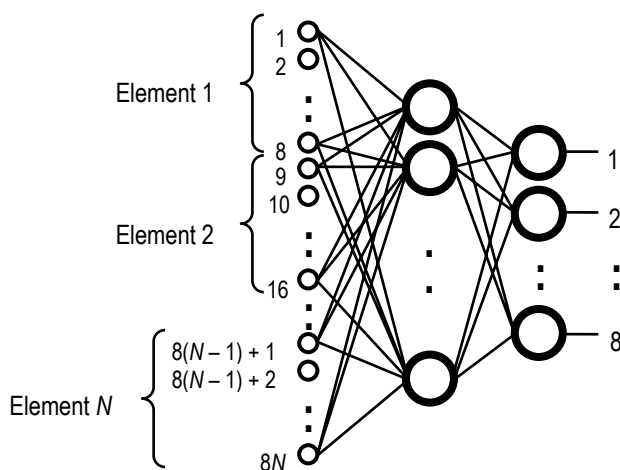


Fig. 1

Proceeding from convenience of realization neural models on the existing computer-based means dimension of 1 byte has been accepted for elements of the input data stream that is not basic restriction of parameters of considered model. Accordingly, the quantity ANN inputs was defined as $8N$, where N – quantity of data elements on which forecasting (that is context length), and the quantity of outputs was fixed and was equal 8 (fig. 1). Coding of inputs and target signals of a network has great value for process of ANN learning because finally it defines finish of learning. Zero bits of inputs and target signals of a network were coded by value “-0,5”, and set

bits – "+0,5" as use of such values allows to consider process of learning completed only at concurrence of signs of the received and required values of a network without achievement of a concrete numerical result.

Influence of context length.

During ANN functioning it was learned of forecasting of the values appearing in an input data stream on contexts with length from 1 up to 15 elements. Here start value answers minimally possible length of a context and the range of its change is defined by known methods of contextual modelling [Ватолин, 2002]. It is established, that quality of

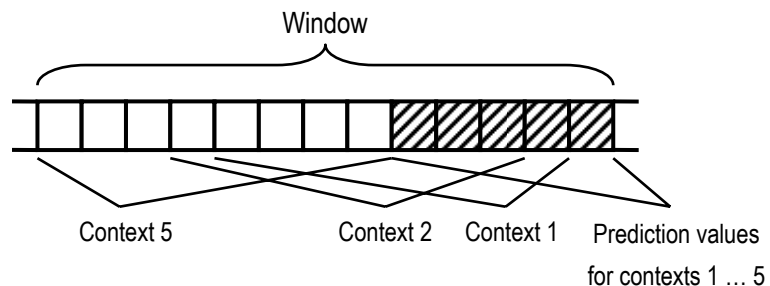


Fig. 2

forecasting in a greater measure depends on length of a context for the data, described by strongly pronounced internal structure (texts in natural languages and texts of programs) and in a smaller measure – for data of type "an analog signal" (the digitized images and a sound). For text data dependence of optimum length of a context on volume of data is revealed that is coordinated with experimental data. At use of contexts of optimum length the quantity of correct forecasts has increased on the average on (25 – 45) %.

Window learning mode.

As a possible way of increase of quantity of correct forecasts window mode of ANN learning is considered at which learning is carried out not on one context, but on set of the overlapped consecutive contexts that are directly previous of prediction value (fig. 2). Researches were spent for windows with a width from 2 up to 50 contexts. The range of change width of a window is chosen experimentally. The basic problem of formation of learning set for a window mode of learning was elimination of inconsistent contexts [Тарасенко, 2001],

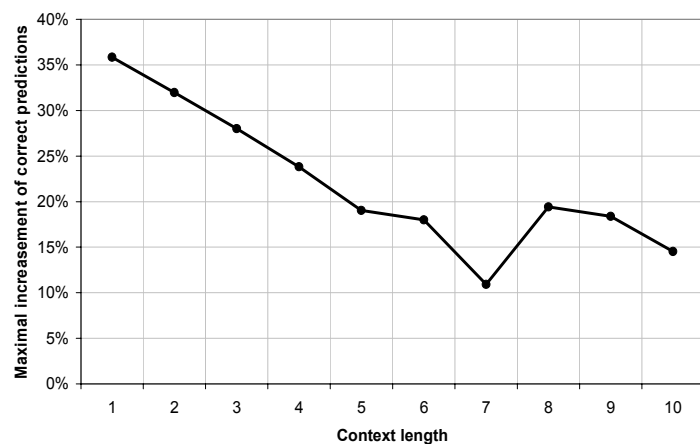


Fig. 3

the presence of which cycles process of learning and decreases speed of functioning of model sharply. Elimination of inconsistent contexts was carried out as follows: if there were two or more conterminous contexts with various prediction values in a window the nearest to prediction value was left and all previous were removed.

As a result of test of neural model in a window mode of learning it has been established, that the increase in width of a window allows raising quantity of correct forecasts (on 40 % for text data), however it reduces speed of ANN learning, and speed of model as a whole. Besides the increase in length of a context leads to reduction of number of the correct forecasts connected with increase of width of a window (fig. 3). Thus the window mode of learning is favourable only at use of contexts of small length (1 ... 4 elements) under condition of elimination of inconsistent contexts.

Selective learning

The analysis of change of quality of forecasting during model functioning has shown, that neural model functioning on all set of contexts with any width of a window is characterized by initial splash in quantity of correct forecasts and its following decrease to some level. It is supposed, that such behaviour of quality of forecasting shows that only some subset from all contexts used for ANN learning represents laws, characteristic for all set of data, and the others contexts have casual character and lead to decrease in quality of training process.

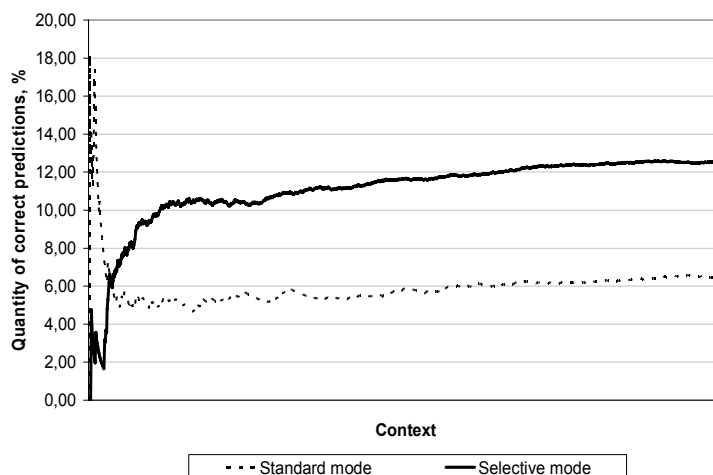


Fig. 4

For check of the given assumption learning of a network has been led in two modes: 1) standard (on all set of contexts); 2) selective (on set of contexts after which forecasting was successful). As a result it was revealed, that at selective learning change of quantity of correct forecasts was absolutely different unlike a standard mode (fig. 4); thus the stable increase in quality of forecasting during all process of ANN training and functioning is observed.

Essential lack of a selective learning mode is dependence of quality of forecasting process on an initial condition of model, which is on a choice of initial values of ANN weight factors. The reason of such dependence is that at a selective mode training of a network begins with the first context for which the correct forecast has been executed, thus the probability of such forecast depends on initial values of ANN weight factors. In connection with that casual values are usually appropriated to ANN weight factors, there is a probability of their unsuccessful choice, as entails sharp decrease in speed, and quality of ANN learning also.

The account of influence of an initial condition of model on forecasting process is offered to be carried out by one of the next two ways. The first way assumes to carry out learning of a network on contexts for which the forecast has been executed with the error that is not exceeding given. Thus, in view of that modelled data are represented by binary codes, for an estimation of an error it is expedient to use Hemming distance. At a choice of Hemming distance it is important to consider, that when admissible error of forecasting increases, the quantity of the casual contexts increases also and it makes worse quality of forecasting, and when admissible error of forecasting decreases, the probability of the duly beginning of ANN learning decreases also.

The second way assumes application of the mixed mode of learning at which in the beginning of ANN functioning it is trained in a usual mode, which is on all meeting contexts ("start" of model), and after a time learning becomes selective. The choice of length of data on which "start" of model is carried out depends on type of data, and also their structure, defined by a solved problem and demands special studying.

Conclusion

The choice of a context length for learning set should be carried out in view of type of modelled data.

The choice of a context length without taking into account type of modelled data does not allow getting optimum quantity of correct forecasts that can be compensated by increase in width of a learning window.

Forming of training set from small contexts (1 ... 4 elements), and also with use of windows in width from 7 up to 10 contexts is preferable to increase in number of correct forecasts at the solving of problems of lossless compression for various types of data.

The greatest number of correct forecasts in the data stream can be achieved at the use of a selective mode of learning with preliminary "start" of model or with use of forecasts within the limits of Hemming distance no more than 2 bits.

Bibliography

- [Jiang, 1999] Jiang J. Image compression with neural networks – A survey. *Signal Processing: Image Commun.* 1999. – V. 14. – P. 737 – 760.
- [Cottrell, 1987] Cottrell G. V., Munro P., Zipser D. Image compression by back propagation: An example of extentional programming // *Proc. 9th Annual Confer., Cognitive Soc.* – 1987. – P. 461 – 473.
- [Verma, 1999] B. Verma, M. Blumenstein, S. Kulkarni. A New Compression Technique Using an Artificial Neural Network // *Journal of Intelligent Systems.* – 1999. – Vol. 9. – P. 39 – 53.
- [Ватолин, 2002] Ватолин Д., Ратушняк А., Смирнов М., Юкин В. Методы сжатия данных. Устройство архиваторов, сжатие изображений и видео. – М.: ДИАЛОГ-МИФИ, 2002. – 384 с.
- [Rissanen, 1981] Jorma Rissanen, Glen G. Langdon. Universal modelling and coding. // *IEEE transaction on information theory.* 1981. V. IT – 27. № 1. – P. 12 – 33.
- [Кричевский, 1989] Кричевский Р. Е. Сжатие и поиск информации. – М.: Радио и связь, 1989. – 168 с.
- [Bell, 1989] Bell T. C., Moffat A. M. A note on the DMC data compression scheme. // *Comput. J.* – 1989. – V. 32. – P.16–20.
- [Witten, 1989] Bell T. C., Witten I. H., Cleary J. G. Modeling for text compression // *ACM Computer Surveys.* – 1989. V. 24. – P. 555 – 591.
- [Ker-Chang Chang, 1993] H. Ker-Chang Chang, Shing-Hong Chen. A New Locally Adaptive Data Compression Scheme Using Multilist Structure // *Comput. J.* – 1993. V. 36.
- [Хайкин, 2006] Хайкин С. Нейронные сети: полный курс, 2-е издание.: Пер. с англ. – М.: Издательский дом "Вильямс", 2006. – 1104 с.
- [Ежов, 1998] Ежов А. А., Шумский С. А. Нейрокомпьютинг и его применение в экономике и бизнесе (серия "Учебники экономико-аналитического института МИФИ" под ред. проф. В. В. Харитонова). – М.: МИФИ, 1998. – 244 с.
- [Schmidhuber, 1996] Schmidhuber J. Sequential neural text compression // *IEEE transaction on Neural Networks.* – 1996. – V. 7 (1). – P. 142 – 146.
- [Тарасенко, 2001] Тарасенко Р. А., Крисилов В. А. Предварительная оценка качества обучающей выборки для нейронных сетей в задачах прогнозирования временных рядов. // *Тр. Одес. политехн. ун-та.* – Одесса, 2001. – Вып. 1. С. 25 – 28.

Authors' Information

Yu. L. Ivaskiv - National Aviation University, Doctor of Science; professor; P.O.Box: 03058, Kosmonavta Komarova Ave., bl. 1, Kiev, Ukraine; e-mail: post@nau.edu.ua.

V. V. Levchenko - Industrially-Economic College of National Aviation University, post-graduate student; P.O.Box: 03126, Geroev Sevastopolya St., bl. 44 - A, 53, Kiev, Ukraine; e-mail: vv@voliacable.com

СИСТЕМА ОПТИМИЗАЦИИ НА ОСНОВЕ ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ, ГЕНЕТИЧЕСКОГО АЛГОРИТМА И НЕЙРОСЕТЕВЫХ МЕТАМОДЕЛЕЙ

Павел Афонин

Аннотация: В работе рассматривается задача оптимизации на основе имитационных моделей сложных систем. Основное внимание уделено оптимизации с использованием генетических алгоритмов. Отмечается необходимость построения и использования метамоделей в процессе поиска решения. Представлена архитектура системы оптимизации на основе имитационного моделирования, генетического алгоритма и нейросетевых метамоделей.

Ключевые слова: Имитационное моделирование, оптимизация, генетический алгоритм, мета модель, нейронная сеть.

Введение

На сегодняшний день большое число задач оптимизации сложных систем из различных областей науки и техники не могут быть решены с помощью аналитического моделирования. Во многих случаях единственным методом решения является метод имитационного моделирования [Лоу и Кельтон, 2004]. Задача оптимизации на основе имитационного моделирования формулируется следующим образом: необходимо найти значения входных переменных (факторов), оптимизирующих основной выходной показатель системы (отклик). При этом предполагается, что функция отклика не может быть вычислена аналитически, но может быть рассчитана с помощью имитационного моделирования, т.е. с помощью проведения имитационного эксперимента с моделью сложной системы.

Оптимизация на основе имитационного моделирования заключается в совместном использовании имитационной модели (ИМ) сложной системы и алгоритма оптимизации. С помощью ИМ рассчитывается значения отклика для различных комбинаций значений факторов, которые предлагает алгоритм оптимизации. Поисковый алгоритм оптимизации, в свою очередь, используя значения отклика, пытается улучшить решение. Следует подчеркнуть, что одной из основных проблем оптимизации на основе имитационного моделирования является проблема длительности прогонов ИМ на языках имитационного моделирования. Кроме того, существует проблема предварительной сходимости алгоритма оптимизации.

В данной работе описаны методы и алгоритмы оптимизации на основе имитационного моделирования. Представлены наиболее популярные в настоящее время программные пакеты оптимизации. Для решения проблемы сокращения времени на поиск решения отмечается необходимость построения и использования метамоделей. Основное внимание уделено метамоделям на основе нейронных сетей (НС). Описаны методы построения и использования нейросетевых метамоделей совместно с алгоритмом оптимизации. Проведен анализ факторов, которые могут оказывать влияние на качество получаемого решения, исходя из чего, предложена архитектура системы оптимизации на основе имитационного моделирования.

Методы и средства оптимизации на основе имитационного моделирования

Наиболее важными свойствами, которыми должны обладать системы оптимизации на основе имитационного моделирования, являются: качество получаемых решений (хотя его трудно определить на практике, поскольку истинный оптимум не известен) и время поиска, необходимое для их получения. Время поиска решения зависит от количества конфигураций системы, которые необходимо моделировать, а также от времени прогона одной конфигурации системы.

Необходимо отметить, что применение в качестве алгоритма оптимизации точных математических методов оптимизации, обеспечивающих нахождение оптимального решения, не всегда целесообразно, поскольку имитационная модель является копией реальной системы с некоторой степенью точности. При этом использование точных методов, как правило, требует значительных вычислительных затрат, что во многих случаях является критичным или вовсе недопустимо. Поэтому в большинстве случаев в качестве алгоритма поисковой оптимизации лучше использовать методы, которые не обязательно гарантируют достижение точного оптимума, а находят близкие к оптимальным решения и при этом обеспечивают быструю поисковую сходимости алгоритма.

На сегодняшний день существует несколько программных пакетов оптимизации имитационного моделирования, которые используют средства имитационного моделирования совместно с различными методами поиска решений. Далее представлена краткая информация об этих пакетах с указанием их поставщиков, программ моделирования, которые они поддерживают, а также используемых процедур поиска решений:

- AutoStat AutoSimulations (Inc AutoMod), программы моделирования: AutoSched, процедуры поиска: эволюционные стратегии;
- OptQuest Optimization (Technologies, Inc.), программы моделирования: Arena, Micro Saint, QUEST процедуры поиска: поиск с рассеиванием, поиск с запрещением, нейронные сети;
- OPTIMIZ (Visual Thinking International Ltd.), программы моделирования: SIMUL8, процедуры поиска: эволюционные стратегии, нейронные сети;
- SimRunner2 (PROMODEL Corp.), программы моделирования: MedModel, ProModel, ServiceModel процедуры поиска: эволюционные стратегии, генетические алгоритмы;
- WITNESS Optimizer (Lanner Group, Inc.), программы моделирования: WITNESS, процедуры поиска: имитация отжига, поиск с запрещением.

Как можно видеть, в большинстве пакетов оптимизации в качестве процедур поиска решений используются эволюционные стратегии и генетические алгоритмы (ГА) [Goldberg, 1989]. На сегодняшний день эти методы хорошо себя зарекомендовали как универсальные алгоритмы глобального поиска, которые позволяют находить квазиоптимальные решения за приемлемое время.

Использование метамоделей

Как было сказано выше, одной из основных проблем оптимизации на основе имитационного моделирования является проблема длительности прогонов имитационной модели. Прогон ИМ для некоторых больших систем на языках имитационного моделирования может достигать нескольких часов [Лоу и Кельтон, 2004]. Реализация серии прогонов таких систем, необходимой для алгоритма оптимизации, за разумное время не представляется возможным.

Наиболее применимым на практике способом решения данной проблемы является использование метамоделей. Метамоделю принято называть приближенную математическую модель, полученную в результате экспериментов с имитационной моделью с целью замещения последней при оптимизации. Основными методами построения метамоделей являются регрессионные модели и искусственные нейронные сети (НС), к которым в последнее время проявляется большой интерес, благодаря их мощной аппроксимирующей способности. Одной из первых фундаментальных работ по построению нейросетевых метамоделей в имитационном моделировании является работа [Piergeval, 1992].

Существуют различные схемы реализации поиска решения с помощью эволюционных вычислений и нейросетевых метамоделей. В частности, в работе [Jin and other, 2002] описаны две схемы организации эволюционных вычислений с помощью метамоделей: схема на основе контроля особей и схема на основе контроля поколений. В первом случае задается число контролируемых особей, для которых на каждом поколении работы алгоритма осуществляется расчет функции пригодности (ФП) с помощью прогона ИМ.

Для остальных особей в популяции расчет ФП реализуется с использованием метамоделей. Во втором случае задается число контролируемых поколений, для которых реализуется расчет ФП всех особей популяции с использованием ИМ. Для неконтролируемых поколений расчет ФП осуществляется с помощью метамоделей. При этом непосредственно стратегия поиска решения и обучение НС реализуются с помощью ковариационной матрицы адаптации.

В работах [Rasheed and Hirsh, 2000] и [Persson and other, 2007] представлена схема поиска решения основанная на стратегии информированности операторов ГА. Данная стратегия заключается в генерации большого числа потомков в операторах скрещивания и мутации генетического алгоритма и последующего расчета их ФП с помощью метамоделей. Приводятся результаты экспериментов с моделями сложных систем, которые показывают эффективность выбранной стратегии.

В работе [Kern and other, 2006] описано использование различных эволюционных стратегий совместно с локальными метамоделями. Проведены исследования для большого числа классических тестовых функций. Представлены результаты экспериментальных данных, которые говорят об эффективности тех или иных метамоделей.

Архитектура системы оптимизации на основе имитационного моделирования, генетического алгоритма и нейросетевых метамоделей

В результате анализа описанных выше схем поиска решений с использованием метамоделей было выявлено, что каждая из них обладает определенными достоинствами и недостатками. С целью комбинации достоинств методов, предлагается объединить три описанные схемы, а именно, схему на основе контроля особей, схему на основе контроля поколений и схему, основанную на стратегии информированности операторов в одну систему оптимизации. Определим основные параметры системы оптимизации: 1. число контролируемых особей; 2. число контролируемых поколений; 3. число потомков в операторах скрещивания и мутации; 4. способ генерации потомков в операторах скрещивания и мутации; 5. число прогонов ИМ для расчета ФП одной особи; 6. Способ подбора обучающих данных для НС.

На сегодняшний день не существует общей стратегии по выбору данных параметров при оптимизации конкретных систем. Также следует подчеркнуть, что для организации эффективного поиска решения и корректной сходимости оптимизационного алгоритма должна иметься возможность подстройки параметров системы оптимизации в процессе поиска решения. С этой целью предлагается использовать экспертную систему (ЭС). Рассмотрим систему оптимизации на основе имитационного моделирования, которая разработана опираясь на предложенную ранее архитектуру системы [Афонин, 2006], путем введения экспертной системы и необходимых связей (рис.1).



Рис.1. Система оптимизации на основе имитационного моделирования

Опишем связи блоков системы. Связь 1 реализует поиск решения на основе генетического алгоритма и ИМ. Связь 2 реализует поиск решения на основе метамоделей. Связь 3 осуществляет формирование обучающих примеров для нейронной сети. Они представляют собой преобразование входов ИМ в выходы, которые были получены с помощью прогона предложенных решений блоком оптимизации на ИМ. Связь 4 реализует передачу обучающих примеров для обучения нейронной сети. В общем случае не все примеры из БД могут использоваться для обучения НС на определенном этапе поиска решения. Связь 5 определяет необходимое число прогонов ИМ для одной особи, которую предлагает ГА. Связь 6 реализует стратегию поиска на основе метамоделей (контролируемые особи, контролируемые поколения). Связь 7 определяет, какие примеры могут входить в БД – в общем случае входят все примеры, т.е. которые описывают все прогоны ИМ. Связь 8 Определяет какие примеры будут использоваться для обучения НС на текущем этапе поиска решения. Также в задачи ЭС входит определение числа потомков и способа генерации потомков в операторах скрещивания и мутации на каждом этапе поиска решения.

Заключение

В данной работе представлена архитектура системы оптимизации на основе имитационного моделирования, генетического алгоритма и нейросетевых метамоделей. Выявлены факторы, которые могут оказывать влияние на качество получаемого решения. Отмечена необходимость использования экспертной системы для организации правильного взаимодействия компонентов системы и определения параметров оптимизации в процессе решения. Представлены современные методы, алгоритмы и программные пакеты оптимизации на основе имитационного моделирования. В дальнейшем предполагается детальная проработка и исследование отдельных взаимосвязей и компонентов системы.

Библиография

- [Лоу и Кельтон, 2004] А.Лоу, Д.Кельтон. Имитационное моделирование: Пер. с англ. (3-е изд.) – СПб.: BHV, 2004.
- [Goldberg, 1989] D.E.Goldberg. Genetic Algorithms in Search, Optimization and Machine Learning. – New York: Addison-Wesley Publishing Company, Inc., 1989.
- [Pierreval, 1992] H.Pierreval. Training a neural network by simulation for dispatching problems, Proceedings of the Third Rensselaer International Conference on Computer Integrated Engineering, New York, 1992, 332–336.
- [Jin and other, 2002] Y.Jin, M.Olhofer; B.Sendhoof. A framework for evolutionary optimization with approximate fitness functions. IEEE Transactions on Evolutionary Computation, Vol.6, No.5, pp.481-494, 2002.
- [Rasheed and Hirsh, 2000] K.Rasheed, H.Hirsh, 2000. Informed operators: Speeding up genetic-algorithm-based design optimization using reduced models. In Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2000), Las Vegas, Morgan Kaufmann, pp. 628–635, 2000.
- [Persson and other, 2007] A.Persson, H.Grimm, A.Ng. Metamodel-assisted global search using a probing technique. In Proceedings of The IAENG International Conference on Artificial Intelligence and Applications (ICAIA'07), 2007.
- [Kern and other, 2006] S.Kern, N.Hansen, P.Koumoutsakos. Local meta-models for optimization using evolution strategies. In proceedings of PPSN, 2006.
- [Афонин, 2006] П.В.Афонин. Гибридные системы интеллектуального имитационного моделирования // Новости искусственного интеллекта. – 2006. – №1.
-

Информация об авторе

Павел Афонин - к.т.н., доцент кафедры «Компьютерные системы автоматизации производства» МГТУ им. Баумана; 2-я Бауманская, д.5, Москва, Россия; e-mail: pavlafon@yandex.ru

ДИАГНОСТИРОВАНИЕ НА НЕЙРОННЫХ СЕТЯХ В СИСТЕМЕ ГОМЕОПАТ

Лариса Катеринич, Александр Провотар

Аннотация: Предлагается нейронная сеть для решения задачи диагностирования в системе ГОМЕОПАТ, а также алгоритмы ее обучения.

Ключевые слова: искусственный интеллект, нейронные сети, обучение нейронных сетей.

Введение

Как известно, в результате изучения структуры головного мозга и механизмов его работы были созданы новые вычислительные модели, а именно, искусственные нейронные сети (НС). Задачи автоматизации процессов управления на базе исследований в области НС являются актуальными и по сей день. НС позволяют решать прикладные задачи, такие как, распознавание образов, моделирование, быстрое преобразование информации (параллельные вычислительные процессы), идентификация, управление, создание экспертных систем [Терехов, 2002, Барский, 2004]

Теоретически НС может решать широкий круг задач в конкретной предметной области (так как она является прототипом модели головного мозга человека), но создать единую универсальную НС для конкретной предметной области в настоящее время практически не возможно, так как не существует единого алгоритма построения (функционирования) НС. В настоящее время для решения конкретной группы задач из заданной предметной области используются НС определенной структуры и с определенными алгоритмами обучения.

Как известно, каждый нейрон имеет ряд качественных характеристик, таких как состояние (возбужденное или заторможенное), входные и выходные связи. Однонаправленные входные связи, соединенные с выходами других нейронов называют синапсами, а выходная связь данного нейрона, с которой сигнал (возбуждения или торможения) поступает на синапсы следующих нейронов, называют аксоном [Терехов, 2002]. Общий вид нейрона приведен на рис. 1.

По сути, работа каждого нейрона является относительно простой. Как правило, на вход нейрона поступает набор сигналов $X = [x_1, x_2, x_3, \dots, x_n]$. Каждый из сигналов может быть выходом другого нейрона или источника. Каждый входной сигнал умножается на соответствующий угловой коэффициент $W = [w_1, w_2, w_3, \dots, w_n]$. Он соответствует силе синапса биологического нейрона.

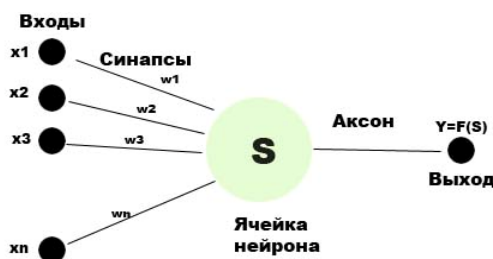


Рис. 1.

Произведение $w_i * x_i$ суммируются и поступают на суммирующий элемент. Для инициализации сети вход x_0 ($x_0 = +1$) и весовые коэффициенты синаптических связей w_0 вводятся специально. Состояние нейрона в текущий момент времени определяется, как взвешенная сумма его входов:

$$S = \sum_{i=1}^n x_i w_i + x_0 w_0$$

Выходом нейрона есть выход его состояния $Y = F(S)$. Функция F является функцией активации. Она является монотонной, непрерывно дифференцируемой на интервале либо $(-1,1)$, либо $(0,+1)$ [Терехов, 2002].

В многослойных нейронных сетях (МНС) выходы базовых элементов каждого слоя поступают на входы всех базовых элементов следующего слоя. Функция активации $F(S)$ выбирается одинаковой для всех нейронов сети. В [1] МНС определена в такой символической форме $N_{n_0, n_1, \dots, n_k}^k$, где k – число слоев в сети, n_0 – число входов сети; $n_i (i = 1, k - 1)$ – число базовых элементов в i -х промежуточных слоях, n_k – число базовых элементов в выходном k -ом слое и одновременно число выходов q_1, \dots, q_{n_k} МНС. Промежуточный a -й слой имеет n_a нейронов. Связи между базовыми элементами в слое отсутствуют. Выходы базовых элементов a -го слоя поступают на входы нейронов только следующего $(a+1)$ -го слоя. Выход для любого нейрона определен в виде:

$$q_i^{(a)} = f\left(\sum_{j=1}^{n_{a-1}} w_{i,j}^{(a)} q_j^{(a-1)} + w_{i,0}^{(a)} q_0^{(a-1)}\right) = f(s_i^{(a)}).$$

Специализированные НС в системе ГОМЕОПАТ

Разработкой математических методов решения медицинских задач ученые занимаются уже много лет. Эффективность подобных математических методов можно проследить по ряду медицинских диагностических систем, которые были разработаны в последнее время. Общей чертой подобных систем является зависимость от конкретных методов обработки групповых данных, слабо применимых к единичным объектам, а также особенностям медицинской информации [Горбань, 1998, Осовский, 2002].

Нейронные сети (НС) являются удобным инструментом для представления информационных моделей. В общем случае, сеть принимает некоторый входной сигнал из внешнего мира и пропускает его через себя с преобразованиями в каждом нейроне. Таким образом, в процессе прохождения сигнала по связям сети происходит его обработка, результатом которой является определенный выходной сигнал.

Для проектирования нейронной сети в системе ГОМЕОПАТ [Проватар, 2000] была выбрана наиболее распространенная структура нейронных сетей - многослойная. Эта структура подразумевает что каждый нейрон произвольного слоя связан со всеми выходами (аксонами) нейронов предыдущего слоя или со всеми входами НС в случае первого слоя. Другими словами сеть имеет следующую структуру слоев: входной, промежуточный (скрытый) и выходной. Такие нейронные сети также называют полносвязными [Барский, 2004].

Для решения задачи диагностирования в системе ГОМЕОПАТ используется НС следующей архитектуры (рис.2):

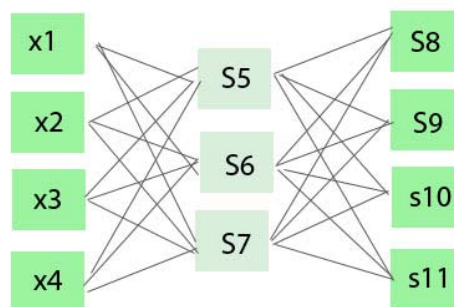


Рис. 2

Задача обучения МНС в классическом виде может быть представлена так. Пусть задана некоторая последовательность x^* входных данных. Необходимо найти такое решение x , при котором можно

классифицировать вновь представленные входные данные. Критерий $R(x, x^*)$ определяет качество решения. Множество решений x определяется выбором алгоритма настройки весовых коэффициентов $w_i^{(a)}$. При такой постановке задачи процесс обучения сводится к получению наилучшего решения из множества возможных. Другими словами, обучение МНС – это процесс накопление информации x^* и параллельно процесс выбора решения x .

Алгоритм обучения

НС системы ГОМЕОПАТ использует алгоритм обратного распространения, суть которого заключается в распространении сигналов ошибки от выходов НС к ее входам в направлении, обратном прямому распространению сигналов в обычном режиме работы (режим распознавания). Другими словами, используя технологии последовательной настройки нейронов, начиная с последнего (выходного слоя) и заканчивая настройкой элементов первого слоя. Обучение НС может быть проведено необходимое число раз. Для обучения используется так называемое δ -правило, которое заключается в реализации стратегии обучения «с учителем». Если через y^* обозначить необходимый выход нейрона, где y – реальный выход, то ошибка обучения вычисляется по следующей формуле $\delta = y^* - y$ в алгоритме градиентного спуска с весовым коэффициентом

$$w_i(k+1) = w_i(k) - \gamma \delta x_i, \quad \gamma > 0,$$

где γ – коэффициент «усиления алгоритма», x_i – i -й вход синаптической связи нейрона.

Алгоритм обучения НС для решения задач диагностики в системе ГОМЕОПАТ состоит из следующей последовательности шагов:

В контексте предметной области строится вектор входных сигналов: $x = \langle x_1, x_2, \dots, x_n \rangle$, где x_1, x_2, \dots, x_n симптомы пациента.

Вектор соответствующий правильным значением (необходимым) $y^* = \langle y_1, y_2, \dots, y_n \rangle$ формируется экспертом предметной области.

Выполняется алгоритм прямого распространения сигнала x через сеть. В результате выполнения алгоритма определяются весовые суммы S_{jn} и активаторы для каждой ячейки.

Выполняется алгоритм обратного распространения сигнала через ячейки выходного и промежуточного слоя. Происходит вычисление ошибок δ_0 для выходной ячейки и δ_i для скрытых ячеек.

Выполняется обновление весов нейронов в сети, где $W_{i,j}$ – весовой коэффициент синаптической связи, соединяющий i -й нейрон слоя $n-1$ с j -нейроном слоя n ; S_{jn} – взвешенная сумма выходных сигналов слоя n (аргумент активационной функции); X_{ij} – i -й вход нейрона j слоя n ; n_i – число нейронов в слое.

Обучение НС состоит в представлении обучающих примеров из небольшой группы желаемых действий. Это достигается путем выполнения алгоритма обратного распространения с учетом желаемого результата и действительного результата.

Сеть работает в двух режимах: в режиме обучения и в режиме распознавания. В режиме обучения производится формирование так называемых логических цепочек. В режиме распознавания НС по конкретным входным сигналам с высокой степенью достоверности определяет, какие действия предпринять.

Обучение нейронной сети проходят на ограниченном количестве примеров, затем ей позволяют самостоятельно генерировать поведение в других ситуациях. Способность генерировать правильную

реакцию на различные симптомы, не входящие в набор обучающих, является ключевым фактором при создании НС.

Данные для тестирования представляют собой несколько сценариев с набором действий. В результате сеть должна рассчитывать реакцию на входы и выполнять действие, которое будет похожим на обучающие сценарии.

Для обучения сети использовались примеры, приведенные в таблице 1.

Таблица 1. Примеры обучения нейронной сети

Симптом 1	Симптом 2	Симптом 3	Симптом 4	Препарат
2	1	0	1	P1
1	1	0	1	P2
0	0	1	0	P3
1	0	1	1	P4
0	0	1	1	P5

Чтобы протестировать НС, сети должны быть представлены новые примеры. Это позволяет определить, как сеть будет реагировать на сценарии, о которых ей ничего не известно. Подобные тесты позволяют узнать насколько качественно НС может реагировать на непредвиденные ситуации и выполнять нужные действия. Чтобы протестировать НС, ей были представлены новые примеры из таблицы 2.

Таблица 2. Примеры ввода данных для тестирования нейронной сети

Симптом 1	Симптом 2	Симптом 3	Симптом 4	Препарат
1	1	0	1	P1
1	2	0	1	P2
0	0	1	1	P3
1	1	0	1	P4
0	1	0	1	P5

Заключение

Отличительным свойством системы ГОМЕОПАТ является то, что она следует стратегии конструирования (и затем проверки дифференциального диагноза), применяемой человеком-клиницистом. Такая модель диагностики включает двухступенчатую процедуру, которая сводится к тому, что сначала выдвигаются гипотезы заболевания на основании введенных данных о пациенте (процесс „снизу-вверх”), а затем производится их оценка с помощью дополнительных симптомов, которые должны быть присущи предполагаемым заболеваниям (процесс „сверху-вниз”). При этом, поиск необходимой информации для формирования образа заболевания осуществляется в базе знаний системы с помощью приведенных выше специализированных нейронных сетей. В перспективе для решения задач идентификации и распознавания предполагается использование нечетких нейронных сетей различной архитектуры.

Библиография

- [Терехов, 2002] В.А. Терехов, Д.В. Ефимов, И.Ю. Тюкин. Нейросетевые системы управления. – Москва: «Радиотехника», 2002. – 467с.
- [Барский, 2004] А.Б. Барский. Нейронные сети: распознавание, управление, принятие решений. – Москва: «Финансы и статистика», 2004. – 398с.

-
- [Горбань, 1998] А.Н.Горбань, В.Л.Дунин-Барковский, А.Н.Кирдин и др. Нейроинформатика. – Новосибирск: «Наука», 1998.
- [Осовский, 2002] С. Осовский. Нейронные сети для обработки информации. – М: «Финансы и статистика», 2002. – 365с.
- [Провотар, 2000] Провотар А.И., Дудка Т.Н., Гошко Б.М. Применение метода резолюций на семантических сетях //Проблемы программирования. –2000. –№ 1–2.
-

Информация об авторах

Катеринич Лариса Александровна – ассистент факультета кибернетики Киевского национального университета имени Тараса Шевченко, katerinich@rambler.ru

Провотар Александр Иванович – доктор физико-математических наук, профессор, заведующий кафедрой информационных систем факультета кибернетики Киевского национального университета имени Тараса Шевченко, aprowata@unicyb.kiev.ua

NEURAL NETWORK APPROACH PREDICTION OF THE TYPE OF A COURSE OF MULTIPLE SCLEROSIS BY THE CLINICAL CHARACTERISTICS OF ITS DEBUT

Inna Panchenko, Tetyana Shatovska

Keywords: *Multilayer Perceptron, Multiple Sclerosis.*

ACM Classification Keywords: *F.2.1 Numerical Algorithms and Problems*

Introduction

Methods of Intellectual Data Analysis are not obtained commonly in medicine because of small data warehouses with different length cases, combination of various data types (integer attributes, qualitative ones, scales of measurements, intervals, and binary parameters), the missed data abundant. A proportion of object quantity and researched parameters does not conform to principles of statistics that interpreted of the received results from its importance inconvenient.

The aim of our research is the Neural Network (NN) prognosis of the illness debut affiliation with the types of Multiple Sclerosis (MS) development.

Analysis of Multiple Sclerosis

Expert has sorted sampling under consideration into three MS classes: Primary Progressive Type (PPT), Relapsed-Remitted Type (RT) and Secondary Progressive Type (SPT). The last two ones intersect – some cases of RT transform to SPT, which means for researcher their only probabilistic separability [Lang, 1997].

Data warehouse includes temporal and clinical parameters of MS debut of 52 PPT patients, 46 RT patients and 68 SPT cases.

Characteristics of the MS debut are:

1. Age of debut.
2. Gender.
3. Risk factors before a debut: stress, infections, insolation, somatic diseases, operations, cranial traumas, spinal traumas, labors, abortions, others.
4. Harbingers: before debut, long-term ones.
5. Temps of development of a debut: hours, within a week, weeks, and months.

6. Character of a debut: monosyndromic, oligosyndromic, polysyndromic.
7. Syndromes of debut: retrobulbar [optic] neuritis, pyramidal, cerebellar, sensitive, brainstem, sphincters disorders, others.
8. Onset of symptoms order: simultaneous, fractional.
9. Duration of a debut: under 1 month, under 3 months, above 3 months.
10. Debut way out: under 2 wks, 2 wks – 1 month, above 1 month.
11. Severity of debut: mild, average, serious.

Age of the beginning of disease was integer value (yrs), data in all sub-groups of parameters 3-11 represented as binary.

According to the probabilistic separability of the MS development types the using of Probabilistic NN seems to be native, but it showed higher baseline error then Multilayer Perceptron (the average value for verification group was 0.3681), because of that the NN Multilayer Perceptron (MLP) was chosen [Grisby, 1994].

According to goal of our research, supervised learned two-layer architecture error back propagation algorithm MLP selected. The number of nodes in its hidden layer was determined as the half-number of input (forty clinical characteristics) and output (three types of MS) variables [Fayyad,1996].

Due to indexes of entropy ranging on the importance, only following sub-group parameters, which were influence to belonging to the MS type, used by the MLP for training:

1. Acute (under week) rate of MS development.
2. A short (under 1-month) debut.
3. Prolonged (above 1 month) debut regression.
4. Sex of patients.
5. Slowed (months) rate of MS development.
6. Rapid (under 2 wks) debut regression.
7. Moderate (2 wks – 1 month) debut regression.
8. Debut of average (less than 3 months) duration.
9. Stroke-like rate of debut development.
10. Presence of infectious background as a risk factor.
11. Fractional onset of attributes of MS manifestation.

Researched sample results (table 1):

Table 1 - Results of training, testing and verification of MLP

	Training sample			Test sample			Verifying sample		
	RT	PPT	SPT	RT	PPT	SPT	RT	PPT	SPT
Total of patients	24	23	36	9	14	18	13	14	14
Correctly	10	20	31	4	11	15	6	13	12
Incorrectly	14	3	5	5	3	3	7	1	2
Unknown	0	0	0	0	0	0	0	0	0
RT	10	0	3	4	0	1	6	0	2
PPT	2	20	2	1	11	2	0	13	0
SPT	12	3	31	4	3	15	7	1	12

According to the table, difference of expert estimation and MLP training result was minimal at the analysis of progressive MS types. Therefore, from 23 training sample PPT cases, 20 examples have recognized correctly,

but three ones MLP has rated to SPT. From 14 cases of the test sample three ones also have classified as SPT and the result of verifying sample were 13 PPT cases and one SPT case.

When classifying the examples, which the expert has considered as SPT type, from 36 cases of the training sample two ones have defined as PPT, and three ones have to as RT. In the test sample of 14 cases, two ones have recognized as PPT, and one as RT. In the verifying sample of 14 cases, two ones defined to RT.

At RT, MLP could not distinguish the most of examples correctly. At the MLP training 12 of 24 patients have been referred to SPT, and two ones to PPT; in test sample 4 from 9 patients were classified as SPT, and one as PPT; the verifying sample has been divided as 6 patients with RT and 7 with SPT [Han, 2000].

It is significant, that divergences MLP with the expert in classification PPT concerned only SPT, i.e. essentially another progressive type of MS development whereas discrepancies between SPT and RT were opposite, and that is indirect evidence of an ambiguous belonging to the class - type of development.

Perhaps, classification of RT as SPT testifies to probable future transformation of RT to SPT.

Conclusion

Thus, based on application of a NN MLP, classification attributes of the MS debut were established. These attributes allow regarding progressive types of development, especially PPT, as stable clinical forms with good chance for forecasting whereas Relapsed-Remitted Type characterize instability of class belonging which means for the majority of patients ability of transformation to Secondary Progressive Type of disease.

Current problem seems to us to require further research.

Bibliography

[Lang, 1997] Lang E.W., Pitts L.H., Damron S.L., Rutledge R. Outcome after severe head injury: Analysis of prediction based upon comparison of neural network versus logistic regression analysis. *Neurological Research* 1997; 19. P. 274- 280.

[Grisby, 1994] Grisby J., Kooker R., Hershberger J. Simulated neural network to predict outcomes, cost and length of stay among orthopaedic rehabilitation patients. *Arch. Phys. Med. Rehabil.* 1994. - Vol. 75. P. 1077-1082.

[Fayyad,1996] Fayyad U. M., G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy. *Advances in Knowledge Discovery and Data Mining*. AAAI/MIT Press, 1996

[Han, 2000]J. Han and M. Kamber. *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 2000.

Authors' Information

Inna Panchenko – Kharkiv National University of Radioelectronics, Department of Artificial Intellect, P.O.Box: Lenin avenue 14, Ukraine, 61166; e-mail: inna_pan_o@mail.ru

Tetyana Shatovska– Kharkiv National University of Radioelectronics, Computer Science department, P.O.Box: Kharkiv-116, Lenin av. 14, Ukraine; e-mail: tanita_uk@mail.ru

FINDING AN APPROPRIATE PARTITION ON THE SET OF ARGUMENTS OF A PARTIAL BOOLEAN FUNCTION TO BE DECOMPOSED

Arkadij Zakrevskij

Abstract: A hard combinatorial problem is investigated – the two-block decomposition, generally non-disjunctive, of partial Boolean functions. The key task is regarded: finding such a weak partition on the set of arguments, at which the considered function can be decomposed. Solving that task is essentially speeded up by the way of preliminary discovering traces of the sought-for partition. Boolean and ternary vectors and matrices are used, with efficient parallel combinatorial operations over them.

Keywords: Partial Boolean function, non-disjunctive decomposition, appropriate partition, combinatorial search, traces of the solution.

Introduction

A task is considered concerning the problem of functional decomposition of Boolean functions, which was set originally in papers [Povarov, 1954], [Ashenurst, 1959], [Curtis, 1962]. Let's formulate it as follows.

Suppose a Boolean function $f(\mathbf{x}) = f(x_1, x_2, \dots, x_n)$ is given. It is required to decompose it, having presented as the following composition of two functions g and h of smaller number of variables:

$$f(\mathbf{x}) = g(h(\mathbf{u}, \mathbf{w}), \mathbf{w}, \mathbf{v}).$$

The sets of arguments given in the vector form are connected by that with the relations $\mathbf{x} = \mathbf{u} \cup \mathbf{w} \cup \mathbf{v}$, $\mathbf{u} \cup \mathbf{w} = \mathbf{u} \cup \mathbf{v} = \mathbf{w} \cup \mathbf{v} = \emptyset$, and the couple of sets \mathbf{u} and \mathbf{v} determines a weak partition on set \mathbf{x} , designated as \mathbf{u}/\mathbf{v} . It is accepted to name such composition sequential two-block. It is illustrated by Figure 1 for the case, when $\mathbf{u} = (x_1, x_2, x_3)$, $\mathbf{w} = (x_4, x_5)$ and $\mathbf{v} = (x_6, x_7)$.

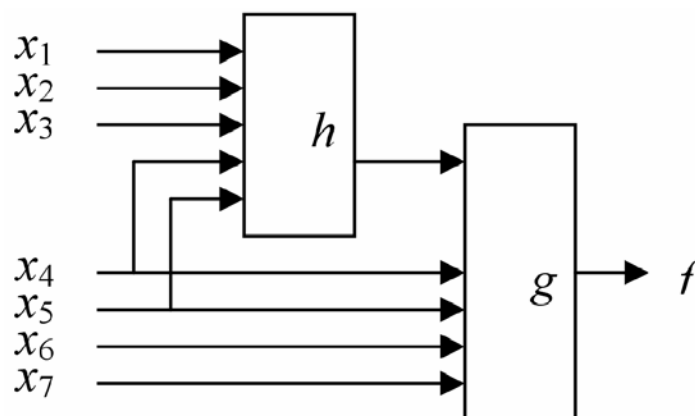


Fig. 1. Example of composition

To solve the formulated task it is necessary, first of all, to find such a weak partition \mathbf{u}/\mathbf{v} , at which the variables of set \mathbf{u} enter in number of arguments of function h only, and variables of \mathbf{v} – only in number of arguments of

function g . The conditions $|u| > 1$ and $|v| > 0$ should be fulfilled also - otherwise the composition will appear trivial (exists always). Let's name this partition *appropriate*, and the function $f(\mathbf{x})$ – separable, or decomposable at the given partition.

The finding of appropriate partition is a difficult task, for which solution an effective combinatorial algorithm was offered [Zakrevskij, 2006], in application to completely specified Boolean functions. This task becomes even more complicated, when the function $f(\mathbf{x})$ appears to be partial, being defined not on all sets of values of variables from set \mathbf{x} . Just this case is considered below.

It was shown [Zakrevskij, 2006], that the probability of decomposability of a random completely defined Boolean function fast tends to zero with growth of number of variables n , so already at $n > 9$ such a function, most likely, is not decomposable. In case of partial functions this probability arises with growth of uncertainty, however even in this case it remains small enough.

Taking into account the given remark, let's assume, that it is known beforehand, that the considered function $f(\mathbf{x})$ is separable, being obtained as a result of composition $g(h(\mathbf{u}, \mathbf{w}), \mathbf{w}, \mathbf{v})$ of some two Boolean functions g and h on a weak partition u/v on the set of arguments \mathbf{x} . It is required to detect (to recognize) this partition, after which the obtaining of functions g and h is not a difficult task.

A method of checking a partial Boolean function for decomposability at some given weak partition was offered in [Zakrevskij, 2007]. In the case, when this appropriate partition is not known a priori, it is possible to organize its search, sorting out different weak partitions and checking the function on decomposability at them. However, such way is rather labor-consuming, as the number of different weak partitions on the set of variables is approximated from above by value 3^n , fast growing with increase of number of variables n .

In the present paper the method of search for appropriate partition u/v by its traces is suggested, which sufficiently cuts down the number of analyzed partitions. Originally it was designed for completely specified Boolean functions [Zakrevskij, 2006], but here it is extended on the case of partial Boolean functions.

Basic operations in Boolean space

The parallelism of component-wise operations above long Boolean vectors is used in the offered method of search for appropriate partitions, and that essentially accelerates the fulfilled calculations.

It is possible to consider Boolean functions $f(\mathbf{x}) = f(x_1, x_2, \dots, x_n)$ and $g(\mathbf{x}) = g(x_1, x_2, \dots, x_n)$ of n variables as appropriate subsets of units of Boolean space $\{0, 1\}^n$. Let's represent them by Boolean 2^n -vectors \mathbf{f} and \mathbf{g} , displaying such vectors below in the more convenient for visual perception matrix form. In such representation the two-place Boolean operations $f \vee g$, $f \wedge g$, $f \oplus g$, $f \rightarrow g$ are easily implemented as parallel component-wise operations over appropriate Boolean vectors designated $\mathbf{f} \vee \mathbf{g}$, $\mathbf{f} \wedge \mathbf{g}$ (or, simpler, $\mathbf{f}\mathbf{g}$), $\mathbf{f} \oplus \mathbf{g}$, $\mathbf{f} \rightarrow \mathbf{g}$. Including these operations in the formed basis, we shall supplement them by some operations of interaction between different components within the framework of one Boolean vector.

Let's remind, that the function $f(\mathbf{x})$ can be represented as Shannon disjunctive decomposition by an arbitrary variable x_i , which coefficients f_{i0} and f_{i1} are Boolean functions obtained as a result of substitution of variable x_i by values 0 or 1 :

$$f(\mathbf{x}) = \bar{x}_i f_{i0} \vee x_i f_{i1},$$

Using vector representation of the function, we shall designate these operations accordingly through $\mathbf{f} - i$ and $\mathbf{f} + i$. They are easily implemented in the Boolean space on couples of elements adjacent by the variable x_i . When executing the operation $\mathbf{f} - i$ both elements of the couple gain the value of the element defined by the condition $x_i = 0$, at execution of the operation $\mathbf{f} + i$ – the value of the other element corresponding to value 1 of variable x_i .

Let's show examples of such operations, and also of their compositions.

	f	f-2	f-2-5	
	-----	-----	-----	4
	---- ----	---- ----	---- ----	3
	-- -- -- --	-- -- -- --	-- -- -- --	2
	- - - - -	- - - - -	- - - - -	1
	0110110101011110	0101111101011111	0101111101011111	
	0010010000010110	0000010100000101	0101111101011111	
	1100101001110001	1111101001010000	1111101001010000	
	0100010111010011	0101010111110000	1111101001010000	
	1101110111101110	1010101000010001	0000000000110011	
	0100010001100110	0101010100110011	0000000000110011	
	1010101000010001	1010101000010001	1100110011111111	
	0101010100110011	0101010100110011	1100110011111111	
6 5	f+3	f+3+6	f+1-3+5	

By interaction of adjacent units there are implemented also operation $Inv_i f$ of inverting the function f at the variable x_i (adjacent elements interchange their values), and so-called operations of symmetrization $S_i^* f$, in which both elements get value defined by the two-place operation $*$ above their initial values [Zakrevskij, 1963]. As a result of these operations the function

$$f(x) = \bar{x}_i f_{i0} \vee x_i f_{i1}$$

is transformed correspondingly into functions

$$\bar{x}_i f_{i1} \vee x_i f_{i0}, \quad \bar{x}_i (f_{i0} * f_{i1}) \vee x_i (f_{i0} * f_{i1}).$$

Examples of these operations are shown below.

	f	$Inv_4 f$	$S_4^\vee f$	
	-----	-----	-----	4
	---- ----	---- ----	---- ----	3
	-- -- -- --	-- -- -- --	-- -- -- --	2
	- - - - -	- - - - -	- - - - -	1
	0110110101011110	01011111001101101	0111111101111111	
	0010010000010110	0001011000100100	0011011000110110	
	1100101001110001	0111000111001010	1111101111111011	
	0100010111010011	1101001101000101	1101011111010111	
	0000110000001100	0011001100110011	0111111101111111	
	0000000000000000	0011001000110010	0111111101111111	
	1100000000110000	1011101110111011	1111111111111111	
	0000000011000011	1001011010010110	1111111111111111	
6 5	$S_1^{\wedge} f$	$S_4^{\oplus} f$	$S_{4,5}^{\vee} f$	

Here $S_{4,5}^{\vee} f$ means composition $S_4^{\vee} (S_5^{\vee} f)$.

Search by traces. Triads and fragments

The method of decomposition suggested below is based on the following reasons which key moments are given in the form of assertions. They were formulated before for the case of completely specified Boolean functions [Zakrevskij, 2006], but became valid when partial Boolean functions are considered.

Suppose two partitions u/v and u^*/v^* are given, such that $u^* \subseteq u$ and $v^* \subseteq v$. Let's speak, that partition u^*/v^* submits to partition u/v , and call it a *trace* of u/v ,

Assertion 1. If a partial Boolean function $f(\mathbf{x})$ is decomposable at partition \mathbf{u}/\mathbf{v} , it is decomposable as well at partition $\mathbf{u}^*/\mathbf{v}^*$.

Corollary. If the function $f(\mathbf{x})$ is not decomposable at partition $\mathbf{u}^*/\mathbf{v}^*$, it is not decomposable also at partition \mathbf{u}/\mathbf{v} .

Let's assume $|\mathbf{u}| = k$ and $|\mathbf{v}| = m$. Partition with $k = 2$ and $m = 1$ we shall term as a *triad*. It is the simplest of partitions, at which some nontrivial decomposition can take place.

Assertion 2. A partial Boolean function is decomposable, if and only if it is decomposable if only at one of triads.

Therefore in the offered method the search for partition \mathbf{u}/\mathbf{v} starts with the search of its traces on the set of triads, i.e. with looking for an appropriate triad. The needed checking of triads can be fulfilled fast enough, as their number is not large, being significantly less than the number of all weak partitions.

Assertion 3. The number of triads is equal to $C_n^2 (n-2) = \frac{n(n-1)(n-2)}{2}$.

Suppose, that some appropriate triad $(x_p, x_q)/x_r$ is detected. If it submits to the required partition \mathbf{u}/\mathbf{v} , the latter can be found, having put for the beginning $\mathbf{u} = (x_p, x_q)$ and $\mathbf{v} = (x_r)$, and then sequentially expanding these two sets, sorting out remaining variables and testing them on possibility of inclusion into set \mathbf{u} or \mathbf{v} .

By reviewing some concrete triad \mathbf{u}/\mathbf{v} the Boolean space $M = \{0, 1\}^n$, where the partial Boolean function $f(\mathbf{x})$ is presented, is divided into $2^{|\mathbf{u}|}$ intervals corresponding to different values of vector \mathbf{w} . On each of them the corresponding coefficient f_i of disjunctive decomposition of the function on variables of set \mathbf{w} is given. It represents some partial Boolean function of variables x_p, x_q, x_r . As a matter of convenience of subsequent reasoning we shall present each of these coefficients by a ternary matrix size 4×2 , which rows correspond to values of the two-component vector \mathbf{u} , and columns – to values of the one-component vector \mathbf{v} . Let's designate this matrix T_i and name it a *fragment*. Thus, the 2^n -element ternary matrix representing the function $f(\mathbf{x})$, is decomposed into 2^{n-3} eight-element fragments specifying functions $f_i(x_p, x_q, x_r)$.

A concrete example of such splitting into eight fragments for a partial Boolean function $f(a, b, c, d, e, f)$ and triad $(a, b)/c$ is shown below.

									f
									e
									d
									c
	10	0-	1-	11	0-	10	-1	10	
	-1	00	-1	1-	00	-1	1-	-0	
	0-	1-	-0	10	10	0-	11	-1	
	1-	01	11	-1	00	10	-0	11	
									b a

Assertion 4. The function $f(\mathbf{x})$ can be decomposed at triad $(x_p, x_q)/x_r$, if and only if each of the coefficients $f_i(x_p, x_q, x_r)$ also is decomposable at the same triad.

It follows from here, that the probability of decomposability of function $f(\mathbf{x})$ at a concrete triad is equal to γ^k , where k is the number of coefficients equal 2^{n-3} and γ – the probability of decomposability of one coefficient. In the case of a completely specified Boolean function the last probability is approximated by the value $1/3$, and with growth of uncertainty decreases. Nevertheless, the probability of decomposability of the function $f(\mathbf{x})$ quickly decreases with growth of the number of variables in it.

Checking triads for fitness

So, a triad is appropriate, if each of fragments of the corresponding splitting of the ternary matrix is suitable. That means, the partial function $f(\mathbf{x})$ can be completely defined in such a way, that each fragment will contain no more than two types of Boolean rows (each having equal rows). In other words, a fragment is suitable, if it contains no more than two classes of compatible rows. Remind, that two ternary rows are compatible, i.e. they could become

equal by changing "–"-values of some components for 1 or 0, if they are not orthogonal. It follows from here, that the fragment is suitable, if the graph of orthogonality of its rows is bichromatic [Harary, 1969].

Let's offer the following way of checking fragments with the purpose of detection of suitable ones among them. Any fragment contains four rows, therefore the graph of orthogonality has four vertices. It is bichromatic, if it has no cycle of length three. Let's select arbitrary two different vertices. If such a cycle exists, then one of the selected vertices will belong to it. Therefore, it is enough to test each of these two vertices on belonging to a cycle of length three. If such belonging will not be revealed, graph is bichromatic, and the triad is suitable.

Necessary and sufficient condition of entering a vertex, i.e. corresponding row, in a cycle of length three could be formulated as follows: among rows orthogonal to the given one, there exist mutually orthogonal rows.

For example, the first of the shown below fragments appears to be suitable, and the second – no, as there is a cycle of length three, composed by three last rows: each of them is orthogonal to the other two.

10	–1	1
–1	00	2
0–	1–	3
01	01	4

The offered way is implemented by the following algorithm, which is remarkable by that it checks on fitness simultaneously all 2^{n-3} fragments generated by the given triad, and finds out by that if the triad is appropriate. The partial function $f(\mathbf{x})$ is represented by a couple of Boolean vectors \mathbf{f}^0 and \mathbf{f}^1 , in first of which by 1s are marked the values 0 of the function and in the second - values 1. The splitting of space into fragments is fulfilled by the triad $(x_p, x_q) / x_r$.

To begin with, first rows of fragments are checked, which form the initial coefficient f^- of decomposition of the function $f(\mathbf{x})$ on variables x_p and x_q . The rows orthogonal to the checked up row, are marked by value 1 in the computed vector \mathbf{g} , and their values are fixed by the couple of vectors \mathbf{h}^0 and \mathbf{h}^1 , checked up further for orthogonality. Alike the initial vectors \mathbf{f}^0 и \mathbf{f}^1 , they are Boolean vectors with 2^n components.

$$\begin{aligned} \mathbf{h}^0 &:= (\mathbf{f}^0 - p) - q && \text{Getting the initial coefficient } f^- \\ \mathbf{h}^1 &:= (\mathbf{f}^1 - p) - q && \\ \mathbf{g} &:= S_r^\vee (\mathbf{h}^0 \mathbf{f}^1 \vee \mathbf{h}^1 \mathbf{f}^0) && \text{Finding coefficients orthogonal to } f^- \\ \mathbf{h}^0 &:= S_u^\vee (\mathbf{f}^0 \mathbf{g}) && \text{Computing their intersection} \\ \mathbf{h}^1 &:= S_u^\vee (\mathbf{f}^1 \mathbf{g}) && \end{aligned}$$

If it turns out that $\mathbf{h}^0 \mathbf{h}^1 \neq \mathbf{0}$, the triad is accepted as not appropriate. In case if $\mathbf{h}^0 \mathbf{h}^1 = \mathbf{0}$ the final rows of fragments are checked, which constitute the final coefficient f^+ .

$$\begin{aligned} \mathbf{h}^0 &:= (\mathbf{f}^0 + p) + q && \text{Getting the final coefficient } f^+ \\ \mathbf{h}^1 &:= (\mathbf{f}^1 + p) + q && \\ \mathbf{g} &:= S_r^\vee (\mathbf{h}^0 \mathbf{f}^1 \vee \mathbf{h}^1 \mathbf{f}^0) && \text{Finding coefficients orthogonal to } f^+ \\ \mathbf{h}^0 &:= S_u^\vee (\mathbf{f}^0 \mathbf{g}) && \text{Computing their intersection} \\ \mathbf{h}^1 &:= S_u^\vee (\mathbf{f}^1 \mathbf{g}) && \end{aligned}$$

If $\mathbf{h}^0 \mathbf{h}^1 \neq \mathbf{0}$, then the triad is not appropriate. On the other hand, if $\mathbf{h}^0 \mathbf{h}^1 = \mathbf{0}$, the triad is accepted as appropriate.

Example. Let's return to regarding the partial Boolean function $f(a, b, c, d, e, f)$, representing it by a couple of Boolean vectors (rolled up into matrices) \mathbf{f}^0 and \mathbf{f}^1 :

$$\mathbf{e} := S_s^\vee \mathbf{g}$$

$$\mathbf{h}^0 := S_u^\vee (\mathbf{f}^0 \mathbf{e})$$

$$\mathbf{h}^1 := S_u^\vee (\mathbf{f}^1 \mathbf{e})$$

and if $\mathbf{h}^0 \mathbf{h}^1 = \mathbf{0}$, then s is included into \mathbf{v} by implementing operations

$$\mathbf{v} := \mathbf{v} \cup \{s\}, \mathbf{g} := \mathbf{e}.$$

So the final value of set \mathbf{v} is found.

Expanding set \mathbf{u} . The maximum expansion of set \mathbf{u} is found similarly. If it is known, that the required partition is strict, it is possible to put $\mathbf{u} = \mathbf{x} / \mathbf{v}$ and, probably, to test the function for decomposability, as the algorithm used is heuristic. Let's remark, however, that the probability of obtaining by this algorithm erroneous solution fast tends to zero with growth of the number of variables n .

If the required partition could be not strict, it is necessary to test all elements from initial value of set $\mathbf{x} \setminus (\mathbf{u} \cup \mathbf{v})$ for the possibility of including them into set \mathbf{u} .

Check of the immediate element s can be fulfilled by the following heuristic algorithm, which partly implements the procedure circumscribed in [Zakrevskij, 2007]. The algorithm considers the initial coefficient f^- of the function f decomposition by the current value of set \mathbf{u} , finds orthogonal to it coefficients, checks them for compatibility and, in case of compatibility, includes element s in set \mathbf{u} without further check.

$$\mathbf{e} := \mathbf{u} \cup \{s\}$$

$$\mathbf{h}^0 := \mathbf{f}^0 - \mathbf{e}$$

$$\mathbf{h}^1 := \mathbf{f}^1 - \mathbf{e}$$

$$\mathbf{g} := S_v^\vee (\mathbf{h}^0 \mathbf{f}^1 \vee \mathbf{h}^1 \mathbf{f}^0)$$

$$\mathbf{h}^0 := S_u^\vee (\mathbf{f}^0 \mathbf{g})$$

$$\mathbf{h}^1 := S_u^\vee (\mathbf{f}^1 \mathbf{g})$$

If $\mathbf{h}^0 \mathbf{h}^1 = \mathbf{0}$, then s is included into \mathbf{u} by operation $\mathbf{u} := \mathbf{e}$.

In such a way the set \mathbf{u} is found and, therefore, the whole partition \mathbf{u}/\mathbf{v} .

Note, that the operation of looking for coefficient f^- is presented in this algorithm in abbreviated form, by expressions $\mathbf{h}^0 := \mathbf{f}^0 - \mathbf{e}$ and $\mathbf{h}^1 := \mathbf{f}^1 - \mathbf{e}$, instead of more detailed

$$\mathbf{h}^0 := (\dots((\mathbf{f}^0 - e_1) - e_2) - \dots) - e_t,$$

$$\mathbf{h}^1 := (\dots((\mathbf{f}^1 - e_1) - e_2) - \dots) - e_t,$$

where $\mathbf{e} = (e_1, e_2, \dots, e_t)$.

Conclusion

In this paper, the heuristic algorithm is offered for finding such weak two-block partition on the set of variables of a partial Boolean function, on which the function can be decomposed. The algorithm is effective, if there exists a good solution "«hidden" in vector representation of the function of many variables. In this case the search of the partition is reduced to recognition of the latter.

Acknowledgement

This work was supported by Belarusian Republican Fond of Fundamental Researches (Project **Φ07MC-034**).

Bibliography

- [Ashenurst, 1959] Ashenurst R.L. The decomposition of switching functions. – Proc. International Symposium on the Theory of Switching, Part 1. – Harvard University Press, Cambridge, 1959, pp. 75-116.
- [Curtis, 1962] Curtis H.A. Design of switching circuits. – Van Nostrand, Princeton, N. J., 1962.
- [Povarov, 1954] Povarov G.N. About functional decomposition of Boolean functions. – Reports of the AS of USSR, 1954. – V. 4, No 5 (in Russian).
- [Harary, 1969] Frank Harary. Graph theory. – Addison-Wesley Publishing Company : Reading, Massachusetts; Menlo Park, California; London; Don Mills, Ontario. 1969.
- [Zakrevskij, 1963] Zakrevskij A.D. Universal system for solving problems the type relay system synthesis. – Annals of Siberian Physical-Technical Institute, 1963. – Issue 42, pp. 9-37 (in Russian).
- [Zakrevskij, 2006] Arkadij Zakrevskij. Decomposition of Boolean functions by looking for tracks of a good solution. – Proceedings of XII-th International Conference Knowledge-Dialogue-Solution (KDS–2006), June 20-30, 2006, Varna, Bulgaria. pp. 211-216.
- [Zakrevskij, 2007] Zakrevskij A.D. Decomposition of partial Boolean functions – checking for decomposability at a given partition. – Informatics, 2007, 1(13), (in Russian).
-

Author's Information

Arkadij Zakrevskij - United Institute of Informatics Problems of the NAS of Belarus, Surganov Str. 6, 220012 Minsk, Belarus; e-mail: zakr@newman.bas-net.by

ВЕРИФИКАЦИЯ ЛОГИЧЕСКИХ ОПИСАНИЙ С ФУНКЦИОНАЛЬНОЙ НЕОПРЕДЕЛЕННОСТЬЮ

Людмила Черемисинова, Дмитрий Новиков

Резюме. Задача верификации, заключающаяся в доказательстве поведенческой эквивалентности двух описаний одного и того же устройства, рассматривается для случая, когда одно из них поведенчески не полностью определено. В такой постановке задача верификации сводится к проверке эквивалентности логических описаний на области задания того описания, которое определено не на всей области возможных значений аргументов. Рассматриваются пути решения этой задачи для разных форм задания описаний. Предлагаемые методы основаны на быстрых булевых вычислениях над булевыми и троичными векторами большой размерности.

Ключевые слова: верификация, булевы вычисления, автоматизация проектирования, моделирование.

Введение

Задача верификации [1] заключается в доказательстве поведенческой эквивалентности двух описаний одного и того же устройства, представляющих разные проектные решения, полученные в ходе логического проектирования. Эта задача в литературе традиционно рассматривается для случая, когда оба описания поведенчески полностью определены и представляют структурные реализации одного и того же устройства. Соответственно задача верификации сводится к проверке эквивалентности пары комбинационных схем. В настоящее время усилиями ведущих научных групп университетов и крупных фирм, занимающихся проектированием дискретных устройств, разработаны [1 – 4] и разрабатываются эффективные методы решения задачи верификации в такой постановке.

В настоящей работе задача верификации рассматривается для случая, когда одно из описаний поведенчески не полностью определено. Этот случай возникает на начальных этапах проектирования, когда проектируемое устройство частично определено, т. е. существуют наборы значений входных переменных, которые при штатном функционировании устройства никогда не появляются на входах устройства. Соответственно поведение такого устройства на этих наборах в процессе логического синтеза может быть доопределено произвольным образом. Исходное логическое описание верифицируемого комбинационного устройства с функциональной неопределенностью может быть представлено в виде системы частично определенных булевых функций.

Рассматривается случай, когда комбинационная структура, полученная в процессе декомпозиции, минимизации и/или синтеза, представляет собой многоблочную структуру, каждый блок которой описывается системой полностью определенных булевых функций. Задача проверки функциональной эквивалентности двух логических описаний трансформируется в этом случае в задачу проверки реализуемости одного логического описания с функциональной неопределенностью другим описанием, полностью определенным.

Задача верификации сводится к проверке эквивалентности логических описаний на области определения описания с неопределенностью, задающего систему частично определенных булевых функций. Рассматриваются пути решения этой задачи для разных форм задания системы частично определенных булевых функций: на наборах и интервалах значений аргументов. Предлагаются методы верификации, в основе которых лежат быстрые булевы вычисления над булевыми и троичными векторами большой размерности.

Основные определения

В настоящей работе рассматривается случай, когда одно описание представляет собой систему $F = \{f_1(X), f_2(X), \dots, f_m(X)\}$ ($X = \{x_1, x_2, \dots, x_n\}$) частично определенных булевых функций, второе – многоуровневую многовыходную комбинационную схему S в базисе элементов трех типов: инверторы, многоместные конъюнкторы и многоместные дизъюнкторы. К такому типу схем может быть преобразована многоуровневая схема из любой сложности элементов или блоков. Задача решается для двух основных видов задания системы F частично определенных булевых функций:

- 1) на наборах значений входных переменных из X ;
- 2) на интервалах в n -мерном булевом пространстве переменных из X .

В первом случае каждую частично определенную булеву функцию $f_i(X)$ можно задать парой множеств M_i^1 и M_i^0 наборов значений переменных, на которых она принимает значения 1 и 0 соответственно, предполагая, что на остальных наборах ее значение не определено. Система F булевых функций

представляется в этом случае парой матриц: булевой \mathbf{B} , задающей все наборы из $M = \bigcup_{i=1}^m (M_i^1 \cup M_i^0)$,

и троичной \mathbf{T} , задающей значения функций на них – “1”, “0”, если они определены, и “–”, если не определены. Матрицы \mathbf{B} и \mathbf{T} имеют по l строк (где l – число наборов в множестве M) и по n и m столбцов соответственно.

Во втором случае каждая частично определенная булева функция $f_i(X)$ представляется парой множеств U_i^1 и U_i^0 , но не наборов, а интервалов значений переменных, каждый из которых задает в общем случае более чем один набор. Точнее интервал ранга k включает в себя 2^{n-k} наборов значений n переменных, понимая под рангом интервала число его компонент, имеющих значение 0 или 1. Система F

представляется в этом случае парой троичных матриц: \mathbf{U} , задающей интервалы из $U = \bigcup_{i=1}^m (U_i^1 \cup U_i^0)$, и

\mathbf{T} , задающей значения функций на них – “1”, “0”, если они определены, или “–”, в противном случае.

Задание функции на интервалах в отличие от задания на наборах имеет следующие особенности. Интервалы $u_i, u_j \in U$ могут пересекаться (наборы – нет). Значение “–” элемента t_i матрицы T может означать, что 1) значение функции f_j не определено на всем интервале u_i (на всех входящих в него наборах); 2) функция f_j не принимает одно и то же значение на всем интервале u_i : в этом интервале существует, как минимум, два набора, на которых она имеет разные значения из множества $\{1, 0, -\}$. Таким образом, значение “–” компоненты t_i матрицы T не всегда означает, что значение функции f_j не определено (в том смысле, что оно может быть при реализации доопределено произвольным образом) на всем интервале u_i , как при первом типе задания функций. Значение “–” говорит лишь о том, что функция f_j не принимает одно и то же значение на всем интервале u_i .

Преобразование представления многоблочной структуры

Каждый блок многоблочной структуры S является многовыходным и задается системой дизъюнктивных нормальных форм (ДНФ), представляемой парой матриц: трюичной, строки которой задают элементарные конъюнкции, и булевой, в которой единичная компонента на пересечении i -й строки и j -го столбца равна 1, если i -я конъюнкция входит в j -ю ДНФ. Множество входных переменных структуры в целом совпадает с множеством X , а m реализуемых выходных функций $y_i(X)$, являются полностью определенными. Условие реализуемости системы F структурой S заключается в том, что для всех $f_i(X) \in F$ и соответствующих им функций $y_i(X)$ должно выполняться: $M_{f_i}^1 \subseteq M_{y_i}^1$ и $M_{f_i}^0 \subseteq M_{y_i}^0$, т. е. на области определения каждой функции $f_i(X)$ значения функций f_i и y_i должны совпадать.

Каждый блок структуры S можно рассматривать как трехуровневую многовыходную логическую схему, первый уровень которой составляют инверторы, второй – многоместные конъюнкторы, а третий уровень – многоместные дизъюнкторы. Перенумеруем полюсы многоблочной структуры, начиная с ее входных полюсов, и далее выходных полюсов инверторов, конъюнкторов и дизъюнкторов всех блоков. Припишем им переменные z_i , выделив входные и выходные полюсы многоблочной структуры. В результате многоблочная структура может рассматриваться как многовыходная логическая схема C из инверторов, конъюнкторов и дизъюнкторов, на выходных полюсах которой должны быть реализованы функции исходной системы частично определенных булевых функций и которая состоит из инверторов, дизъюнкторов и конъюнкторов. Предлагаемый ниже метод годится также и для сетей из любых других элементов, реализующих симметрические логические операции $\varphi(z_1, z_2, \dots, z_k)$.

Ранжируем (пронумеруем) элементы схемы C таким образом, чтобы любая межэлементная связь соединяла выходной полюс элемента с меньшим номером с входным полюсом элемента с большим номером. Необходимым и достаточным условием ранжируемости схемы является отсутствие в ней контуров. Это условие заведомо выполняется для структуры каждого блока, предполагается, что многоблочная структура, представляемая схемой C , в целом этому условию также удовлетворяет.

Задача верификации для случая задания системы на наборах значений переменных

Идея предлагаемого метода проверки реализуемости системы частично определенных булевых функций многоблочной структурой S состоит в моделировании поведения многовыходной логической схемы на области задания системы F . Условие реализуемости сводится к проверке для каждого набора значений переменных $\mathbf{b}_i \in \mathbf{B}$ условия неортогональности булева вектора значений $y_i(\mathbf{b}_i)$ выходных функций структуры S трюичному вектору $\mathbf{t}_i \in \mathbf{T}$.

Будем использовать идею двоичного параллельного моделирования [5], проводя моделирование многовыходной логической схемы C сразу на всех входных наборах из множества M . При параллельном моделировании схемы на l наборах состояние каждого полюса (включая входные и выходные) схемы представляется булевым вектором размерности l . Таким образом, каждый, вектор представляет состояния одного полюса для всех l входных состояний схемы, а совокупность одноименных компонент всех векторов соответствует состоянию всех полюсов схемы для одного и того же входного набора.

В начале моделирования имеется упорядоченное множество n булевых векторов размерности l , представляющих состояния n входных полюсов во всех l наборах и задаваемых столбцами матрицы \mathbf{B} . Затем последовательно просматриваются элементы (в порядке возрастания их номеров) предварительно ранжированной схемы S , реализующие функции $\varphi(z_{1i}, z_{2i}, \dots, z_{ki})$, и выполняется функция φ над ее аргументами $z_{1i}, z_{2i}, \dots, z_{ki}$. Так как каждому из аргументов z_{ji} функции соответствует булев вектор \mathbf{z}_{ji} , то операция сводится к выполнению покомпонентной операции φ над булевыми векторами $\mathbf{z}_{1i}, \mathbf{z}_{2i}, \dots, \mathbf{z}_{ki}$. Результатом операции является новый вектор \mathbf{z}_i , также размерности l .

После просмотра последнего элемента схемы будут найдены реакции схемы на все наборы значений входных переменных, входящие в область M определения системы F . При этом каждая выходная функция y_i схемы имеет определенное значение (0 или 1) на всех наборах значений входных переменных, в частности и на тех, на которых определенное значение имеет и соответствующая функция $f_i \in F$. Остается только сравнить значения функций значения y_i и f_i на области $M_i^1 \cup M_i^0$ на ортогональность, это сводится к проверке, не ортогональны ли следующие пары векторов: троичный вектор \mathbf{t}_i , соответствующий i -му столбцу матрицы \mathbf{T} , и булев вектор \mathbf{z}_p , соответствующий i -му выходному полюсу схемы. Реализуемость системы F многоблочной структурой S имеет место, если все эти пары векторов не ортогональны. В случае ортогональности некоторой пары можно путем обратного прослеживания логической схемы S найти причину, ответственную за нарушение условия реализуемости: определить блок структуры S , ее выход или ДНФ системы задающей, задающей его функциональное описание, и наконец, конъюнкции этой ДНФ.

Таким образом задача проверки реализуемости системы частично определенных булевых функций многоблочной структурой сводится к выполнению булевых вычислений над векторами (последовательностями бит) одной и той же (но произвольной) размерности.

Задача верификации для случая задания системы на интервалах значений переменных

Возможны два пути решения задачи верификации системы частично определенных булевых функций, заданных на интервалах значений переменных: 1) сведение ее к рассмотренному выше случаю верификации системы частично определенных булевых функций путем перехода от задания функций на интервалах к заданию на наборах; 2) решение этой задачи на области, заданной множеством интервалов.

Первый путь решения целесообразно использовать в том случае, когда мало число интервалов множества M , имеющих ранг меньший, чем n , и эти ранги близки к n . В этом случае можно надеяться, что область задания системы частично определенных булевых функций не возрастет резко по размеру. Второй путь решения целесообразно использовать в том случае, когда число интервалов ранга, меньшего, чем n , велико и эти интервалы имеют ранг, значительно меньший, чем n . В этом случае при переходе к заданию функций на наборах значений переменных размер области задания системы может возрасти настолько, что задача верификации может стать практически не решаемой. В работе рассматривается решение задачи верификации системы частично определенных булевых функций, заданных на интервалах значений переменных.

Как и в случае задания системы функций на наборах значений аргументов будем проводить параллельное моделирование многовыходной логической схемы, но не наборах значений переменных, а на интервалах из множества U . При таком моделировании состояние каждого полюса схемы представляется не булевым, а в общем случае троичным вектором. Таким образом, каждый, вектор представляет состояния одного полюса для всех l входных состояний схемы, а совокупность одноименных компонент всех векторов соответствует состоянию всех полюсов схемы для одного и того же интервала значений входных переменных. При этом значение "–" i -й компоненты этого вектора говорит в общем случае лишь о том, что функция, реализуемая полюсом, имеет разные значения на разных наборах i -го интервала.

В начале моделирования имеется упорядоченное множество n интервалов l -мерного булева пространства, задаваемых столбцами матрицы U . Последовательно просматриваются элементы предварительно ранжированной схемы C , реализующие функции $\varphi_i(z_{1i}, z_{2i}, \dots, z_{ki})$, и выполняется функция φ_i над ее аргументами $z_{1i}, z_{2i}, \dots, z_{ki}$. Так как каждому из аргументов z_{ji} функции соответствует интервал z_{ji} , то операция сводится к выполнению покомпонентной операции φ_i над троичными векторами $z_{1i}, z_{2i}, \dots, z_{ki}$. Результатом операции является новый троичный же вектор z_i . Ниже приводится определение основных операций над троичными переменными для случая принятой выше интерпретации неопределенного значения “-”:

a:	0	0	0	-	-	-	1	1	1
b:	0	-	1	0	-	1	0	-	1

\bar{a} :	1	1	1	-	-	-	0	0	0
$a \vee b$:	0	-	1	-	-	1	1	1	1
$a \wedge b$:	0	0	0	0	-	-	0	-	1

После окончания процесса моделирования проверяется, реализуется ли система F функций многоблочной структурой S . При моделировании схемы C на интервалах значений переменных, эта проверка может потребовать значительных затрат по времени. Сначала троичный вектор t_i значений каждой функции $f_i \in F$ на всех интервалах задания системы F сравнивается с троичным вектором z_p , соответствующим i -му выходному полюсу схемы. Возможны три случая.

1. Векторы t_i и z_p ортогональны по j -й компоненте. Делается вывод, что схема C не реализует функцию f_i .
2. Вектор t_i поглощает вектор z_p , т.е. все компоненты вектора t_i , значения которых отличны от “-”, совпадают по значению с соответствующими компонентами вектора z_p . Делается вывод, что схема C реализует функцию f_i .
3. Значение j -й компоненты вектора z_p равно “-”, тогда как значение этой компоненты в векторе t_i равно 1 или 0. Невозможно дать однозначный ответ на вопрос реализует ли схема C функцию f_i .

В третьем случае необходим дополнительный анализ, позволяющий найти причину несоответствия откликов выхода y_p схемы и значения функции f_p на наборах значений входных переменных из интервала t_i . Самый простой метод заключается в повторном моделировании схемы C , но уже на наборах значений переменных из этого интервала. Более изощренный метод заключается в оперативном анализе результатов моделирования, проводимом в его процессе. Если в результате вычислений на некотором шаге получается неопределенное значение на некотором полюсе z_q схемы C , то соответствующий интервал значений входных переменных дробится на подинтервалы, на которых z_q принимает определенные значения. Это делается для того, чтобы предотвратить дальнейшее распространение неопределенности при моделировании. Например, если $z_q = x_1 x_3 x_4 x_6$, а вычисления ведутся на интервале $u_i = 1 0 - 1 - -$, то $z_q(u_i) = -$, то расщепляем интервал на три следующие: $u_i^1 = 1 0 1 1 - 1$, $u_i^2 = 1 0 0 1 - -$, $u_i^3 = 1 0 - 1 - 0$, на элементах которого z_q принимает одно и то же значение: $z_q(u_i^1) = 1$, $z_q(u_i^2) = z_q(u_i^3) = 0$.

Для того, чтобы избежать излишних расщеплений интервалов при моделировании исходную систему частично определенных булевых функций имеет смысл ортогонализировать. После ортогонализации каждый из интервалов задания функций системы будет обладать тем свойством, что все функции будут принимать одно и то же значение (1, 0, -) на всех его элементах. При этом значение “-” некоторой функции будет означать, что значение этой функции не определено на всем интервале.

Ортогонализация системы частично определенных булевых функций

В работах [6, 7] задача ортогонализации ставится для системы полностью определенных булевых функций $F = \{f_1(X), f_2(X), \dots, f_m(X)\}$, заданных на интервалах значений переменных из X парой матриц: троичной, представляющей интервалы значений переменных и булевой, отмечающей единицами те интервалы, на которых функции принимают значение 1. Задача заключается в нахождении совокупности

взаимно ортогональных полностью определенных булевых функций $\varphi_1(X), \varphi_2(X), \dots, \varphi_r(X)$ такой, что любую функцию $f_i \in F$ можно выразить в виде дизъюнкции некоторых из ортогональных функций φ_j ($j = 1, 2, \dots, r$), причем общее число r этих функций должно быть минимальным. Под ортогональностью функций φ_j и φ_k понимается выполнение условия $\varphi_j \wedge \varphi_k = 0$ при любых значениях аргументов из X . При этом, если функции φ_j взаимно ортогональны, то интервалы, на которых определены разные функции не пересекаются, но интервалы, на которых определена одна и та же функция могут и пересекаться.

Аналогично системам полностью определенных булевых функций задача ортогонализации может быть поставлена и для системы частично определенных булевых функций $F = \{f_1(X), f_2(X), \dots, f_m(X)\}$, если задавать каждую функцию $f_i \in F$ двумя функциями f_i^1 и f_i^0 . В область единичных значений функции f_i^1 включаются те интервалы, на которых функция f_i принимает значение 1, а в область единичных значений функции f_i^0 включаются те интервалы, на которых функция f_i принимает значение 0. На остальной области булева пространства обе эти функции принимают значения 0. Таким образом от системы частично определенных булевых функций F мы переходим к системе F' полностью определенных функций, содержащей удвоенное число функций. Решаем задачу ортогонализации для системы F' одним из известных методов [6, 7], а затем совершаем обратный переход к системе F , но уже ортогонализованной.

Отличительной чертой ортогонализованной системы частично определенных булевых функций является то, что в ней интервалы, на которых хотя бы одна из функций системы принимает разные значения, не пересекаются. Это означает, что значение “–” функции на некотором интервале говорит о том, что эта функция не определена на всех наборах значений аргументов, входящих в интервал, и ее значение на этом интервале может быть при реализации доопределено произвольным образом.

Заключение

Предложенные методы ориентированы на верификацию систем слабо определенных булевых функций – систем, для которых число наборов области их определения M (на наборах которой значение хотя бы одной функции системы определено) существенно меньше числа наборов области их неопределенности, где значения всех функций не определены. Описанные методы годятся для верификации логических описаний большой размерности.

Задача проверки реализуемости системы частично определенных булевых функций многоблочной структурой сведена к булевым вычислениям над троичными и булевыми векторами произвольной размерности. Методы допускают эффективную программную реализацию. Для удобства программирования операций над булевыми и троичными векторами произвольной размерности в языке C++ ранее были разработаны классы CBV [8] и CTM [9].

Сложность $S(M)$ программной реализации описанных методов верификации линейно зависит от общего числа полюсов моделируемой схемы (суммарного числа входных полюсов всех элементов схемы S) и от числа байтов (или 32-разрядных слов), используемых для представления l -разрядного вектора (здесь l – число интервалов или наборов значений аргументов области задания системы частично определенных булевых функций) [10].

Библиография

- Drechsler R. and others. Advanced Formal Verification. – Kluwer Academic Publishers, 2005. – 249 .
- Mishchenko A., Chatterjee S., Brayton R., Eem N. Improvements to Combinational Equivalence Checking // Proc. ICCAD'06, Nov. 5–9, 2006. – San Jose, CA, 2006.
- Ganai M.K., Zhang L., Ashar P., Gupta A., Malik S. Combining strengths of circuit-based and CNF-based algorithms for a high-performance SAT solver // Proc. ACM/IEEE Design Automation Conference, 2002 – P. 747–750.

- Goldberg E., Novikov Y. BerkMin: A fast and robust SAT-Solver // Proc. European Design and Test Conference, 2002. – P. 142–149.
- Закревский А.Д., Поттосин Ю.В., Черемисинова Л.Д. Основы логического проектирования. Кн. 3. Проектирование устройств логического управления. – Мн.: ОИПИ НАН Беларуси, 2006. – 252 с.
- Кузнецов О.П. Ортогональные системы булевых функций и их применение к анализу и синтезу логических сетей // Автоматика и телемеханика, 1970. – № 10. – С. 117–128.
- Поттосин Ю.В., Шестаков Е.А. Ортогонализация системы полностью определенных булевых функций // Логическое проектирование. – Мн.: Ин-т техн. кибернетики НАН Беларуси, 2000. – Вып. 5. – С. 107–115.
- Василькова И.В., Романов В.И. Булевы векторы и матрицы в C++ // Логическое проектирование. – Мн.: Ин-т техн. кибернетики НАН Беларуси, 1997. – С. 150–158.
- Черемисинов Д.И., Черемисинова Л.Д. Троичные векторы и матрицы в C++ // Логическое проектирование. – Мн.: Ин-т техн. кибернетики НАН Беларуси, 1998. – Вып. 3. – С. 146–156.
- Романов В.И. Оптимизация булевых вычислений на программном уровне // Танаевские чтения. Доклады Второй научной конференции (28 марта 2005 г., Минск). – Мн.: ОИПИ НАН Беларуси. 2005. – С. 91–93.

Информация об авторах

Людмила Дмитриевна Черемисинова – д.т.н., главный научный сотрудник Объединенного института проблем информатики Национальной академии наук Беларуси, ул. Сурганова, 6, Минск, 220012, Беларусь, e-mail: cld@newman.bas-net.by

Дмитрий Яковлевич Новиков – аспирант Объединенного института проблем информатики Национальной академии наук Беларуси, ул. Сурганова, 6, Минск, 220012, Беларусь, e-mail: yakov_nov@tut.by

ОБОБЩЁННЫЕ ВАРИАНТЫ ПРЕОБРАЗОВАНИЯ ХОКА: СТАТИСТИЧЕСКИЙ И АЛГЕБРАИЧЕСКИЙ АСПЕКТЫ

Владимир Донченко

Abstract: The problems within mathematical theory of the Hough Transform are represented in the report including those to be the Clustering Problem in the Euclidean spaces along appropriate hyper planes and to be the arguments in the support of the its using in the "multi parametric estimation" variant. As regarding clustering problem explicit expressions for the grouping distances are represented. The Moore-Penrose Generalized Inverse is the base of the corresponding considerations.

Keywords: Hough Transform, clusterization, Moore – Penrose Generalized Inverse, Single Valued Decomposition (SVD), orthogonal projectors, Generalized Inverse for disturbed matrixes, clustering, clustering along hyper planes.

ACM Classification Keywords: G.3 Probability and statistics, G.1.6. Numerical analysis: Optimization; G.2.m. Discrete mathematics: miscellaneous.

Вступление

Преобразование Хока (ПГ (и) , [Hough,1962]), появившееся и интенсивно развивающееся как инженерный метод обработки изображений (см., например, [Xu , Oja, 1992]), утверждается [Donchenko, 2003] как равноправный метод описания неопределенности, наряду с детерминированным, статистическим, минимаксным (с гарантированной точностью), нечётким.

Что касается последнего из упомянутых методов: нечёткого, – то его место и возможности определяются, – и в значительной мере расширяются – статистической интерпретацией [Donchenko, 2006] и модификацией классического определения введением объекта нечеткости [там же]. Собственно статистическая интерпретация представляет собой статистическую модель классической нечёткости.

В работе [Duda R.O, Hart P.,1972] задача выделения прямых в рамках ПХ(и) была сформулирована как задача кластеризации. Она была сформулирована в виде задачи распределения элементов выборки, по подходящим гиперплоскостям заданной размерности. Определение необходимого количества гиперплоскостей и их описание являлось одним из элементов задачи.

Заметим, что, хотя само появление ПХ (и) было ответом на потребности практики в выделении нескольких прямых на контурном изображении, вопрос, касающийся обоснования возможности его использования в «мультипараметрическом» варианте: для «оценивания» нескольких параметров, представленных в выборке, оставался открытым. Отсутствие категорического ответа на вопрос о возможности использования ПХ (и) для «мультипараметрического оценивания» определило предложение [Risse,1989] использовать технику ПХ (и) рекуррентно. В соответствии с его подходом на каждом шагу такого применения ПХ (и) выделяется одна прямая: та, которая «наиболее выраженным образом» представлена на изображении; из выборки удаляются все точки, выделенной прямой, после чего процедура выделения «наиболее выраженной прямой» повторяется в применении к редуцированной выборке.

В предлагаемом докладе представлено обоснование возможности применения ПХ (и) в его статистической формализации в виде ПГ (с) (статистической модели преобразования Хока) для использования в режиме «мультипараметрического оценивания», по крайней мере,– для достаточно естественных распределений ошибок наблюдения, к которым относятся симметричные, выпуклые по Андерсону [Anderson,1955] распределения. К последним относятся, в частности, гауссовские распределения с нулевым мат ожиданием и симметричные с треугольной плотностью. Кроме того, предложен вариант решения задачи кластеризации по гиперплоскостям из уже упомянутой выше работы [Duda R.O, Hart P.,1972] в ослабленном варианте: когда не задаётся размерность гиперплоскостей, по которым необходимо провести кластеризацию. Полный вариант решения упомянутой задачи можно найти в работе [Кириченко, Донченко., 2007], а также в соответствующем докладе упомянутых авторов в материалах текущей серии конференций в Варне. Применение техники псевдообращения по Муру – Пенроузу ([Moore,1920], [Penrose,1955]) (см., например [Алберт, 1977]) позволяет явным образом описывать формируемые гиперплоскости, а также – явный вид расстояния до них. Заметим, что в аналогичных задачах классификации для статистических совокупностей на основе корреляционных матриц (см., например, [Варнік,1998]) соответствующие расстояния вычисляются приближённо в ходе выполнения подходящей рекуррентной вычислительной процедуры.

В работе приводится определение псевдообращения через SVD-представление матрицы, а также определение основных операторов, рассматриваемых в связи и определяемые через псевдообращение.

ПГ (с) в варианте «мультипараметрического» оценивания

Возможность использования ПГ (и) в «мультипараметрическом» варианте оценивания следует из базовых результатов математической теории ПГ (и) – из ПГ (с) - теории (см., например, [Donchenko, 2003]). Напомним, что в рамках ПГ (с), предметом анализа является выборка (последовательность наблюдений)

$s_i = s_{x_i} = (x_i, y_i), i = \overline{1, N}$, в которой компоненты каждой из пар связаны соотношением: $y_i = g_{\theta_i}(x_i) + \varepsilon_i, i = \overline{1, N}$, где $g_{\theta}, \theta \in \Theta$ – параметрическое семейство отображений между евклидовыми пространствами, $\varepsilon_i, i = \overline{1, N}$ – независимые ошибки наблюдений. Преобразование Хока выборки называется последовательность $L_i = L_{s_i} = \{ \theta \in \Theta : y_i = g_{\theta}(x_i) \}, i = \overline{1, N}$, каждый элемент которой называется преобразованием Хока соответствующего наблюдения. Основой анализа ПГ (с) является аккумуляторная функция (АФ) $A_N(\pi)$ или нормированная АФ (НАФ) $A_N^{(f)}(\pi)$ как функция множества, заданная на подходящей совокупности $\pi \in \Pi$ (совокупности множеств-зондов) подмножеств множества параметров Θ соотношениями, соответственно

$$A_N(\pi) = \sum_{i=1}^N \delta(\pi \cap L_i), A_N^{(f)}(\pi) = \frac{1}{N} \sum_{i=1}^N \delta(\pi \cap L_i).$$

Основой вывода о корректности использования ПГ (и) в «мультипараметрическом» варианте являются теоремы о предельном поведении нормированных значений НАФ при неограниченном увеличении объёма выборки и неограниченном уменьшении геометрических размеров множеств-зондов. В качестве множеств-зондов Π в рассматриваемой ситуации рассматривается либо совокупность всех замкнутых шаров $S_{\rho}(\theta) = \theta + \rho S$ радиуса $\rho > 0$ с центром в $\theta \in \Theta$, либо множества $\pi = V_{\rho}(\theta) = \theta + \rho V$ с замкнутым симметричным выпуклым множеством V единичного радиуса, содержащим шар ненулевого радиуса с центром в нуле. Такая нормировка может рассматриваться как для общего варианта размерности евклидовых пространств, так и, в частности, для скалярных значений функций параметрического семейства. Ниже приведены две теоремы именно для этого случая.

Теорема 1. Пусть для каждого $x \in X$ функции параметрического семейства со скалярными значениями являются непрерывно дифференцируемыми относительно θ в области определения Θ ; пусть также плотности распределений ошибок наблюдений $h_{x, \theta^{(0)}}(z), x \in X, \theta^{(0)} \in \Theta$ являются непрерывными по z и определяются только комбинацией аргумент-параметр, представленной в наблюдении; пусть в наблюдениях представлено конечное множество параметров $\Theta^{(0)} = \{ \theta_1^{(0)}, \dots, \theta_K^{(0)} \}$ та конечное множество значений аргументов $X^{(0)} = \{ x_1^{(0)}, \dots, x_K^{(0)} \}$ с предельными ненулевыми частотами $f_{k,q}^{(0)}, k = \overline{1, K}, q = \overline{1, Q}$ для соответствующих комбинаций аргумент-параметр в выборке. Тогда

$$\lim_{\rho \rightarrow 0} \rho^{-1} A_{\infty}(S_{\rho}(\theta)) = 2 \sum_{k=1}^K \sum_{q=1}^Q f_{kq}^{(0)} h_{kq} (g_{\theta}(x_q^{(0)}) - g_{\theta_k^{(0)}}(x_q^{(0)})) \| \text{grad}_{\theta} g_{\theta}(x_q^{(0)}) \|, \quad (1)$$

где $h_{kq} = h_{x_q^{(0)}, \theta_k^{(0)}}, k = \overline{1, K}, q = \overline{1, Q}$,

а A_{∞} - предельное значение НАФ при неограниченном увеличении объёма выборки.

Если множество S заменить на V , то соответствующий результат приобретает вид.

Теорема 2. В условиях и обозначениях теоремы 1

$$\lim_{\rho \rightarrow 0} \rho^{-1} A_{\infty}(S_{\rho}(\theta)) = 2 \sum_{k=1}^K \sum_{q=1}^Q \varphi_q f_{kq}^{(0)} h_{kq} (g_{\theta}(x_q^{(0)}) - g_{\theta_k^{(0)}}(x_q^{(0)})) \| \text{grad}_{\theta} g_{\theta}(x_q^{(0)}) \|, \quad (2)$$

где $\varphi_q : \varphi_q \in (0, 1), q = \overline{1, Q}$.

Несложно убедиться, что в (1), (2) множитель - плотность распределения для $\theta \in \Theta$ в ПГ (с) отвечает преобразованию Хока истинного (без ошибок) наблюдения, «размытого» плотностью ошибки наблюдения. Если плотность ошибки наблюдения имеет максимум в нуле, как это имеет место для симметричных, выпуклых по Андерсону [Anderson, 1955] распределений, то максимумы предельной функции будут отвечать максимумам для «безошибочной» НАФ. И, таким образом, ПХ можно использовать в «мультипараметрическом» режиме, поскольку в ПХ оценивание параметров, представленных в выборке осуществляется на основе максимумов АФ или НАФ.

Кластеризация по гиперплоскостям: постановка задачи

В задаче о кластеризации по гиперплоскостям каждый из векторов признаков $x(1), \dots, x(n)$ из пространства признаков R^m может принадлежать одной из двух – для простоты рассмотрений – гиперплоскостей $\Gamma(k) = x_k + L_k \subseteq R^m, k = 1, 2$ (x – смещение (L подпространство гиперплоскости)). Требуется выделить указанные плоскости и отнести каждый из векторов к той, к которой он принадлежит. Вариантом указанной задачи является такой, в которой дополнительно фиксируется размерность s : $s \leq m$ – для каждой из гиперплоскостей $\Gamma(k) = x_k + L_k, k = 1, 2$.

Эти гиперплоскости подлежат определению на основе внутренней структуры имеющегося набора векторов $x(1), \dots, x(n)$ с соответствующим их разбиением на две части:

$$x(i_1), \dots, x(i_{n_1}) \in \Gamma(1),$$

$$x(j_1), \dots, x(j_{n_2}) \in \Gamma(2):$$

$$\{i_1, \dots, i_{n_1}\} \cup \{j_1, \dots, j_{n_2}\} = \{1, 2, \dots, n\} \quad n_1 + n_2 = n.$$

Вариантом такой задачи является и задача Duda&Hart'a 1972 года. Задача группировки по гиперплоскостям является одним из вариантов оценивания нескольких зависимостей заданного параметрического семейства, представленных в выборке [Donchenko, 2003]. В такой постановке речь идёт о семействе аффинных преобразований между евклидовыми пространствами произвольных размерностей, когда определяющим является структура возможного пространства значений, а не представление оператора в том и ли ином базисе. Стандартным образом, предлагаемый к рассмотрению алгоритм кластеризации носит характер рекуррентной процедуры в ходе, которой после первично произвольного разбиения на две части происходит «рафинирование» исходных частей, в ходе которого осуществляется освобождение от «не отвечающих совокупности» элементов. Аппарат псевдообращения по Муру-Пенроузу позволяет получить явные формулы для расстояний от подходящих гиперплоскостей, которые отвечают тому или иному разбиению. Отметим, что указанные расстояния эффективно и явным видом описываются как в варианте, когда проверяемый элемент не исключается из совокупности (см. ниже), так и при его исключении [Кириченко, Донченко., 2007]. В последнем варианте процедура кластеризации аналогична статистической процедуре, которая называется Jack Knife (складной нож) [Эфрон, 1988]. Говоря об эффективности и адекватности применения аппарата псевдообращения для рассматриваемой задачи, отметим также, что формулы вычисления расстояний носят рекуррентный по исключаемым элементам вид. Достижение указанной рекуррентности обеспечивается теорией возмущения псевдообратных матриц [Кириченко, Лепеха, 1997].

Кластеризация по гиперплоскостям: вспомогательные определения и утверждения

Псевдообращение A^+ по Муру-Пенроузу для $m \times n$ матрицы A может определяться одним из нескольких эквивалентных способов, среди которых отметим определение через сингулярное представление матриц (SVD-разложение), когда псевдообращение определяется соотношением

$A^+ = \sum_{i=1}^r x_i y_i^T \lambda_i^{-1}$, которое определяется элементами SVD-представления исходной матрицы:

$A = \sum_{i=1}^r y_i x_i^T \lambda_i$, в котором: $\lambda_1^2 \geq \dots \lambda_r^2 > 0$ – общий набор ненулевых собственных чисел;

$y_i, i = \overline{1, r}$ и $x_i, i = \overline{1, r}$ – ортонормированные наборы собственных векторов матриц $P(A), P(A^T)$, $A^T A$: соответственно, а $r = \text{rank } A = \text{rank } A^T$.

Псевдообращение позволяет в явном виде выписать две пары ортогональных проекторов (ОП) $P(A), P(A^T)$: $P(A) = A^+ A, P(A^T) = A^T A^+ = A A^+$ и $Z(A) = E_n - P(A), Z(A^T) = E_m - P(A^T)$ (P- и Z-проекторы соответственно). Первая пара представляет собой ОП на подпространства L_{A^T}, L_A являющиеся множествами значений A^T, A соответственно, вторая – на ортогональные к ним подпространства соответственно $L_{A^T}^\perp, L_A^\perp$. Очевидно образом, $L_{A^T}^\perp = \text{Ker } A$. Заметим также, что каждое из подпространств L_{A^T}, L_A является линейной оболочкой соответственно векторов-столбцов и векторов-строк матрицы A .

Не ограничивая общности, будем считать, что $x(1), \dots, x(n)$ из пространства признаков R^m могут принадлежать одной из двух – для простоты рассмотрений – гиперплоскостей $\Gamma(k) = x_k + L_k \subseteq R^m, k = 1, 2$ (x – смещение, L подпространство гиперплоскости). Подпространства могут задаваться, как множества значений подходящих матриц $A(k), k = 1, 2$. В этом случае будут использоваться обозначения как $L_k = L_{A(k)}, k = 1, 2$.

Лемма. Пусть подпространства гиперплоскостей являются множествами значений матриц $A(k), k = 1, 2$ соответственно. Тогда расстояния соответствия $\rho(x, \Gamma(k)), k = 1, 2$ произвольного вектора $x \in R^m$ до каждой из двух гиперплоскостей $\Gamma(k), k = 1, 2$ определяются соотношением

$$\rho(x, \Gamma_k) = (x - x_k)^T Z(A^T(k))(x - x_k), k = 1, 2$$

Кластеризация по гиперплоскостям: основной алгоритм

Сам алгоритм кластеризации по гиперплоскостям состоит в выполнении следующих шагов.

1. Первичное разбиение на две совокупности произвольным образом.
2. Построение смещений $x_k, k = 1, 2$. Соответствующие смещения могут быть реализованы или как средние по векторам каждой из отобранных совокупностей или в виде произвольных представителей каждой из них.
3. Построение матриц каждой из совокупностей $A(k), k = 1, 2$ из векторов столбцов каждой из групп, центрированных соответствующими векторами смещений, построенных на шаге 2.
4. Опребделение гиперплоскостей $\Gamma(k), k = 1, 2$ как таких, которые задаются смещениями, вычисленными на шаге 2, и подпространствами $L_k = L_{A(k)}, k = 1, 2$
5. Вычисление расстояний элементов каждой из совокупностей до каждой из двух построенных
6. гиперплоскостей с использованием результатов леммы 2.
7. Перераспределение векторов между совокупностями и повторением рекуррентных шагов.

Перераспределение элементов можно реализовывать разными способами с введением подходящих параметров алгоритма

Литература

- [Anderson, 1955] Anderson T.W. The integral of a symmetric random functions over symmetric convex sets // Proc. Amer. Math. Soc. – № 6.–1955.–P.170-175.
- [Donchenko, 2003] Donchenko V.S. Hough Transform and Uncertainty.//Proceedings International Conference "Knowledge Dialog – Solution". – V.–June 16-23, 2003.–Varna (Bulgaria). – P.391-395.
- [Donchenko, 2003] Donchenko V.S. Hough Transform and Uncertainty.//Proceedings International Conference "Knowledge Dialog – Solution". – V.–June 16-23, 2003.–Varna (Bulgaria). – P.391-395.
- [Donchenko, 2006] Donchenko V.S. Fuzzy sets: abstraction axiom, statistical interpretation, observation of fuzzy sets. // International Journal on Information Theory and Applications.–Vol.13, №3– 2006(Bulgaria). – 233-238.
- [Duda R.O, Hart P., 1972] Duda R.O, Hart P. Using the Hough Transform to detect Lines and Curves in pictures.//Communications of ACM.–v.15.–1972.–p.11-15.
- [Hough, 1962] Hough P.V.C. Method and Means for Recognizing Complex Patterns. - U.S. Patent 3069354, 1962.
- [Moore, 1920] Moore E.H. On the reciprocal of the general algebraic matrix.//Bull. Amer. Math. Soc. 26, 1920. p. 394-395.
- [Penrose, 1955] Penrose R. A generalized inverse for matrices.// Proc. Camb. Philosophical Soc. - 51, 1955. – p.406-413.
- [Risse, 1989] Risse T. Hough Transformation for line recognition: complexity for evidence accumulation and cluster detection.// Computer Vision, Graphics and Image Processing. – Vol.46 –1989.–p.327-345.
- [Xu, Oja, 1992] Xu L., Oja E. Further Developments on RHT: Basic Mechanisms, Algorithms, and Computational Complexities.// ICPR (92): Proceedings, International Conference on Pattern Recognition. – Vol.1. –1992.– P.125-132.
- [Алберт, 1977] Алберт А. Регрессия, псевдо инверсия, рекуррентное оценивание. –М.: Наука, 1977.–305 с.
- [Кириченко, Донченко., 2007] Кириченко Н.Ф., Донченко. В.С. Псевдообращение в задачах кластеризации.// Киб. и СА.- №4, 2007– С.98-122.

Информация об авторе

Владимир С. Донченко – Киевский национальный университет имени Тараса Шевченко, факультет кибернетики, профессор, Украина, e-mail:voldon@unicyb.kiev.ua

АЛГЕБРАИЧЕСКИЙ JACK KNIFE: КЛАСТЕРИЗАЦИЯ ПО ГИПЕРПЛОСКОСТЯМ

Николай Кириченко, Владимир Донченко

Abstract: The Clustering Problem, when the clustering is carried out along appropriate hyper planes, was investigated. Different variants of the "correspondence distances" have been proposed and investigated including the algebraic Jack Knife one. Efficacy, constructivism and explicit representation for the objects under investigation are provided with the Generalized Inverse Technique including the Generalized Inverse Disturbances. The approach represented is the one to be the variant of the Hough Transform in its mathematical. The Collection of important for the applications results regarding Generalized Inverse is also represented in the paper.

Keywords: clusterization, Hough Transform, Moore – Penrose Generalized Inverse, Single Valued Decomposition (SVD), orthogonal projectors, Generalized Inverse for disturbed matrixes, clustering, clustering along hyper planes.

ACM Classification Keywords: G.3 Probability and statistics, G.1.6. Numerical analysis: Optimization; G.2.m. Discrete mathematics: miscellaneous.

Вступление

Статья посвящена применению методов псевдо обращения (см. например, [Алберт, 1977]) и, в особенности их развития в виде теории возмущения псевдо обращения [Кириченко, Лепеха, 1997, 2002], в решении задач кластеризации. Задача кластеризации в такой постановке может рассматриваться как вариант преобразования Хока [Donchenko, 2003]. При таком подходе определяющим является рассмотрение расстояний не до ограниченных множеств, которые отвечают элементам, отнесённым к тому или иному классу, а до подходящих гиперплоскостей, которые ими порождаются. Отметим, что в статистических подходах (см., например [Varpiк, 1998]) гиперплоскости связываются с корреляционными матрицами подходящих распределений вероятностей в пространстве признаков. Вычислительные алгоритмы для расстояния от гиперплоскостей использовались, например, в работе [Haykin, 1999]. В предлагаемой работе применение аппарата псевдо обращения позволяет выписать явные формулы «расстояний соответствия» в том числе и в варианте алгебраического Jack Knife'a [Эфрон, 1988].

Важными в применении техники псевдообращения являются эквивалентные варианты определения псевдо обращения, прямые [Алберт, 1977] и обратные [Кириченко, Лепеха, 1997, 2002], формулы Гревилля, формулы псевдообращения для замены строки или столбца матрицы [Кириченко, Лепеха, 1997, 2002], [Кириченко, Донченко, 2005], а также формулы возмущения для Z - и R -операторов [Кириченко, Донченко, 2007]. В первой части предлагаемой работы приводится подборка некоторых таких результатов.

Вторая часть посвящена постановке и решению задачи разбиения наличной совокупности векторов на два кластера. Предложены различные варианты определения и вычисления «расстояний соответствия», которые имеют естественную геометрическую интерпретацию и эффективно описываются в терминах псевдо обращения. Явный вид формул, определяющих расстояния соответствия позволяет построить эффективные алгоритмы кластеризации.

Постановка задачи

В задаче о кластеризации по гиперплоскостям каждый из векторов признаков $x(1), \dots, x(n)$ из пространства признаков R^m может принадлежать одной из двух – для простоты рассмотрений – гиперплоскостей $\Gamma(k) = x_k + L_k \subseteq R^m, k = 1, 2$ (x – смещение, L подпространство гиперплоскости). Требуется выделить указанные плоскости и отнести каждый из векторов к той, к которой он принадлежит. Вариантом указанной задачи является такой, в которой дополнительно фиксируется размерность $s: s < m$ – для каждой из гиперплоскостей $\Gamma(k) = x_k + L_k, k = 1, 2$.

Эти гиперплоскости подлежат определению на основе внутренней структуры имеющегося набора векторов $x(1), \dots, x(n)$ с соответствующим их разбиением на две части:

$$x(i_1), \dots, x(i_{n_1}) \in \Gamma(1),$$

$$x(j_1), \dots, x(j_{n_2}) \in \Gamma(2):$$

$$\{i_1, \dots, i_{n_1}\} \cup \{j_1, \dots, j_{n_2}\} = \{1, 2, \dots, n\} \quad n_1 + n_2 = n.$$

Вариантом такой задачи является и задача Duda&Hart'a 1972 года. Задача группировки по гиперплоскостям является одним из вариантов оценивания нескольких зависимостей заданного параметрического семейства, представленных в выборке [Donchenko, 2003]. В такой постановке речь идёт о семействе аффинных преобразований между евклидовыми пространствами произвольных размерностей, когда определяющим является структура возможного пространства значений, а не представление оператора в том и ли ином базисе. Стандартным образом, предлагаемый к рассмотрению алгоритм кластеризации носит характер рекуррентной процедуры в ходе, которой после первично произвольного разбиения на две части происходит «рафинирование» исходных частей, в ходе которого осуществляется

освобождение от «не отвечающих совокупности» элементов. Аппарат псевдообращения по Муру-Пенроузу позволяет получить явные формулы для расстояний от подходящих гиперплоскостей, которые отвечают тому или иному разбиению. Отметим, что указанные расстояния эффективно и явным видом описываются как в варианте, когда проверяемый элемент не исключается из совокупности, формирующей гиперплоскость, так и при его исключении. В последнем варианте процедура кластеризации аналогична статистической процедуре, которая называется Jack Knife (складной нож) [Эфрон, 1988]. Говоря об эффективности и адекватности применения аппарата псевдообращения для рассматриваемой задачи, отметим также, что формулы вычисления расстояний носят рекуррентный по исключаемым элементам вид. Достижение указанной рекуррентности обеспечивается теорией возмущения псевдообратных матриц [Кириченко, Лепеха, 1997].

Вспомогательные определения и утверждения

Псевдообращение A^+ по Муру-Пенроузу для $m \times n$ матрицы A может определяться одним из нескольких эквивалентных способов, среди которых отметим определение через сингулярное представление матриц (SVD-разложение), когда псевдообращение определяется соотношением:

$A^+ = \sum_{i=1}^r x_i y_i^T \lambda_i^{-1}$, которое определяется элементами SVD-представления исходной матрицы:

$A = \sum_{i=1}^r y_i x_i^T \lambda_i$, в котором: $\lambda_1^2 \geq \dots \lambda_r^2 > 0$ – общий набор ненулевых собственных чисел;

$y_i, i = \overline{1, r}$ и $x_i, i = \overline{1, r}$ – ортонормированные наборы собственных векторов матриц $P(A), P(A^T)$, $A^T A$: соответственно, а $r = \text{rank } A = \text{rank } A^T$.

Псевдообращение позволяет в явном виде выписать две пары ортогональных проекторов (ОП) $P(A), P(A^T)$: $P(A) = A^+ A, P(A^T) = A^T A^+ = A A^+$ и $Z(A) = E_n - P(A), Z(A^T) = E_m - P(A^T)$ (P- и Z- проекторы соответственно). Первая пара представляет собой ОП на подпространства L_{A^T}, L_A являющиеся множествами значений A^T, A соответственно, вторая – на ортогональные к ним подпространства соответственно $L_{A^T}^\perp, L_A^\perp$. Очевидны образом, $L_{A^T}^\perp = \text{Ker } A$. Заметим также, что каждое из подпространств L_{A^T}, L_A является линейной оболочкой соответственно векторов-столбцов и векторов-строк матрицы A .

Важными в связи с рекуррентными формулами псевдообращения: формулами позволяющими записывать соответствующий оператор при добавлении или вычёркивании строки или столбца матрицы, – являются также \square -операторы, определяемые соотношениями: $R(A) = A^+ A^{T+}, R(A^T) = A^{+T} A^+$. К таким формулам псевдообращения относятся прямые [Алберт, 1977] и обратные формулы Гревилля (Greville) [Кириченко, Лепеха, 1997]. Такой же характер имеют формулы возмущения псевдообратных матриц [Кириченко, Лепеха, 1997], дающие представление $A^+(a, b) = (A + ab)^T$ через A, A^+, a, b , [там же], а также [Кириченко, Донченко, 2005]. В последней работе также приведены рекуррентные формулы для P-, Z- и R- операторов, описывающие их вид при замене строки или столбца матрицы, для которой они рассматриваются. Формулы, описывающие вид соответствующих операторов при возмущении матрицы, представлены следующей леммой [Кириченко, Донченко., 2007].

Лемма 1.

1. Для векторов a и b^T линейно не зависящих от, соответственно, столбцов и строк матрицы A , т.е. при выполнении условий $a^T Z(A^T) a > 0, b^T Z(A) b > 0$, справедливы следующие соотношения:

$$Z(A + ab^T) = Z(A) + \frac{Z(A)bb^T Z(A)}{b^T Z(A)b};$$

$$Z((A + ab^T)^T) = Z(A^T + ba^T) = Z(A^T) + \frac{Z(A^T)aa^T Z(A^T)}{a^T Z(A^T)a};$$

$$R(A + ab^T) = R(A) - R(A) \frac{bb^T Z(A)}{b^T Z(A)b} - \frac{Z(A)bb^T}{b^T Z(A)b} R(A) - cA^+ ab^T Z(A) - cZ(A)ba^T A^{+T} + \\ + \frac{A^+ aa^T A^{+T}}{a^T Z(A^T)a} + \frac{b^T R(A)ba^T Z(A^T)a + (1 + b^T A^+ a)^2}{a^T Z(A^T)a [b^T Z(A)b]^2} Z(A)bb^T Z(A),$$

$$\text{где } c = \frac{1 + b^T A^+ a}{a^T Z(A^T)a b^T Z(A)b}.$$

2. Для вектора a линейно зависимо от столбцов матрицы A , а вектора b^T – линейно не зависимо от строк матрицы таким образом, что, – для упрощения представления результата, – $b \perp L_{A^T}$, т.е. при выполнении условий $a^T Z(A^T)a = 0, b^T Z(A)b = \|b\|^2$, справедливы соотношения:

$$Z(A + ab^T) = Z(A) + \frac{k_{A,a,b} k_{A,a,b}^T}{\|k_{A,a,b}\|^2} \frac{bb^T}{\|b\|^2},$$

где:

$$k_{A,a,b} = A^+ a \frac{b}{\|b\|^2},$$

$$Z((A + ab^T)^T) = Z(A^T + ba^T),$$

$$R(A + ab^T) = I_n - \frac{kk^T}{\|k\|^2} R(A) I_n - \frac{kk^T}{\|k\|^2}.$$

3. Для векторов a и b^T одновременно линейно зависимых от соответственно столбцов и строк матрицы A , при условии падения ранга возмущённой матрицы: $\text{rank}(A + ab^T) = \text{rank} A - 1$, т.е. при выполнении условий: $a^T Z(A^T)a = 0, b^T Z(A)b = 0, b^T A^+ a = 1$, справедливы следующие соотношения:

$$Z(A + ab^T) = Z(A) + \frac{A^+ a a^T (A^+)^T}{a^T R(A^T)a},$$

$$Z((A + ab^T)^T) = Z(A^T + ba^T) = Z(A) + \frac{(A^+)^T bb^T A^+}{b^T R(A)b},$$

$$R(A + ab^T) = A^+(a,b)A^{+T}(a,b),$$

$$\text{где: } A^+(a,b) = A^+ \frac{A^+ aa^T R(A^T)}{a^T R(A^T)a} - \frac{R(A)bb^T A^+}{b^T R(A)b} + cA^+ ab^T A^+, \quad c = \frac{b^T R(A)A^+ a}{a^T R(A^T)a b^T R(A)b}.$$

4. Для векторов a и b^T одновременно линейно зависимых от, соответственно, столбцов и строк матрицы A , но при условии неизменности ранга возмущённой матрицы по сравнению с рангом A , т.е. при выполнении условий

$$a^T Z(A^T) a = 0, b^T Z(A) b = 0, b^T A^+ a \neq 1,$$

справедливы следующие соотношения:

$$Z(A + ab^T) = Z(A), Z((A + ab^T)^T) = Z(A^T + ba^T) = Z(A^T),$$

$$R(A + ab^T) = R(A) \frac{A^+ ab^T R(A)}{1 + b^T A^+ a} \frac{R(A) ba^T A^{+T}}{1 + b^T A^+ a} + \frac{b^T R(A) b}{1 + b^T A^+ a} A^+ aa^T A^{+T}.$$

Лемма 2.

Пусть подпространства гиперплоскостей являются множествами значений матриц $A(k), k = 1, 2$ соответственно. Тогда расстояния соответствия $\rho(x, \Gamma(k)), k = 1, 2$ произвольного вектора $x \in R^m$ до каждой из двух гиперплоскостей $\Gamma(k), k = 1, 2$ определяются соотношением:

$$\rho(x, \Gamma_k) = (x \quad x_k)^T Z(U_s^T(k))(x \quad x_k), k = 1, 2, \quad (1)$$

$$\text{Где } U_s(k) = \begin{cases} \sum_{i=1}^r y_i(k) x_i^T(k) \lambda_i(k) & \text{разм. } s \text{ не задана} \\ \sum_{i=1}^s y_i(k) x_i^T(k) \lambda_i(k) & \text{разм. } s \text{ задана} \end{cases}, r = 1, 2.$$

Кластеризация по гиперплоскостям

Сам алгоритм кластеризации по гиперплоскостям состоит в выполнении следующих шагов.

1. Первичное разбиение на две совокупности произвольным образом.
2. Построение смещений $x_k, k = 1, 2$. Соответствующие смещения могут быть реализованы или как средние по векторам каждой из отобранных совокупностей или в виде произвольных представителей каждой из них.
3. Построение матриц каждой из совокупностей $A(k), k = 1, 2$ из векторов столбцов каждой из групп, центрированных соответствующими векторами смещений, построенных на шаге 2.
4. Определение гиперплоскостей $\Gamma(k), k = 1, 2$ как таких, которые задаются смещениями, вычисленными на шаге 2, и подпространствами $L_k = L_{A(k)}, k = 1, 2$
5. Вычисление расстояний элементов каждой из совокупностей до каждой из двух построенных гиперплоскостей с использованием результатов леммы 2.
6. Перераспределение векторов между совокупностями и повторением рекуррентных шагов.

Перераспределение элементов можно реализовывать разными способами с введением подходящих параметров алгоритма

Кластеризация по гиперплоскостям – модификация расстояний

Можно показать, что расстояния до гиперплоскостей леммы 2, являются вариантом квадратичной формы с весовыми коэффициентами, $\lambda_i^2, i = 1, r$. Ниже приводится вариант расстояний до гиперплоскостей, реализующий идею взвешивания следующим образом. Модифицированные расстояния ρ_R в этом случае определяются соотношением:

$$\rho_R(x, \Gamma(k)) = \frac{1}{\text{tr}R(A^T(k)A(k))} (x \quad x_k)^T R(A^T(k))(x \quad x_k), k = 1, 2. \quad (2)$$

Кластеризация по гиперплоскостям □ алгебраический Jack Knife

При проверке элементов совокупностей на соответствие вычислением расстояний по формулам (1) или (2) тестируемые элементы принимают участие в формировании гиперплоскостей, представляющих кластеры. Поэтому целесообразной является такая процедура проверки соответствия, при которой тестируемый элемент кластера, исключается из числа объектов, которые определяют исследуемый кластер. В статистике такая процедура исключения носит название "Jack Knife" (складной нож) [Эфрон, 1988]. Поэтому процедуру тестирования на принадлежность кластеру с исключением тестируемых элементов из описания кластера будем называть алгебраическим Jack Knife'ом.

Заметим, что естественным является вариант кластеризации, когда исключение элемента приводит к падению ранга матрицы $A(k), k = 1, 2$. Псевдообращение даёт конструктивную явную формулу проверки соответствующего условия.

Исключение тестируемых элементов из кластера изменяет как сдвиг (центр кластера), так и линейное подпространство кластера. Формулы (1),(2) при таком исключении, очевидным образом, переписываются в виде, для изменённых смещений (будем считать их средними) и изменённых матриц: $x_k^{(0)}, A^{(0)}(k), k = 1, 2$ соответственно.

Теория возмущения псевдообратных матриц [Кириченко, Лепеха, 1997]. Даёт возможность эффективной организации процедуры «отсеивания», в которой критерий замены строится на основе леммы 1 и имеет вид, определяемый следующей теоремой.

Теорема. В условиях падения ранга

$$\rho(x_j(k), \Gamma_j^{(0)}(k)) = \frac{n_k^2}{\|E_m\| \frac{q_j(k)q_j^T(k)}{\|q_j(k)\|^2} \sum_{l \neq j} q_l(k)\|^2}, j = \overline{1, n_k}, k = 1, 2,$$

Где $x_j(k), \Gamma_j^{(0)}(k) j = \overline{1, n_k}, k = 1, 2$ – исключаемые элементы каждой из совокупностей и гиперплоскости, отвечающие «усечённым» совокупностям, а $q_j(k), j = \overline{1, n_k}, k = 1, 2$ столбцы с номером $j, j = \overline{1, n_k}$ в каждой из матриц $A^+(k), k = 1, 2$.

Литература

- [Donchenko, 2003] Donchenko V.S. Hough Transform and Uncertainty.//Proceedings:X International Conference "Knowledge Dialog – Solution". – V. – June 16-23, 2003. – Varna(Bulgaria). – P.391-395.
- [Найкин, 1999] Neural networks. A comprehensive Foundation. – New Jersey: Prentice Hall 07458. – 1999. – 842 p.
- [Moore, 1920] Moore E.H. On the reciprocal of the general algebraic matrix.//Bull. Amer. Math. Soc. – 26, 1920. – p. 394-395.
- [Penrose, 1955] Penrose R. A generalized inverse for matrices.// Proc. Camb. Philosophical Soc.- 51, 1955.– p.406-413.
- [Vapnik, 1998] Vapnik V.N. Statistical Learning Theory.–New York: Wiley.– 1998.
- [Алберт, 1977] Алберт А. Регрессия, псевдо инверсия, рекуррентное оценивание. □ М.: Наука, 1977.–305 с.
- [Кириченко, Донченко, 2005] Кириченко Н.Ф., Донченко В.С., Задача термінального спостереження динамічних системи: множинність розв'язків та оптимізація. //Ж. Обч. та пр. мат. – Вип.3, 2005, с. 63-78.
- [Кириченко, Донченко., 2007] Кириченко Н.Ф., Донченко. В.С. Псевдообращение в задачах кластеризации.// Киб. и СА.- №4, 2007– С.98-122.
- [Кириченко, Лепеха, 1997]. Кириченко Н.Ф., Лепеха Н.П. Псевдо обратные и проекционные матрицы в применении к исследованию задач управления, наблюдения и идентификации.//Киб. и СА.- №2, 1997– С.98-122.
- [Кириченко, Лепеха., 2002] Кириченко Н.Ф., Лепеха Н.П. Псевдо обратные и проекционные матрицы в применении к исследованию задач управления, наблюдения и идентификации.// Киб. и СА.- №2, 1997– С.98-122.
- [Эфрон, 1988] Эфрон Б. Нетрадиционные методы многомерного статистического анализа. – М.: Фин. и стат., 1988. – 263 с.

Информация об авторах

Николай Ф. Кириченко – Институт кибернетики им.В.М.Глушкова НАН Украины, ведущий научный сотрудник, профессор.

Владимир С. Донченко – Киевский национальный университет имени Тараса Шевченко, факультет кибернетики, профессор, Украина, e-mail:voldon@unicyb.kiev.ua

МЕТАЭВРИСТИЧЕСКИЙ МЕТОД ДЕФОРМАЦИЙ ДЛЯ РЕШЕНИЯ ЗАДАЧ КОМБИНАТОРНОЙ ОПТИМИЗАЦИИ

Леонид Гуляницкий

Аннотация: Исследуются проблемы разработки приближенных алгоритмов комбинаторной оптимизации. Приводится формализация понятия комбинаторных объектов, позволяющая строго формально определить понятие задачи комбинаторной оптимизации. Предлагается метаэвристический алгоритм комбинаторной оптимизации, названный *H*-методом, метаэвристика которого использует аналогии с известным в недифференцируемой непрерывной оптимизации методом Нелдера-Мида. В качестве встроенной процедуры использован *G*-алгоритм – один из алгоритмов стохастического локального поиска. Обсуждаются вопросы практической реализации алгоритмов предложенного *H*-метода. Приводятся результаты вычислительного эксперимента по решению квадратичных задач о назначениях из известной библиотеки.

Ключевые слова: комбинаторная оптимизация, метаэвристики, популяционные методы, стохастический локальный поиск, квадратичная задача о назначении.

Введение

Среди прикладных проблем моделирования видное место занимают задачи выбора оптимальных решений в дискретных (в частности, конечных) пространствах, чем обуславливаются потребности в исследовании и внедрении моделей и методов комбинаторной оптимизации (КО). Для решения возникающих сложных задач КО, а также для решения задач повышенной размерности необходима разработка новых общих вычислительных схем (в том числе с распараллеливанием вычислений) и исследование их эффективности. Поскольку большинство практически важных задач КО относится к числу *NP*-трудных, основное внимание уделяется разработке и исследованию приближенных алгоритмов. Прогресс в повышении производительности вычислительной техники привел в последние годы к появлению так называемых метаэвристических (гибридных) методов оптимизации, применение которых направлено на повышение точности решения задач КО [1,2].

Метаэвристические методы основаны на использовании двух техник: общая схема строится на базовом методе (метаэвристике), в которую включается та или иная встроенная процедура. Большинство разработанных метаэвристических методов относятся к классу популяционных алгоритмов [2], т.е. алгоритмов, в которых, в отличие от траекторных, на каждой итерации обрабатывается не один, а сразу несколько вариантов решения.

В работе [3] предложен алгоритм комбинаторной оптимизации, предназначенный для использования как на традиционных ЭВМ, так и многопроцессорных вычислительных комплексах (МВК), который назван методом деформированных многогранников (МДМ). В его вычислительной схеме осуществляется

глобальное сканирование пространства решений решаемой задачи, чем повышается вероятность нахождения более точных вариантов решения.

Ниже излагается метаэвристический метод КО, названный *H*-методом, который в качестве метаэвристики использует схему МДМ, а встроенная процедура основана на одной из модификаций алгоритмов стохастического локального поиска.

Общая схема *H*-метода

В настоящее время общепринятым является определение задачи КО, предложенное Пападимитриу и Стайглицем [1,2,4]: необходимо найти $x^* \in X$ такое, что

$$x_* = \arg \min_{x \in D \subseteq X} f(x), \quad (1)$$

где X – конечное (или, возможно, счетное бесконечное) пространство решений задачи, D – его подпространство, определяемое ограничениями задачи, $f: X \rightarrow \mathbb{R}^1$ – заданная целевая функция задачи.

Такое определение однозначно относит к задачам КО все проблемы оптимизации на конечных множествах, однако в случае бесконечных пространств не позволяет четко классифицировать задачи оптимизации по структуре элементов пространства решений. Берж [5] предложил формализовать понятие комбинаторной конфигурации следующим образом: пусть имеем m, n – натуральные, а также два множества $U = \{1, \dots, m\}$, $V = \{v_1, \dots, v_n\}$, причем на V задан некий строгий порядок $v_1 < \dots < v_n$ (т.е. V – цепь).

Определение 1. Комбинаторной конфигурацией называется отображение $\varphi: U \rightarrow V$, которое удовлетворяет некоторому комплексу ограничений Λ .

Из определения следует, что при фиксированных m и n число комбинаторных конфигураций конечно. Выбор ограничений в Λ позволяет описывать различные комбинаторные конфигурации.

Предлагается следующее обобщение схемы Бержа. Пусть заданы $Y = \{1, \dots, m\}$, Z – дискретное, в частности, конечное пространство (назовем его образующим), φ – гомоморфизм, $\varphi: Y \rightarrow Z$, удовлетворяющий некоторой системе ограничений Ω . Напомним, что под дискретным пространством понимается множество, состоящее из изолированных точек.

Определение 2. Под комбинаторным объектом κ будем понимать триаду $\kappa = (\varphi, \tilde{X}, \Omega)$, где \tilde{X} – базовое пространство.

Определение 3. Назовем комбинаторными объектами 1-го порядка такие комбинаторные объекты, у которых базовое пространство совпадает с образующим:

$$\kappa = (\varphi, X_{(1)}, \Omega),$$

где $X_{(1)} \equiv Z$.

Нетрудно убедиться, что если Z – это конечная цепь, то такие комбинаторные объекты совпадают с комбинаторными конфигурациями в смысле Бержа [5].

Определение 4. Комбинаторными объектами k -го порядка ($k > 1$) назовем комбинаторные объекты

$$\kappa = (\varphi, X_{(k)}, \Omega),$$

где $X_{(k)} \subseteq X_{(k-1)} \cup X^k$.

Возвращаясь к оптимизационной задаче (1), дадим следующее

Определение 5. Задача (1) называется задачей КО, если пространство ее решений X – это пространство, элементами которого являются комбинаторные объекты.

В ряде развитых алгоритмов КО для избежания концентрации поиска в ограниченной подобласти пространства решений задачи X и повышения точности получаемых решений используются процедуры

возмущения (как в повторяющемся локальном поиске [6]) или скрещивания и мутации, как в генетических алгоритмах (ГА) или меметических алгоритмах (МА) [1]. Заметим, что подобные процедуры порождают подмножества вариантов решения, которые не согласованы с топологией пространства X . В то же время, примером подобного согласования является упомянутый МДМ. Осуществляемое в нем использование специальных отрезков дает возможность синтезировать поиск в окрестностях и глобальное сканирование пространства решений X , причем процедура сканирования, в отличие от общих операторов возмущения или рекомбинации в большинстве других метаэвристических методов, определена конкретно.

На основе синтеза идей МДМ и G-алгоритмов [7] (а также подхода на основе популяций, который был использован, например, при разработке ГА и фронтальных алгоритмов [8]), предлагается схема гибридного алгоритма, названного H -методом [9].

В алгоритмах МДМ существенную роль играет понятие d -отрезка $/x, y/$, соединяющего произвольные две точки $x, y \in X$, $X=(X, d)$ – метрическое пространство с метрикой d [8].

Определение 6. Назовем d -отрезком $/x, y/$ упорядоченную совокупность точек $x_i \in X, i=1, \dots, k$, которые удовлетворяют условию: $d(x, x_i) + d(x_i, y) = d(x, y)$ для всех $i=1, \dots, k$, причем $x_1 = x, x_k = y$, а $d(x, x_i) < d(x, x_{i+1}), i=1, \dots, k-1$; при этом не существует точки $z \in X$ такой, что $d(x_i, z) + d(z, x_{i+1}) = d(x, x_{i+1}), z \neq x_i, z \neq x_{i+1}, i=1, \dots, k-1$.

Определение 7. d -интервалом $\langle x, y \rangle$ назовем упорядоченную совокупность $/x, y/ \setminus \{x, y\}$.

Дальше будем рассматривать такие пространства вариантов решения X , для которых $d(x, y) = b \cdot h$ (b – натуральное, $h > 0$), причем если $d(x, y) > h$, то интервал $\langle x, y \rangle \neq \emptyset$. Без ущерба для общности положим, что $D \equiv X$, т.е. X будет играть роль пространства допустимых решений.

Вычислительная схема H -метода в терминах эволюционных вычислений представлена на рис., где **G_Search** – процедура G-алгоритма [7], осуществляющего поиск локального оптимума (не исключается возможность использования и другого метода поиска локально оптимального типа).

```

procedure H (x);
begin
  h := 0; P0 := ∅;
  for i:=1 to m do
    x := некоторый начальный вариант решения;
    x := G_Search (x);
    P0 := P0 ∪ x;
  end for; {сформирована начальная популяция P0 }
  repeat
    P := Ph;
    for i:= 1 to k do
      ОтборДляВариации (x, y ∈ P);
      ПостроениеПолуинтервала <x, x~/ : y ∈ <x, x~/ & f(x) > f(y);
      z := arg min {f(u) : u ∈ <x, x~/ \ Ls(y), y ∈ <x, x~/ };
      z := G_Search (z);
      P := P ∪ z;
    end for; {сформирована временная популяция из m+k точек}
    for i:=1 to 1 do
      ОтборДляМутации (x ∈ P);
      z := Мутация (x, предыстория);
      z := G_Search (z);
      P := P ∪ z;
    end for; {сформирована временная популяция из m+k+1 точек}
    Ph+1 := ОтборПопуляции (P);
    h := h+1;
  until не выполняется условие завершения;
  x := arg min {f(u) : u ∈ P};
  return x;
end

```

Рис. Схема H -метода

Основные параметры H -метода:

m – число особей в популяции (в терминах эволюционных вычислений);

k – количество пар особей, выбираемых для исследования (проведения полуинтервалов от текущей точки до максимально удаленной и поиска минимума на этих прямых);

l – количество особей, которые подлежат мутации;

s – радиус метрической окрестности $L_s(y)$, точки которой, находящиеся на построенном полуинтервале, исключаются из рассмотрения (в большинстве случаев s – натуральное);

x^∞ – точка, максимально удаленная в пространстве вариантов решений от исходной точки x (диаметрально противоположная в пространстве X точка).

Кроме численности популяции m , другие параметры могут изменяться от итерации к итерации, то есть определяться динамически с учетом хода процесса поиска экстремумов.

Как и в алгоритме повторяющегося локального поиска или МА, предлагаемый метод оперирует с локальными экстремумами. Их множество P играет роль, аналогичную популяции в ГА или МА – поэтому использованные в нем три процедуры отбора могут реализовываться по аналогии с эволюционными алгоритмами. Принципиальное отличие алгоритма H -метода – глобальный характер поиска в пространстве решений X путем нахождения субоптимального решения исходной задачи на основе решения подзадачи вида:

$$z = \arg \min_{u \in \langle x, x^\infty \rangle \setminus L_s(y)} f(u), \quad (2)$$

где $u \in \langle x, x^\infty \rangle$.

Вопросы реализации ключевых аспектов вычислительной схемы

При практическом применении алгоритмов H -метода было отмечено, что поскольку выбранные точки x и y являются локальными экстремумами, то часто может возникать ситуация, когда $z=y$. Во избежание этого и вводится параметр $s > 0$ – значение радиуса метрической окрестности $L_s(y)$, точки которой исключаются из рассмотрения в подзадачах (2). Понятно, что для большинства комбинаторных пространств целесообразно выбирать $s \geq 1$, исключая тем самым возможность определения точки y в качестве субоптимального решения в указанной подзадаче.

Как и в алгоритмах локального поиска, при решении подзадач (2) можно избежать повторного вычисления значений целевой функции: стоит вычислять лишь разность $\Delta = f(x) - f(y)$, учитывая тот факт, что отрезок составляют лишь соседние точки – во многих случаях вычисления разности имеет существенно более низкую трудоемкость.

В отличие от операторов скрещивания или других неструктурированных операторов рекомбинации предлагаемый подход не позволяет "стягиваться" точкам текущей популяции P , поэтому отпадает необходимость в диверсификации результатов, которая обычно производится в ГА или МА.

После этапов вариации и мутации образуется временное множество P , количество точек в которой ограничено величиной $m+k+l$. Задачей процедуры *ОтсевПопуляции* является уменьшение числа точек в P снова к величине m .

Использование предложенной схемы позволяет порождать семейство алгоритмов КО. Заметим, что при $m=1, k=0, l > 0$ получаем схему повторяющегося локального поиска [6], а при $l = 0$ – алгоритм без использования возмущений.

Отметим также, что механизм порождения субоптимального решения на полуинтервале позволяет провести отдаленную аналогию с алгоритмом рассеянного поиска [10]: его можно представлять как

действие специфического оператора рекомбинации, который формирует потомка от трех родителей – точек $x, y, x^\infty \in X$.

Другой путь порождения алгоритмов – за счет альтернативных способов конкретизации (с учетом опыта применения и специфики решаемой задачи) следующих аспектов общей схемы метода, приведенной на рис.

1) Правило отбора очередной пары точек $x, y \in P$ для вариации:

- выбор таких точек, которые отвечают наилучшему и наихудшему значению функции пригодности среди точек P ;
- случайный выбор точек с вероятностью, пропорциональной пригодности или другим показателям;
- выбор "наихудшей точки" x и случайный выбор точки y , причем при этом следует учитывать и удаленность точек x и y – как в пространстве решений задачи, так и по значениям целевой функции.

Во всех случаях целесообразно, как показал опыт решения задач КО, в качестве начальной точки x выбирать ту точку из отобранной пары, у которой "худшее значение" целевой функции. Иными словами, для задач минимизации более эффективно поиск осуществляется в случае, когда выполняется неравенство $f(x) > f(y)$.

2) *Отбор для мутации*: в самом распространенном случае используется случайный отбор l точек. Сама мутация обычно заключается в возмущении нескольких компонентов заданного варианта решения, выбранных случайным образом.

3) Стратегия формирования новой популяции (*Отбор Популяции*), которая должна в итоге состоять из m точек, из образованной временной популяции объемом $m+k+l$ точек:

- отбор m самых пригодных точек;
- включение потомков на основе критерия Метрополиса;
- замена всех m родителей лучшими из $k+l$ потомков, $m \leq k+l$.

При определении функции пригодности для конкретной точки из P кроме значения целевой функции уместно дополнительно учитывать еще и такие показатели:

- число процедур отбора, в которых данная точка была выбрана для вариации;
- число потомков, которые уже отбирались в обновленные популяции;
- "время жизни", т.е. число осуществленных итераций, в течение которых точка не выбывала из множества P .

4) Критерием завершения вычислительного процесса может быть:

- превышение заданного числа H_{max} обновлений популяции P (алгоритм завершает работу, если $h > H_{max}$);
- стабилизация рекордного или среднего значения целевой функции в точках множества P :

$$\bar{f} = \frac{\sum_{x \in P} f(x)}{m};$$

- стремление к нулю разброса значений целевой функции в точках множества P : $\sigma = \max\{f(x): x \in P\} - \min\{f(x): x \in P\}$;
- достижение ситуации, когда все пары точек из очередной популяции уже принимали участие в формировании полуинтервалов – если используется однозначный способ построения метрических полуинтервалов $\langle x, x^\infty \rangle$.

5) *Способ построения полуинтервалов $\langle x, x^\infty \rangle$* . Рассматриваемые конечные комбинаторные метрические пространства X имеют две такие характерные особенности.

Во-первых, для произвольного $x \in P$ "максимально удаленной" является точка, удовлетворяющая условию: $x^\infty = \max \{d(x, y) : y \in X\}$, где $d(x, y)$ – заданная на пространстве X метрика. Отметим, что часто это значение равно диаметру пространства X .

Во-вторых, во многих комбинаторных метрических пространствах между двумя несоседними точками $x, y \in P$ бывает возможным построение не одного, а сразу нескольких d -отрезков. Поэтому при реализации H -алгоритмов следует различать случай, когда из множества возможных d -отрезков $/x, y/$, $d(x, y) > h$, всегда по конкретному правилу построения выбирается только один отрезок, или когда могут рассматриваться/строиться все возможные отрезки. В зависимости от этого может модифицироваться и правило "*ОтсевДляВариации*": если рассматривается один возможный полуинтервал $\langle x, x^\infty /$, причем $y \in \langle x, x^\infty /$, то вершины x и y после их выбора следует отметить как "отработанные", поскольку их повторный выбор становится нецелесообразным. Если же имеется возможность строить все или несколько возможных полуинтервалов $\langle x, x^\infty /$, то повторный отбор некоторой пары x, y становится возможным, если при этом будет обеспечено построение именно нового полуинтервала.

Отметим, что при решении задач КО большой размерности процедура оптимизации вдоль полуинтервала может оказаться достаточно трудоемкой. В таких случаях можно ввести дополнительный параметр алгоритма, который бы определял часть полуинтервала $\langle x, x^\infty /$, подлежащую просмотру в задачах (2) – то есть определял бы максимально допустимую удаленность точек x и x^∞ в смысле метрического расстояния.

Следует особо отметить, что аналогичный параметр обязательно следует вводить в том случае, когда пространство X является бесконечным, поскольку тогда точка x^∞ действительно устремляется в бесконечность, а полуинтервал превращается в луч, а задача (2) будет иметь бесконечное число вариантов решения.

Вычислительный эксперимент

Поскольку теоретические исследования алгоритмов КО крайне редко позволяют получать результаты, пригодные при решении практических задач, принято анализировать показатели эффективности путем проведения вычислительных экспериментов. С этой целью часто используют "классические" модели КО – прежде всего, задачу коммивояжера и квадратичную задачу о назначениях [1]. В [11] приведены результаты вычислительного эксперимента по сравнению метаэвристического метода G-ГА, основанного на синтезе (гибридизации) схемы G-алгоритмов и ГА, с разными модификациями ГА (в том числе, близкими к МА [1,12]) при решении известных задач коммивояжера. Результаты этого эксперимента продемонстрировали преимущество метода G-ГА, что послужило основанием для использования его в нашем эксперименте по решению серии более трудоемких квадратичных задач о назначениях, заимствованных из известной Интернет-библиотеки [13].

Для решения названных задач был отобран специализированный комбинированный алгоритм, разработанный специально для решения задач на перестановках на основе схем ГА и локальной оптимизации (обозначение КГА), а также лидер упомянутого эксперимента – алгоритм G-ГА [11]. В табл. приведены результаты решения десяти квадратичных задач о назначениях из [13]. Здесь f^* – известное значение целевой функции в точке глобального минимума, f – найденное соответствующим алгоритмом значение целевой функции, δ – относительная погрешность алгоритма (%), t – время счета на ПЭВМ класса Pentium-II (с). В реализованном алгоритме H -метода выбиралось $m=10$, а мутации не производились ($l = 0$). Для квадратичных задач о назначениях меньшей размерности (19–20) лучшие результаты продемонстрировали ГА, но уже для $n > 20$ проявилось преимущество H -метода, который чаще находил точное решения. Из алгоритмов на базе генетических точнее был G-ГА, причем получаемая им точность была высшей по отношению к результатам КГА более чем в два раза.

Использование специально разработанной процедуры автонастройки параметров позволило в дальнейшем H -методу достичь оптимальных значений практически во всех задачах (в задаче LIPA40A – за 150 с, в задачах CHR20A и CHR20C – за 7 и 5 с соответственно).

Таблица 1. Результаты решения КЗН

Название задачи	n	f	КГА			G-ГА			H-метод		
			f	δ	t	f	δ	t	f	δ	t
ELS19	19	1721254	1721254	0	4	1721254	0	6	1721254	0	18
CHR20A	20	2192	2192	0	2	2192	0	6	2396	9	5
CHR20B	20	2298	2298	0	3	2298	0	5	2498	9	6
CHR20C	20	14142	14142	0	5	14142	0	7	16872	19	17
NUG21	21	2438	2447	0.4	5	2438	0	8	2438	0	6
NUG22	22	3596	3614	0.5	8	3598	0.1	7	3596	0	6
LIPA40A	40	31538	31779	0.8	80	31629	0.3	174	31649	0.4	103
LIPA40B	40	476581	485132	1.8	152	480056	0.7	132	476581	0	28
LIPA50A	50	62093	62891	1.3	612	62467	0.6	668	62589	0.8	325
LIPA50B	50	1210244	1246510	3.0	694	1227348	1.4	649	1210244	0	185

Заключение

Предложенный H -метод может быть использован для решения широкого круга задач комбинаторной оптимизации, поскольку он изложен при весьма общих предположениях о решаемой задаче КО. Его вычислительная схема эффективно распараллеливается, что подтверждено итогами вычислительного эксперимента по решению задач КО на кластерном МВК СКИТ, разработанном в Институте кибернетики им. В.М. Глушкова НАН Украины [14].

Важной целью дальнейших теоретических исследований может стать получение условий сходимости и трудоемкости алгоритмов метода, предназначенных для решения конкретных классов задач КО.

Список литературы

1. Hoos H.H., Stützle T. Stochastic Local Search: Foundations and Applications. – San Francisco: Morgan Kaufmann Publ., 2005. – 658 p.
2. Blum C., Roli A. Metaheuristics in Combinatorial Optimization: Overview and Conceptual Comparison // ACM Computing Surveys. – 2003. – 35, No. 3. – P. 268–308.
3. Гуляницкий Л.Ф. Метод деформаций в дискретной оптимизации // Исследование операций и АСУ. – 1989. – Вып. 34. – С. 30–33.
4. Пападимитриу Х., Стайглиц К. Комбинаторная оптимизация. Алгоритмы и сложность. – М.: Мир, 1985. – 512 с.
5. Berge C. Principes de combinatoire. – Paris: Dunod, 1968. – 146 p.
6. Lourenço H. R., Martin O., Stützle T. Iterated local search // Handbook of Metaheuristics: International Series in Operations Research & Management Science, vol. 57 (Eds. F. Glover and G. Kochenberger). – Norwell: Kluwer Academic Publishers, MA, 2002. – P. 321–353.
7. Гуляницкий Л.Ф. Решение задач комбинаторной оптимизации алгоритмами ускоренного вероятностного моделирования // Компьютерная математика. – Киев: Ин-т кибернетики им. В.М. Глушкова НАН Украины, 2004. – №1. – С. 64–72.
8. Сергиенко И.В., Гуляницкий Л.Ф. Фронтальные алгоритмы для многопроцессорных ЭВМ // Кибернетика. – 1981. – №6. – С.1–4.
9. Гуляницкий Л.Ф. Один гибридный алгоритм комбинаторной оптимизации // Abstract of Int. Ukrainian-Polish Workshop "Problems of Stochastic and Discrete Optimization" (May 10-15, 2005, Kaniv, Ukraine). – Kaniv, 2005. – P. 63–65.

10. Glover F. Scatter Search and Path Relinking // New Ideas in Optimization (Eds. D.Corne, M.Dorigo, F.Glover). – McGraw Hill, 1999. – P. 297-316.
11. Гуляницкий Л.Ф. Разработка гибридных методов дискретной оптимизации на основе G-алгоритмов // Компьютерная математика. – Киев: Ин-т кибернетики им. В.М. Глушкова НАН Украины, 2005. №1. С. 143–151.
12. Haupt R.L., Haupt S.E. Practical Genetic Algorithms (2nd Ed.). – Hoboken: John Wiley&Sons, 2004. – 253 p.
13. QAPLIB //www.opt.math.tu-graz.ac.at/qaplib
14. Гуляницкий Л.Ф., Гложик Ю.С. Параллельный метод деформируемых многогранников для решения задач комбинаторной оптимизации // Искусственный интеллект. – 2005. – 4. – С.130–139.

Об авторе

Леонид Гуляницкий (Hulianytskyi) – д.т.н., Институт кибернетики им. В.М.Глушкова НАН Украины, пр-т Глушкова, 40, Киев, 03680, Украина. e-mail: lh_dar@hotmail.com

ОДИН КЛАСС АЛГОРИТМОВ СТОХАСТИЧЕСКОГО ЛОКАЛЬНОГО ПОИСКА

Леонид Гуляницкий, Александр Турчин

Аннотация: Рассматриваются алгоритмы ускоренного вероятностного моделирования, которые относятся к классу методов стохастического локального поиска. Приводится общая вычислительная схема алгоритмов, на основе которой с использованием принципа «золотого сечения» предлагается алгоритм комбинаторной оптимизации, названный GS-алгоритмом. Исследуются вопросы сходимости предложенного алгоритма на основе анализа цепей Маркова. Приведена оценка числа операций, необходимых для достижения глобального решения. Дан обзор применения рассмотренных алгоритмов для решения разных типов задач комбинаторной оптимизации.

Ключевые слова: combinatorial optimization, stochastic local search, simulated annealing, Markov chains.

Введение

Для решения задач комбинаторной оптимизации (КО) широкое применение нашли приближенные алгоритмы. Это объясняется рядом обстоятельств, среди которых выделим следующие: во-первых, все практически важные задачи относятся к числу NP -трудных, так что точное их решение весьма проблематично даже с использованием современных и перспективных компьютеров; во-вторых, целевые функции имеют, как правило, большое количество локальных экстремумов; в-третьих, во многих прикладных проблемах данные задаются с определенными погрешностями, что делает нецелесообразными те существенные вычислительные затраты, которые необходимы для нахождения их точного решения; в-четвертых, положенные в основу разработки приближенных вычислительных схем идеи (метаэвристики) позволяют создавать алгоритмы, которые могут решать не одну, а целый класс близких по формулировке оптимизационных задач.

В дальнейшем под задачей КО будем понимать проблему поиска хотя бы одного аргумента экстремума заданной целевой функции:

$$x_* = \arg \min_{x \in D_x \subseteq X} f(x) \quad (1)$$

где X – комбинаторное (конечное) пространство, D_π – его подмножество, элементы которого удовлетворяют ограничивающим условиям задачи, $f: X \rightarrow \mathfrak{R}^1$ – целевая функция задачи минимизации.

Отметим, что большинство полученных результатов без труда переносится и на тот случай, когда X (или только D_π) – дискретное счетное пространство.

Одним из распространенных и эффективных алгоритмов КО является метод имитационного отжига (МИО) [1]. Опыт применения алгоритмов МИО показал, что получаемые результаты имеют высокую точность, эффективно реализуются на многопроцессорных ЭВМ.

В МИО вероятность принятия очередного решения зависит от степени изменения целевой функции, абсолютная величина которой масштабируется с помощью параметра T (в МИО он называется температурой). Его изменение (температурное расписание) задают так, чтобы по ходу применения алгоритма в исследуемой окрестности вероятности перехода к худшим вариантам решения (т.е. к вариантам с большим значением целевой функции) убывали бы от итерации к итерации, стремясь к нулю. Условия, которые определяют необходимость перехода к новому значению параметра T , в МИО называются условиями равновесия. Наличие такого вероятностного механизма создает предпосылки для выхода в начале поиска алгоритмов МИО из локальных экстремумов, а в конце вычислений способствует концентрации поиска вокруг улучшенных локальных решений.

В тот же время, проведенные многими исследователями вычислительные эксперименты выявили существенные затраты компьютерного времени на поиск решений, а также значительную изменчивость точности получаемых вариантов решений задачи в зависимости от задания значений варьируемых параметров. Необходимость в обобщении вероятностных механизмов и в повышении эффективности алгоритмов дискретной оптимизации привела к разработке алгоритмов ускоренного вероятностного моделирования [2], в которых применена иная вероятностная модель.

G-алгоритмы

В алгоритмах ускоренного вероятностного моделирования, которые получили еще название G-алгоритмы, на каждой итерации также осуществляется построение точек из окрестности текущего варианта и улучшающие варианты всегда принимаются в качестве очередного приближения, а варианты, соответствующие ухудшению (возрастанию – в случае минимизации) целевой функции, также могут быть выбраны с некоторой вероятностью. Однако, в отличие от МИО, значения этих вероятностей рассчитываются аналогично на протяжении всего вычислительного процесса, но изменяется весовое значение, которое определяет условия отсева ухудшающих вариантов.

Для построения вычислительного процесса формируется строго монотонно возрастающая последовательность действительных чисел $\{\mu_t\}$, $0 \leq \mu_0 < \mu_1 < \dots \leq 1$, которые играют роль, в каком-то смысле соответствующую роли параметра T в МИО. Если x^h – это найденный текущий вариант на шаге h , а $L(x)$ – заданная (метрическая) окрестность произвольной точки $x \in X$, то исследуемый вариант $y \in L(x^h)$, для которого $f(y) > f(x^h)$, может быть принят в качестве x^{h+1} с вероятностью перехода $p(x^h, y)$, зависящей от текущих значений величины μ_t .

Пусть $\Phi(x, y)$, $0 \leq \Phi(x, y) \leq 1$, – некоторый функционал, зависящий от значений целевой функции задачи. Тогда определим вероятность перехода от точки x к точке y , $x, y \in X$, так:

$$p \equiv p(x, y) = (1 - \mu_t) \cdot \Phi(x, y)$$

Предлагается следующая обобщенная схема G-алгоритмов для решения задач вида (1), представленная на рис.1.

```

procedure G_Search(x)
begin
   $x^0 :=$  некоторый начальный допустимый вариант решения из  $X$ ;
   $\mu_0 := 0$ ;  $h := 0$ ;  $t := 0$ ;
   $x_{rec} := x^0$ ;  $f_{rec} := f(x^0)$ ;
  while окрестность текущего решения  $L(x^h)$  не просмотрена полностью do
    begin
      while не выполнено условие равновесия do
        begin
           $y :=$  ГенерированиеСледующейТочкиОкрестности  $L(x^h)$ ;
          Вычисление  $\Phi(x^h, y)$ ;
           $p := (1 - \mu_t) \Phi(x^h, y)$ ;
           $\xi := \text{random}[0,1]$ ;
          if  $p \geq \xi$  then
             $h := h + 1$ ;  $x^h := y$ ;
            if  $f_{rec} > f(x^h)$  then
               $x_{rec} := x^h$ ;  $f_{rec} := f(x^h)$ 
            end if
          end if
        end;
        ФормированиеОчередногоЗначения  $\mu_{t+1}$ ;
         $t := t + 1$ ;
      end;
    return  $x = x_{rec}$ ;
  end

```

Рис. 1. Алгоритм ускоренного вероятностного моделирования (G-алгоритм)

Здесь $\text{random}[0,1]$ – датчик случайных чисел из отрезка $[0,1]$, при помощи которого моделируются вероятностные переходы.

Построение конкретного алгоритма осуществляется путем конкретизации таких основных аспектов:

- задание функционала $\Phi(x,y)$;
- механизм построения последовательности $\{\mu_t\}$;
- условия равновесия при данном значении величины μ_t ;
- правило останова.

При задании функционала $\Phi(x,y)$ можно использовать монотонные по значениям целевой функции зависимости, удовлетворяющие двум условиям:

- а) $\Phi(x,y) \rightarrow 1$, если $f(y) \rightarrow f(x)$.
- б) $\Phi(x,y) \rightarrow 0$, если $f(y) \rightarrow \infty$.

Примером функционала такого рода может быть функционал:

$$\Phi(x, y) = \begin{cases} \left[\frac{f(x)}{f(y)} \right]^\beta, & \text{если } f(y) \geq f(x), \\ 1, & \text{в противном случае.} \end{cases}$$

где параметр алгоритма $\beta > 0$ – действительное, в частности, натуральное число.

Во многих практических задачах КО не составляет особых затруднений нахождение верхней границы целевой функции (учитывая, что пространство X во многих случаях конечно), т.е. такой величины f_{\max} , что $f(x) \leq f_{\max}, \forall x \in X$.

В этом случае можно положить

$$\Phi(x, y) = 1 - \left[\frac{f(y) - f(x)}{f_{\max} - f(x)} \right]^\beta.$$

В обычных G-алгоритмах (см. [6] и цитированную там литературу) в качестве функционала $\Phi(x, y)$ выбирается кусочно-линейный функционал

$$\varphi(x, y) = \begin{cases} \min \left\{ 1, 1 - \frac{f(y) - f(x)}{\gamma f(x)} \right\}, & \text{если } f(y) \geq f(x), \\ 0, & \text{если } f(y) < f(x), \end{cases}$$

где параметр $\gamma, \gamma > 0$, – действительное число.

В этом случае при текущем варианте x^h вероятность перехода $p(x^h, y)$ к точке $y \in L(x^h)$ представляется кусочно-линейным функционалом, состоящим из трех фрагментов: при $f(y) < f(x^h)$ его значение равно 1, при $f(x^h) \leq f(y) \leq (1 + \gamma) f(x^h)$ он убывает от 1 до 0, а при $f(y) > (1 + \gamma) f(x^h)$ – принимает нулевое значение (рис. 2).

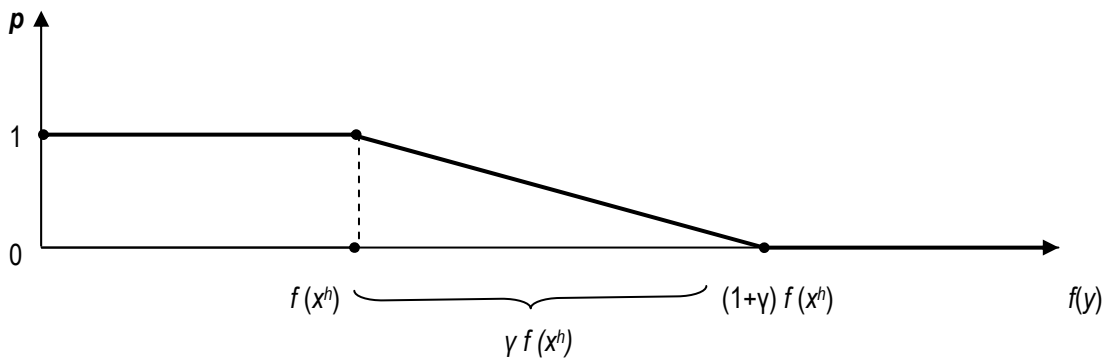


Рис. 2. Вероятность перехода в новую точку окрестности

Таким образом, значение параметра γ определяет величину среднего интервала $\gamma \cdot f(x^h)$, при попадании значений $f(y)$ в который точка y еще может быть выбрана в качестве x^{h+1} . Соответственно, при более значительном ухудшении (увеличении в случае поиска минимума) значений $f(y)$ эта точка всегда удаляется из рассмотрения. Важно отметить, что величина этого интервала автоматически уменьшается при уменьшении $f(x^h)$, что происходит при приближении к более точным вариантам решения задачи, и увеличивается в противном случае.

Кроме этого, использование относительных величин позволяет избежать зависимости от абсолютных значений целевой функции. Тем самым, происходит адаптация вычислительного процесса к динамике изменения значений (ландшафта) целевой функции.

В настоящее время наиболее часто применяются два способа порождения последовательности $\{\mu_i\}$: через определение и использование некоторой строго монотонной функции, обозначаемой G (отсюда и сокращенное название – G-алгоритмы [2]), и на основе использования принципа “золотого сечения” [3].

Пусть $G: [0, 1] \rightarrow [0, 1]$ – некоторая заданная строго монотонная действительнoзначная функция, тогда в G-алгоритмах полагают $\mu_{i+1} = G(\mu_i)$. В качестве функции G на практике обычно выбирают функции вида

$$G_k(x) = \min \{1, (x^{1/k} + H)^k\}, \quad x \geq 0. \quad (3)$$

где $k \in \{1, 2, 3\}$, а $H, 0 < H < 1$, – некоторая малая величина, выбором значения которой достигается скорость следования $G_k(x) \rightarrow 1$ при $x \rightarrow 1$.

Альтернативно, значения μ_i могут формироваться путем деления отрезка $[0, 1]$ с определенным шагом и последовательным выбором построенных точек. Следует заметить, что в случае конечности упомянутой

последовательности после достижения максимального значения (равного единице) алгоритм трансформируется в детерминированный локальный поиск.

Эффективным подходом стало использование правила «золотого сечения», которое (хотя и совсем иным способом) используется и при построении алгоритмов поиска экстремумов непрерывных функций одного аргумента. В нашем случае определяется величина μ_t как меньшая (левая) из двух точек, которые реализовывают «золотое сечение» отрезка $[\mu_t, 1]$ (при условии, что $\mu_0 = 0$, а $\mu_{t+1} = G(\mu_t)$). Таким образом, правило «золотого сечения» задает скорость приближения левой границы этого отрезка к 1, а соответствующие величины μ_t выступают аргументами выбранной функции $G(x)$, например, из (3). Алгоритм, использующий правило «золотого сечения», назовем GS-алгоритмом.

При разработке условий равновесия, учитывая аналогию между параметром "температура" в МИО и μ , целесообразно использовать опыт, накопленный при создании алгоритмов МИО. В частности, равновесие может определяться так: задаются некоторый (натуральный) параметр ν и действительное число $\varepsilon > 0$, а выполнение ν переходов называют прогоном [4]. Если при данной температуре выполнено k прогонов и получены значения f_1, \dots, f_k , то считают, что достигнуто равновесие, если в $(k+1)$ -м прогоне выполняется неравенство

$$|f_{k+1} - f_i| \leq \varepsilon,$$

для некоторого $i \in \{1, \dots, k\}$. Величина f_i для прогона может быть, например, средним значением целевой функции или наилучшим среди тех, которые составляют прогон.

Правилом останова может служить: окончание перебора всех точек в окрестности без реализованного перехода в новую точку; ограничение по продолжительности работы алгоритма; достижение требуемой точности – при известной нижней границе целевой функции. Еще одно из распространенных правил – сравнение разности максимального и минимального значений целевой функции с максимальным значением изменения этой функции при данном значении μ : если это соотношение стремится к единице, то вычисления завершаются.

Следует отметить, что в вычислительной схеме G-алгоритмов достаточно просто учесть многие типы ограничительных условий, которые часто встречаются на практике: соответствующие условия проверки на допустимость могут быть учтены при порождении точек из окрестности текущего варианта. Это позволяет решать не одну, а целый подкласс задач, причем ограничения и некоторые данные задачи могут меняться по ходу ее решения.

По аналогии с алгоритмом повторяющегося локального поиска [1], после завершения работы общей схемы GS-алгоритма можно использовать процедуру возмущения найденного локального решения, а полученный в результате возмущения вариант использовать в качестве нового начального приближения для встроеного GS-алгоритма, порождая тем самым метаэвристический метод, который можно назвать повторяющимся GS-алгоритмом.

Исследование условий сходимости

Изучим вопросы сходимости GS-алгоритма на примере важного класса задач КО – задач на перестановках с транспозиционной метрикой. Для получения оценок скорости сходимости траекторию поиска представим в виде графа. Для этого построим полный взвешенный граф $Gr = (V, H)$, где $V = \{1, \dots, n\}$ является множеством вершин, а H – множеством ребер графа. Пути между двумя точками пространства решений X оптимизационной задачи (1) будут соответствовать последовательность ребер этого графа: каждая из вершин графа, соответствующая $x \in X$, будет инцидентна тем вершинам, которые соответствуют точкам окрестности $L(x)$. Не ограничивая общности, будем рассматривать окрестности минимального радиуса – в большинстве случаев комбинаторных пространств он равен единице: $L(x) = L_1(x)$. Поставим в соответствие окрестности $L(x)$ множество $N(v), v \in V$

следующим образом: если точка x' находилась в окрестности $L(x)$ точки $x \in X$, то вершина, соответствующая x' в графе Gr , будет связана ребром с вершиной-образом точки x . Иными словами, если точки в пространстве X отличались одной транспозицией – то на графе Gr их вершины-образы будут соединены ребром.

В этом случае случайному процессу поиска решения задачи соответствует процесс переходов от одной вершины графа к другой с использованием определенных вероятностных механизмов. В нашем случае степень d графа $Gr = (V, H)$ равна $d = \frac{n(n-1)}{2}$, а диаметр $D = n - 1$.

Тогда при выполнении GS-алгоритма процесс поиска моделируется цепью Маркова, поскольку определение состояния, в которое состоится переход, зависит только от текущего состояния системы. Будем считать, что GS-алгоритм сходится, если соответствующая цепь Маркова как минимум раз будет содержать состояние, соответствующее глобальному оптимуму.

Вероятность перехода в следующее состояние при условии нахождения приближения решения с не меньшим значением целевой функции ограничена снизу величиной μ^s . Обозначим

$$\hat{d} = \left(\frac{1}{d}\right)^{-D}, \hat{\mu} = \prod_{i=1}^D \mu^i.$$

Лемма. Пусть $v \in V$ – вершина, соответствующая некоторому состоянию x . Тогда ожидаемое число шагов, за которое будет достигнута вершина, соответствующая глобальному оптимуму, не превышает $\frac{\hat{d}}{\hat{\mu}}$.

Доказательство. Пусть \underline{x} – глобальный оптимум задачи (1), и пусть в графе $Gr = (V, H)$ этому оптимуму соответствует вершина \underline{v} . Существует путь от вершины v к вершине \underline{v} длины $q < D$. Значит, существует последовательность вершин v_1, v_2, \dots, v_q , которые определяют маршрут.

Если допустить, что любой элемент окрестности $L(x)$ может быть с одинаковой вероятностью выбран в качестве следующего приближения решения, то можно утверждать, что за q шагов вероятность перехода

в состояние \underline{v} будет, как минимум, равной $\left(\frac{1}{d}\right)^q \times \prod_{i=1}^q \mu^i$ и будет выполняться неравенство

$$\left(\frac{1}{d}\right)^q \times \prod_{i=1}^q \mu^i \geq \left(\frac{1}{d}\right)^D \prod_{i=1}^D \mu^i = \frac{\hat{\mu}}{\hat{d}}.$$

Следовательно, вероятность попадания в состояние \underline{v} , начиная с определенного состояния v , будет не меньше, чем $\frac{\hat{\mu}}{\hat{d}}$. А значит, можно утверждать, что ожидаемое число шагов алгоритма, необходимое для

попадания в глобальный оптимум, не буде превышать $\frac{\hat{d}}{\hat{\mu}}$.

Лемма доказана.

Теорема. GS-алгоритм достигает глобального оптимума целевой функции с вероятностью большей, чем $(1 - \frac{1}{C^k})$, за число шагов, не превышающее $Ck \cdot \frac{\hat{d}}{\hat{\mu}}$, и эта оценка не зависит от начального приближения решения ($C = const > 1$).

Доказательство. Для доказательства используем метод математической индукции по k . Положим $Q = C \left[\frac{\hat{\mu}}{\hat{d}} \right]$. Докажем, что вероятность непопадания за kQ шагов в вершину \underline{v} , соответствующую глобальному оптимуму, не превышает $1/C^k$.

Базовый случай соответствует $k=1$ и определенному состоянию x . В соответствии с леммой, необходимое число итераций для достижения вершины \underline{v} не будет превышать $\frac{\hat{d}}{\hat{\mu}}$. Используя неравенство Маркова, получаем:

$$p\left(\frac{\hat{d}}{\hat{\mu}} > Q\right) = \frac{M\left(\frac{\hat{d}}{\hat{\mu}}\right)}{Q} = \frac{\frac{\hat{d}}{\hat{\mu}}}{C \frac{\hat{d}}{\hat{\mu}}} = \frac{1}{C}.$$

Допустим, что утверждение справедливо для всех $k \leq K-1$. Докажем, что утверждение будет справедливым и для $k=K$. Пусть $x_Q, x_{2Q}, \dots, x_{(K-1)Q}$ – последовательность элементов цепи Маркова, построенная за число шагов $Q, 2Q, \dots, (K-1)Q$ соответственно.

Рассмотрим два события:

Событие $H1$: вершина \underline{v} не достигнута за первые Q шагов;

Событие $H2$: вершина \underline{v} не достигнута за следующие $(K-1)Q$ шагов.

В соответствии с таким определением событий вероятность того, что вершина \underline{v} не будет достигнута за KQ шагов, равна $p = P(H2 | H1) \times P(H1)$.

Из теории цепей Маркова известно, что вероятность достижения глобального оптимума за определенное число шагов, начиная с некоторого состояния, зависит только от количества шагов и текущего состояния и не зависит от состояний, посещенных раньше.

Используя этот факт можно утверждать, что вероятность $P(H2 | H1)$ зависит только того состояния, в котором будет находиться цепь Маркова на итерации Q , и от шагов $(K-1)Q$. Значит,

$$P = P(H1) \sum_{i \in V} P(H2 | x_Q = i) \times P(x_Q = i).$$

Вероятность $P(H1)$ не превышает $\frac{1}{C}$ в соответствии с базовым случаем, а вероятность

$P(H2 | x_Q = i)$ не превышает $\frac{1}{C^{K-1}}$ для каждого $i \in V$, исходя из гипотезы математической индукции. Имеем:

$$P \leq \frac{1}{C} \times \frac{1}{C^{K-1}} = \frac{1}{C^K}.$$

Вероятность попадания в глобальный оптимум будет обратной к вероятности непопадания.

Следовательно, можно утверждать, что с вероятностью, большей чем $\left(1 - \frac{1}{C^k}\right)$, будет достигнут

глобальный оптимум целевой функции за число шагов, не превышающее $Ck \cdot \frac{\hat{d}}{\hat{\mu}}$, и эта оценка не зависит от начального приближения решения.

Теорема доказана.

Следствие. При $k \rightarrow \infty$ последовательность построенных приближений решения сходится к глобальному оптимуму по вероятности.

Доказательство этого факта можно получить, исходя из определения сходимости по вероятности.

Особенность полученных оценок состоит в том, что они обобщают известные оценки сходимости алгоритмов МИО (см., например, [5]).

Полученные результаты теоретического исследования сходимости GS-алгоритма могут использоваться и при исследовании иных алгоритмов ускоренного вероятностного моделирования и близких к ним техник поиска.

Практические применения

Разработанные алгоритмы КО могут применяться для решения задач из разных классов, поскольку они предъявляют весьма общие требования к формулировкам этих задач. Были исследованы алгоритмы решения как известных проблем (задач коммивояжера, квадратичных задач о назначениях), так и более специальных задач (задач размещения, складирования продукции, проектирования технических систем, настройки параметров модели при макроэкономическом прогнозировании и др.).

Одним из первых применений G-алгоритмов было решение задачи коммивояжера – наверное, наиболее известной из проблем КО как в плане приложений, так и как полигона для исследования предлагаемых исследователями алгоритмов. В [6] приведены решения ряда известных задач, причем результаты оказались сравнимыми с решениями, полученными с использованием супер-ЭВМ. За последнее время накоплен опыт решения комбинированными алгоритмами, основанными на синтезе генетических алгоритмов и G-алгоритмов.

Широкое применение в различных сферах находит и квадратичная задача о назначениях: разработка радиоэлектронной и цифровой аппаратуры, экономика, планирование и размещение производства, проектирование сложных технических систем и др. Несмотря на то, что эта задача исследуется давно, точное решение даже при использовании супер-ЭВМ вызывает затруднение в общем случае уже для $n > 15-20$, причем поиск ε -приближенного решения является также NP-полной проблемой [1]. Полученные результаты позволили отобрать наиболее эффективные варианты вычислительных схем, на базе которых в дальнейшем были созданы алгоритмы, использующие правило «золотого сечения» [3], а также метаэвристические алгоритмы (они применялись, в частности, для решения задач коммивояжера и квадратичных задач о назначениях [7,8]).

В работе [9] рассматривался один из классов задач гильотинного раскроя – проблемы оптимального размещения прямоугольников фиксированной высоты и разной длины на полубесконечной ленте, состоящей из заданного числа полос той же высоты при наличии ограничений на размещение [10]. Предложен ряд алгоритмов размещения, которые основаны на схемах локальной оптимизации, имитационного отжига и G-алгоритмов. Результаты проведенных вычислительных экспериментов показали, что для рассмотренного типа задач G-алгоритмы позволяют получать лучшие результаты как по точности, так и по трудоемкости по сравнению с другими рассмотренными алгоритмами.

При выборе стратегических аспектов управления запасами на предприятиях и в компаниях возникает проблема оптимального размещения нескольких видов продукции на складе ограниченной вместимости с учетом информации о вероятных потребностях в каждом виде и убытках от неудовлетворенного спроса [11]. Разработанный G-алгоритм сравнивался с тремя модификациям алгоритма локального поиска и имитационного отжига. Полученные результаты обширного вычислительного эксперимента продемонстрировали стабильность и высокую точность решений, полученных G-алгоритмом при меньших по сравнению с имитационным отжигом затратах времени.

Одной из важных задач проектирования и оптимизации телекоммуникационных сетей с технологией АТМ является задача оптимального выбора пропускных способностей каналов связи. В ней задана структура

сети, состоящая из коммутаторов, соединенных каналами связи определенной длины. Заданы также требования в передаче трафика для каждой упорядоченной пара коммутаторов, а для каждого канала связи известны величины общих потоков. Пропускная способность любого канала связи пропорциональна пропускной способности базового канала. Необходимо для всех каналов связи выбрать такое количество базовых каналов, при котором стоимость функционирования сети будет минимальной при выполнении ограничений на показатели качества обслуживания. Для решения названных задач были разработаны и реализованы алгоритмы локального поиска и повторяющегося локального поиска, МИВ, G-алгоритм, а также генетический алгоритм. Анализ результатов вычислительных экспериментов показал, что для разных видов трафика наилучшие по точности решения в большинстве случаев находились G-алгоритмом, сравнимые с ними результаты для задач с отдельными видами трафика показал алгоритм повторяющегося локального поиска [12].

Заключение

В работе описаны алгоритмы, принадлежащие к классу методов стохастического локального поиска. Предложена их модификация, названная GS-алгоритмом, которая использует правило "золотого сечения". Применение оригинальных вероятностных механизмов организации вычислений в соединении с процедурой локального поиска позволяет эффективно решать задачи КО из разных классов. На основе использования цепей Маркова получена вероятностная оценка верхней границы для числа итераций, необходимых для нахождения глобального минимума. Приведены примеры применения алгоритмов рассмотренного типа для решения ряда задач КО. Полученные результаты позволяют рекомендовать разработанные алгоритмы как для решения прикладных задач КО, так и в качестве встроенной процедуры в метаэвристических алгоритмах, прежде всего в алгоритмах, основанных на популяциях [1].

Список литературы

1. Hoos H.H., Stützle T. Stochastic Local Search: Foundations and Applications. – San Francisco: Morgan Kaufmann Publ., 2005. – 658 p.
2. Гуляницкий Л.Ф. Решение задач комбинаторной оптимизации алгоритмами ускоренного вероятностного моделирования // Компьютерная математика. – Киев: Ин-т кибернетики им. В.М. Глушкова НАН Украины, 2004. – №1. – С. 64–72.
3. Гуляницкий Л.Ф., Турчин О.Я. Использование правила "золотого сечения" в алгоритмах вероятностного моделирования // Журн. вычисл. и прикладной математики. – 1998. – № 1(83). – С.26–29 (укр.).
4. Golden B.L., Skiscim C.C. Using simulated annealing to solve routing and location problem // Naval Research Logistics Quarterly. – 1986. – 33, № 6. – P. 273–277.
5. Rajasekaran S. Simulated annealing and nested annealing // J.of Global Optimization.– 2000.–16. –P. 43-56.
6. Гуляницкий Л.Ф. Модифицированные алгоритмы вероятностного моделирования в комбинаторной оптимизации //Технология и методы решения задач прикладной математики. – Киев: Ин-т кибернетики им.В.М.Глушкова АН Украины, 1991. – С. 10–14.
7. Гуляницкий Л.Ф., Турчин О.Я. Об одном подходе к использованию вероятностного моделирования в схеме генетического алгоритма //Тр. Междун. конф. по индуктивному моделированию, Львов 20-25 мая 2002. –Львов: ГНИИ информационной инфраструктуры, 2002. Т 2. – С. 275-281 (укр.).
8. Hulianitsky L., Turchin A. On approach to combining genetic algorithms and accelerated probabilistic modeling algorithms /The Experience of Designing and Application of CAD Systems in Microelectronics: Proc. of the VII Int. Conf. CADMS 2003 (18-22 February 2003, Lviv-Slavske, Ukraine). –Lviv: Publ. House of Lviv Polytechnic National University, 2003. – P. 218–219.
9. Гуляницкий Л.Ф., Гобов Д.А. О сравнении эффективности алгоритмов решения одного класса задач размещения прямоугольников на полубесконечной ленте //Компьютерная математика. – 2003. – №1. С. 88-97.
10. Грицюк Ю.И. Регулярное размещение прямоугольных объектов вдоль полос односторонне ограниченной ленты. – Львов: Изд. дом «Панорама», 2002. – 220 с. (укр.).

11. Гуляницкий Л.Ф. О решении на ПЭВМ одной задачи о складировании продукции // Теория и программная реализация методов дискретной оптимизации. -Киев: Ин-т кибернетики им.В.М.Глушкова АН УССР, 1989. – С.20-25.
12. Huliannytskyi L., Baklan A. Optimization of ATM telecommunication Networks // Int. J. "Information theories & applications". – 2005. – 12, N 4. – P. 328–335.

Об авторах

Леонид Гуляницкий (Huliannytskyi) – д.т.н., Институт кибернетики им. В.М.Глушкова НАН Украины, пр-т Глушкова, 40, Киев, 03680, Украина. e-mail: lh_dar@hotmail.com

Александр Турчин – Институт кибернетики им. В.М.Глушкова НАН Украины, пр-т Глушкова, 40, Киев, 03680, Украина. e-mail: turchin@ua.fm

ОБ ЭФФЕКТИВНОСТИ ФУНКЦИОНАЛОВ ЭМПИРИЧЕСКОГО РИСКА И СКОЛЬЗЯЩЕГО ЭКЗАМЕНА КАК ОЦЕНОК ВЕРОЯТНОСТИ ОШИБОЧНОЙ КЛАССИФИКАЦИИ¹

Виктор Неделько

Abstract: *The goal of the paper is to estimate misclassification probability for decision function by training sample. The simple estimation presented shows how far Vapnik–Chervonenkis risk estimations are off in some special case. The existence of situations when an empirical risk appears to be better risk estimate in comparison with cross validation is investigated.*

Keywords: *pattern recognition, classification, statistical robustness, deciding functions, complexity, capacity, overtraining problem.*

ACM Classification Keywords: *G.3 Probability and statistics, G.1.6. Numerical analysis: Optimization; G.2.m. Discrete mathematics: miscellaneous.*

Введение

Одним из наиболее часто используемых методов оценивания качества решающей функции является скользящий экзамен (cross-validation). Функционал скользящего экзамена, в отличие от эмпирического риска (доля неправильно классифицированных объектов обучающей выборки), является, как известно, несмещенной оценкой риска (вероятности ошибочной классификации).

Вместе с тем, оценка скользящего экзамена имеет относительно большую дисперсию, при этом достаточно точных универсальных оценок этой дисперсии до настоящего времени не найдено. Известные оценки имеют малую точность, и построенные на их основе интервальные оценки для риска оказываются, вообще говоря, не лучше, чем известные оценки Вапника-Червоненкиса (В-Ч) [Вапник, Червоненкис, 1974], основанные на эмпирическом риске.

В настоящей работе будет приведена простая оценка точности оценок В-Ч в некотором частном случае, а также будут исследованы ситуации, когда эмпирический риск является более эффективным функционалом, по сравнению со скользящим экзаменом. Под эффективностью здесь понимается точность оценивания риска.

¹ Работа выполнена при поддержке РФФИ, а также Лаврентьевского гранта СО РАН.

Постановка задачи

Пусть X – пространство значений переменных, используемых для прогноза, а Y – пространство значений прогнозируемых переменных, и пусть C – множество всех вероятностных мер на $D = X \times Y$. Тогда элементом $c \in C$ будет $P_c[D]$. Здесь и далее квадратные скобки используются для указания множества, на σ -алгебре подмножеств которого задана мера.

Решающей функцией назовем соответствие $f: X \rightarrow Y$ и введем для нее функцию потерь:

$$L: Y^2 \rightarrow [0, \infty).$$

Под риском будем понимать средние потери:

$$R(c, f) = \int L(y, f(x)) dP_c[D].$$

Пусть $v = \{(x^i, y^i) \in D \mid i = \overline{1, N}\}$ – случайная независимая выборка из распределения $P_c[D]$.

Эмпирический риск определим как средние потери на выборке:

$$\tilde{R}(v, f) = \frac{1}{N} \sum_{i=1}^N L(y^i, f(x^i)).$$

Пусть $Q: \{v\} \rightarrow \Phi$ – алгоритм построения решающих функций, а $f_{Q,v} \in \Phi$ – функция, построенная по выборке v алгоритмом Q .

Оценкой скользящего экзамена называется величина

$$\bar{R}(v, Q) = \frac{1}{N} \sum_{i=1}^N L(y^i, f_{Q,v'_i}(x^i)),$$

где $v'_i = v \setminus \{(x^i, y^i)\}$ – выборка, получаемая из v удалением i -го наблюдения.

Чтобы по полученному значению эмпирического риска можно было оценивать риск, необходимо так или иначе оценить вероятность уклонения данных величин:

$$\eta(c, \varepsilon) = P\left(\left|\tilde{R}(v, f_{Q,v}) - R(c, f_{Q,v})\right| > \varepsilon\right).$$

Поскольку риск в среднем значительно больше эмпирического риска, возможно оценивать величину:

$$\eta'(c, \varepsilon) = P\left(R(c, f_{Q,v}) > \tilde{R}(v, f_{Q,v}) + \varepsilon\right).$$

Существенная проблема заключается в том, что выражения зависят от c – распределения, которое неизвестно. Решением может быть взятие супремума по всем распределениям и ориентирование таким образом на «наихудшее» распределение.

Кроме самих величин рисков нас будут интересовать их средние.

Обозначим: $F(c, Q) = ER(c, f_{Q,v})$, $\tilde{F}(c, Q) = E\tilde{R}(c, f_{Q,v})$, где математическое ожидание берется по всем выборкам объема N .

Все практически используемые алгоритмы построения решающих функций так или иначе минимизируют эмпирический риск, поэтому последний оказывается смещенной оценкой риска.

Введем функцию максимального смещения:

$$S_Q(\tilde{F}_0) = \hat{F}_Q(\tilde{F}_0) - \tilde{F}_0,$$

где $\hat{F}_Q(\tilde{F}_0) = \sup_{c: \tilde{F}(c, Q) = \tilde{F}_0} F(c, Q)$.

Классификация в дискретном пространстве

Будем рассматривать задачу классификации двух образов.

Пусть X дискретно, то есть $X = \{1, \dots, n\}$, и решающая функция минимизирует эмпирический риск независимо в каждой точке x .

При этом $Y = \{0, 1\}$, функцией потерь будет: $L(y, y') = \begin{cases} 0, & y = y' \\ 1, & y \neq y' \end{cases}$, а риском – вероятность ошибочной классификации.

Тогда вероятностная мера $c \in C$ задается набором вероятностей $c = \{p_j^\omega = P(x = j, y = \omega) \mid j = \overline{1, n}, \omega = \overline{0, 1}\}$.

Чтобы проиллюстрировать степень точности оценок Вапника-Червоненкиса, рассмотрим так называемый «детерминистский» случай, когда в классе решающих функций всегда находится решение, которое классифицирует выборку точно, т. е. $\tilde{R} = 0$.

При этом будем рассматривать асимптотический случай: $\frac{N}{n} = M = \text{const}$, $N \rightarrow \infty$, $n \rightarrow \infty$. Следует подчеркнуть, что мы рассматриваем «асимптотику малых выборок», то есть условия, близкие к условиям подавляющего большинства реальных задач.

Для начала приведем вывод асимптотических выражений для оценок В-Ч, когда эмпирический риск равен нулю.

Вероятность уклонения для риска находится как:

$$P(R - \tilde{R} \geq \varepsilon) = P(\tilde{R} = 0 / R = \varepsilon) = (1 - \varepsilon)^N.$$

Для равномерного уклонения имеем оценку:

$$P\left(\sup_{f \in \Phi} |R - \tilde{R}| \geq \varepsilon\right) < |\Phi| (1 - \varepsilon)^N.$$

Приравняем правую часть величине выбранного уровня значимости η :

$$\ln|\Phi| + N \ln(1 - \varepsilon) = \ln \eta.$$

Легко заметить, что в асимптотике больших N слагаемым $\ln \eta$ можно пренебречь.

Теперь, учитывая, что $|\Phi| = 2^n (1 - e^{-M})$ и $\ln(1 - \varepsilon) \approx -\varepsilon$ получаем:

$$S'_V(0) = \varepsilon = \frac{(1 - e^{-M}) \ln 2}{M}.$$

Множитель $1 - e^{-M}$ есть вероятность ненулевого числа исходов при распределении Пуассона, или средняя доля значений переменной X , с ненулевым числом выборочных точек. Данная поправка отражает тот факт, что число решающих функций, различимых на выборке, определяется именно числом «непустых» значений.

Точное значение смещения риска при $\tilde{F}_0 = 0$ составляет:

$$S_Q(0) = \begin{cases} \frac{1}{2} e^{-M}, & M \leq 1 \\ \frac{1}{2Me}, & M \geq 1 \end{cases}.$$

Данный результат может быть получен следующим образом.

Во-первых, заметим, что ожидание эмпирического риска может быть равно нулю, только если распределения классов не пересекаются, т. е. $p_j^0 p_j^1 = 0$, $j = \overline{1, n}$.

Во-вторых, ошибочное решение возможно только для тех значений X , в которые не попало ни одной выборочной точки, причем вероятность такого события для рассматриваемого асимптотического приближения находится из распределения Пуассона, а вероятность ошибки при условии попадания в такое значение равна $\frac{1}{2}$. Таким образом, вероятность ошибки всецело (считаем, что априорные вероятности классов равны) определяется априорным распределением в X .

Для нахождения распределения, при котором ожидаемая вероятность ошибки максимальна, теперь воспользуемся фактом, доказанным в [Неделько, 2003], а именно, тем, что максимум смещения риска достигается на кусочно-постоянном распределении, с двумя (для рассматриваемого случая) областями постоянства. Более того, исследование показывает, что одна из областей имеет минимально возможный размер.

Таким образом, для искомого распределения в одной из точек X сосредоточена вероятность $1 - \alpha$, а по остальным равномерно распределена вероятность $0 < \alpha \leq 1$.

Тогда (после предельного перехода) ожидаемая вероятность ошибки будет $\frac{1}{2} \alpha e^{-\alpha M}$.

Максимум данного выражения достигается при $\alpha = \min\left(\frac{1}{M}, 1\right)$, откуда и получаем искомое смещение риска.

Отношение $\frac{S'_V(0)}{S_Q(0)} = \frac{2Me(1 - e^{-M}) \ln 2}{M} \xrightarrow{M \rightarrow \infty} 2e \ln 2 \approx 3,77$ показывает степень

погрешности оценок В-Ч, которая, в основном, обусловлена приближением вероятности суммы суммой вероятностей, а также заменой энтропии на емкость.

Вклад последнего фактора можно оценить. Для этого заметим, что для найденного «худшего» распределения доля «непустых» значений переменной X при $M > 1$ на самом деле равна $1 - e^{-1}$,

откуда получаем: $\frac{S'_V(0)}{S_Q(0)} = 2(e - 1) \ln 2 \approx 2,38$.

Сравнение эффективности функционалов

Исчерпывающее сравнение эффективности функционалов в настоящее время не представляется возможным, поскольку на их основе пока не построены оптимальные оценки доверительного интервала для вероятности ошибки. Поэтому будем проводить сравнение на некоторых частных примерах. Нашей целью будет показать, что существуют такие ситуации, когда функционал эмпирического риска оказывается эффективнее. Для этого нам достаточно рассмотреть ряд весьма упрощенных модельных задач.

Пусть в дискретной задаче классификации, сформулированной в предыдущем разделе, распределение в пространстве D задается одним параметром $0 \leq \xi \leq 0,5$, а именно:

$$P(x = j, y = 0) = \frac{\xi}{n}, \quad P(x = j, y = 1) = \frac{1 - \xi}{n}, \quad j = \overline{1, n}.$$

Параметр ξ определяет так называемый байесовский уровень ошибки, т. е. вероятность ошибочной классификации при оптимальной решающей функции.

Решающая функция минимизирует эмпирический риск, т. е.

$$f(x) = \arg \max_{y \in \{0,1\}} \left| \left\{ (x^i, y^i) \in v \mid x^i = x, y^i = y \right\} \right|.$$

В случае, если при заданном x число выборочных точек обоих классов одинаково, в качестве решения принимаются с равной вероятностью 0 и 1.

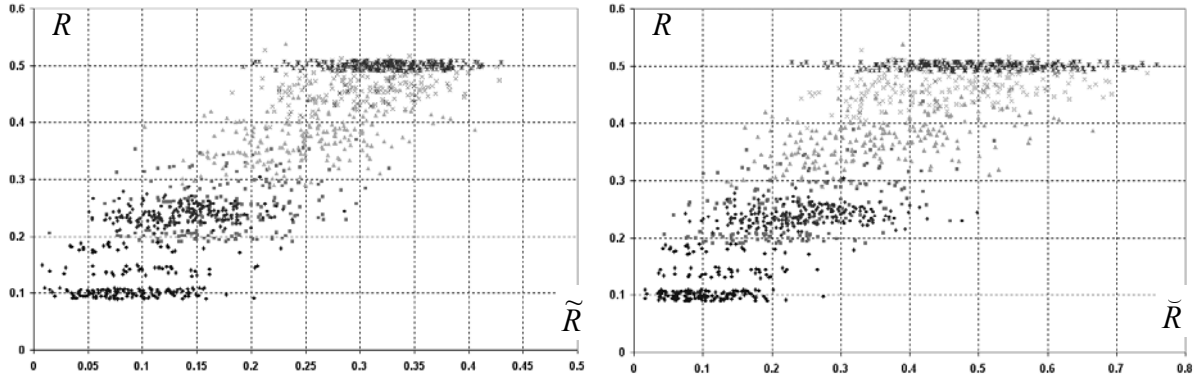


Рис. 1. Точки изображают полученные моделированием случайных выборок пары значений риска и эмпирического риска (левый график), а также риска и функционала скользящего экзамена (правый график). Разные типы маркеров соответствуют значениям параметра ζ : 0,1, 0,2, 0,3, 0,4, 0,5. Объем выборки $N = 50$, мощность пространства $n = 10$.

На рис. 1 приведены результаты моделирования случайных выборок. Видно, что дисперсия функционала скользящего экзамена больше дисперсии эмпирического риска, однако визуально сложно оценить, какой из функционалов позволяет более точно прогнозировать риск.

На практике распределение c неизвестно, поэтому для сравнения эффективности функционалов нужно на их основе построить наилучшие оценки доверительных интервалов. Для рассмотренной дискретной постановки эта задача, по-видимому, решается, однако, достаточно нетривиальна (см. далее), и к настоящему времени решение получить не удалось.

Однако, поставленная нами цель позволяет сделать упрощение и предположить, что c определяется одним параметром ζ и, более того, само случайно, задав при этом некоторое распределение на множестве значений параметра ζ .

Тогда определено полное совместное распределение величин R , \tilde{R} и \check{R} .

Критерием качества (эффективности) функционала \tilde{R} примем

$$\tilde{K} = \int \left(\min_{R^* \in [0,1]} \int |R - R^*| dP(R/\tilde{R}) \right) dP(\tilde{R}).$$

Аналогично

$$\check{K} = \int \left(\min_{R^* \in [0,1]} \int |R - R^*| dP(R/\check{R}) \right) dP(\check{R}).$$

Содержательно данные критерии отражают среднюю точность (в смысле модуля отклонения) прогнозирования риска при использовании соответствующего выборочного функционала.

Можно также измерять погрешность через среднеквадратичное отклонение:

$$\tilde{D} = \sqrt{\int \left(\int (R - E_{\tilde{R}} R)^2 dP(R/\tilde{R}) \right) dP(\tilde{R})},$$

где $E_{\tilde{R}} R$ – условное математическое ожидание риска при заданной величине эмпирического риска.

Аналогично вводится \tilde{D} – среднеквадратичный критерий качества для скользящего экзамена.

На рис. 2 изображены зависимости качества функционалов эмпирического риска и скользящего экзамена от объема выборки при $n = 3$. При этом параметр ζ равновероятно принимает значения 0 и 0,5. Видно, что различие эффективности функционалов в рассмотренном примере невелико, однако, существуют значения параметров, при которых функционал эмпирического риска достоверно эффективнее функционала скользящего экзамена.

Совместное использование функционалов

Оценивать риск можно и при одновременном использовании обоих функционалов.

$$K^{++} = \int \left(\min_{R^* \in [0,1]} \int |R - R^*| dP(R/\tilde{R}, \tilde{R}) \right) dP(\tilde{R}, \tilde{R}).$$

В этом случае эффективность оценивания риска может быть несколько выше, чем при использовании любого из функционалов в отдельности. Это демонстрируется третьей (наиболее светлой) кривой на рис. 2.

Более общим случаем метода скользящего экзамена является отделение для контроля не по одному объекту выборки, а выбор всех сочетаний из k объектов.

Естественно ожидать, что более точная оценка риска может быть получена при использовании полного набора функционалов скользящего экзамена, включая все допустимые значения параметра k .

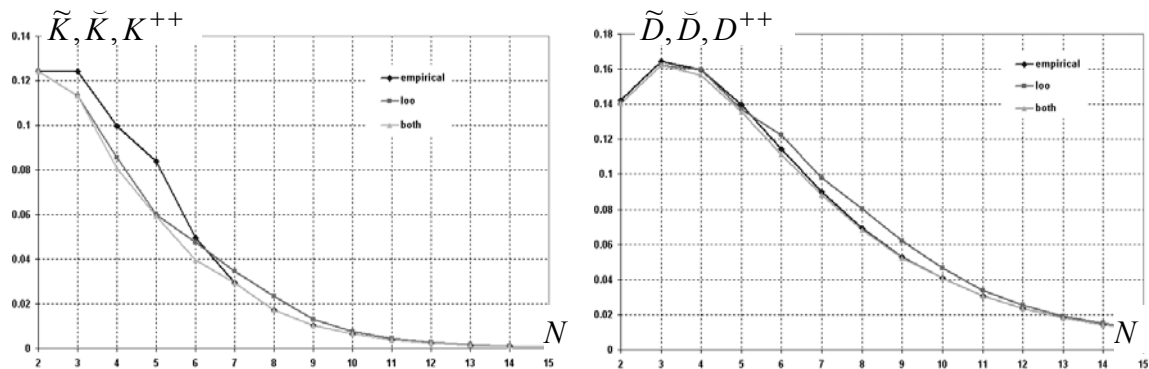


Рис. 2. Сравнение эффективности функционалов эмпирического риска и скользящего экзамена при различном объеме выборки, $n = 3$.

Построение доверительного интервала

Чтобы построить доверительный интервал для риска на основе некоторого функционала, например, эмпирического риска, нужно выбрать критическое множество из множества пар (\tilde{R}, R) .

Зададим критическое множество в виде

$$\{(\tilde{R}, R) \mid R > \hat{R}_\eta(\tilde{R})\},$$

тогда доверительным интервалом будет интервал $[0, \hat{R}_\eta(\tilde{R})]$.

Функция $\hat{R}_\eta(\tilde{R})$ определяется из условий:

$$\forall c, P(R > \hat{R}_\eta(\tilde{R})) \leq \eta, \int_0^{\tilde{R}_{\max}} \hat{R}_\eta(\tilde{R}) d\tilde{R} \rightarrow \min.$$

Первое из условий отражает требование, что вероятность попадания в критическую область не должна превышать заданного уровня значимости. Второе условие отражает предпочтительность по возможности меньших величин прогнозируемого риска. Точное выражение данного условия может корректироваться в зависимости от практических требований к прогнозу.

Точное решение возникающей оптимизационной задачи до сих пор не найдено, но известно приближенное решение, предложенное Вапником и Червоненкисом, которые, однако, не формулировали оптимизационную задачу в представленном виде, а подбирали $\hat{R}_\eta(\tilde{R})$ эмпирически.

Заключение

В работе рассмотрен ряд случаев, когда функционал эмпирического риска эффективнее скользящего экзамена. Само по себе существование таких случаев представляется интересным и говорит о том, что при условии точной оценки смещения эмпирический риск может конкурировать с несмещенной оценкой скользящего экзамена.

Обнаруженное преимущество эмпирического риска наблюдалось лишь в некоторых случаях и было несущественным, поэтому на практике все же более предпочтительным остается использование (при наличии такой возможности) функционала скользящего экзамена.

Задача построения оптимального функционала оценки качества в настоящее время остается открытой.

Литература

- [Вапник, Червоненкис, 1974] Вапник В. Н., Червоненкис А. Я. Теория распознавания образов. М.: Наука, 1974. 415 с.
- [Лбов, Старцева, 1999] Г.С. Лбов, Н.Г. Старцева. Логические решающие функции и вопросы статистической устойчивости решений. Институт математики СО РАН, Новосибирск, 1999, 211 с.
- [Неделько, 2003] V. M. Nedelko. Estimating a Quality of Decision Function by Empirical Risk // LNAI 2734. Machine Learning and Data Mining in Pattern Recognition. Third International Conference, MLDM 2003, Leipzig. Proceedings. Springer-Verlag. pp. 182–187.
- [Неделько, 2004] Неделько В. М. Оценивание смещения эмпирического риска для линейных классификаторов. // Таврический вестник информатики и математики. Изд-во НАН Украины 2004, № 1. С. 47–53.

Информация об авторе

Виктор Михайлович Неделько – с.н.с. лаборатории Анализа данных Института математики СО РАН, 660090, пр-т Коптюга, 4, Новосибирск, Новосибирский госуд. Университет, Россия, e-mail: nedelko@math.nsc.ru

ПРОГНОЗИРОВАНИЕ РАЗНОТИПНОГО ВРЕМЕННОГО РЯДА МЕТОДОМ АДАПТИВНОГО ФОРМИРОВАНИЯ ПРОСТРАНСТВА СОСТОЯНИЙ В КЛАССЕ ЛОГИЧЕСКИХ РЕШАЮЩИХ ФУНКЦИЙ¹

Светлана Неделько

Abstract: *The method of heterogeneous multidimensional time series probabilistic model reconstruction based on adaptive forming a discrete state set is offered. To estimate a deciding function quality some kind of informativity criterion for conditional distribution is used. The algorithm based on the proposed method was implemented and tested on model and applied tasks.*

Keywords: *multidimensional heterogeneous time series, pattern recognition, classification, statistical robustness, deciding functions, complexity.*

ACM Classification Keywords: *G.3 Probability and statistics: time series analysis, Markov processes, multivariate statistics, nonparametric statistics; G.1.6. Numerical analysis: Optimization.*

Введение

Для многих методов прогнозирования многомерного разнотипного временного ряда имеет место проблема резкого увеличения размерности пространства, в котором ведется поиск решающей функции, при увеличении глубины предыстории, от которой зависит прогноз. В этой ситуации приходится либо упрощать класс решающих функций и уменьшать глубину предыстории, либо использовать различные эвристики и усложнять алгоритмы построения решения.

Кроме того, одна из особенностей задачи прогнозирования многомерного временного ряда состоит в том, что прогнозируется одновременно несколько переменных. В большинстве алгоритмов эта особенность не принимается во внимание, и решающие функции строятся для каждой переменной отдельно, без учета зависимости целевых переменных.

В настоящей работе рассматривается алгоритм прогнозирования многомерного разнотипного временного ряда, основанный на адаптивном формировании пространства состояний в классе логических решающих функций. Поиск оптимального с точки зрения критерия информативности разбиения исходного пространства переменных учитывает зависимость целевых переменных и снимает проблему роста размерности пространства, хотя и огрубляет прогноз.

Постановка задачи

Пусть дан n -мерный разнотипный временной ряд $v = \{z^t \mid t = \overline{1, N}\}$, $z^t = (z_1^t, \dots, z_n^t)$, $z_j^t \in Z_j$. Здесь Z_j – множество допустимых значений j -й переменной ряда. В наборе переменных могут присутствовать одновременно непрерывные и дискретные, а также переменные с упорядоченным и неупорядоченным множеством значений. Пространство значений ряда обозначим $Z = \sum_{j=1}^n Z_j$.

Задача состоит в том, чтобы на основе анализа имеющихся эмпирических данных v прогнозировать значения временного ряда в моменты времени $t > N$.

¹ Работа выполнена при поддержке РФФИ, грант 04-01-00858-а.

Будем рассматривать статистическую постановку задачи, когда ν является реализацией некоторого случайного процесса $z(t)$ с дискретным временем. При этом предположим, что процесс задается переходной (условной) вероятностной мерой $P[Z/z(t-1), z(t-2), \dots, z(t-d)]$, определяемой предысторией длины d . Квадратные скобки здесь и далее в аналогичных ситуациях означают, что имеется в виду не мера множества Z , а мера, заданная на некоторой σ -алгебре его подмножеств.

Обычно прогнозирование временного ряда подразумевает построение решающей функции $f: Z^d \rightarrow Z$, которая по заданной предыстории ряда для моментов $t-1, t-2, \dots, t-d$ дает прогнозируемый набор значений переменных ряда для момента t . Мы будем строить аппроксимацию самой переходной вероятности. Этот подход имеет потенциально большую гибкость при оценивании модели, определяющей временной ряд. Знание переходной вероятности позволяет, в том числе, и строить оптимальный прогноз значений.

При этом, классические непараметрические методы оценивания условного распределения требуют относительно большого объема эмпирических данных и предполагают те или иные метрические свойства пространства Z , что в разнотипном случае не вполне оправдано.

В настоящей работе используется основанный на критерии информативности метод оценивания условного распределения в заданном классе кусочно-постоянных распределений.

Данный метод не требует каких-либо метрических свойств пространства Z и позволяет гибко подстраивать сложность модели под объем выборки.

Критерий качества вероятностной модели

Описываемый критерий основан на понятии информативности распределений. Под информативностью здесь понимается степень отличия от априорного распределения. Степень отличия может характеризоваться расстоянием в некоторой выбранной метрике (однако требование выполнения всех свойств метрики, вообще говоря, необязательно).

Зафиксируем некоторое разбиение $\lambda = \{E^\omega \subseteq Z \mid \omega = \overline{1, k}\}$, $\bigcup_{\omega=1}^k E^\omega = Z$, $\omega \neq \varpi \Rightarrow E^\omega \cap E^\varpi = \emptyset$, пространства Z .

Теперь исходному многомерному ряду ν можно сопоставить одномерную символьную последовательность $w = \left\{ \omega^t \mid z^t \in E^{\omega^t}, t = \overline{1, N} \right\}$.

Случайному процессу $z(t)$ будет соответствовать процесс $\omega(t)$, переходные вероятности для которого обозначим

$$P_{\omega_0 | \omega_1, \omega_2, \dots, \omega_d} = P(\omega(t) = \omega_0 / \omega(t-1) = \omega_1, \dots, \omega(t-d) = \omega_d).$$

Также будем использовать совместную вероятность $P_{\omega_0 \dots \omega_d} = P\left(\bigwedge_{\tau=0}^d (\omega(t-\tau) = \omega_\tau)\right) = P\left(\bigwedge_{\tau=0}^d (Z(t-\tau) \in E^{\omega_\tau})\right)$ — вероятность заданной предыстории длины d .

Критерий информативности определим как

$$K(\lambda) = \sum_{\omega_0=1}^k \dots \sum_{\omega_d=1}^k \left| P_{\omega_0} - P_{\omega_0 | \omega_1, \omega_2, \dots, \omega_d} \right| \cdot P_{\omega_1 \omega_2 \dots \omega_d}.$$

Данный критерий есть средний модуль разности между вероятностями перехода и безусловными вероятностями нахождения в состояниях.

После тождественных преобразований имеем:

$$K(\lambda) = \sum_{\omega_0=1}^k \dots \sum_{\omega_d=1}^k \left| P_{\omega_0 \dots \omega_d} - \left(\sum_{\omega_0=1}^k P_{\omega_0 \omega_1 \dots \omega_d} \right) \left(\sum_{\omega_1=1}^k \dots \sum_{\omega_d=1}^k P_{\omega_0 \omega_1 \dots \omega_d} \right) \right|.$$

Для оценки данного критерия по выборке достаточно заменить $P_{\omega_0 \dots \omega_d}$ на $N_{\omega_0 \dots \omega_d} / N$ – частоту реализации предыстории на обучающей последовательности w .

Описание алгоритма

Во многих распространенных алгоритмах прогнозирования временного ряда производится сведение реализации временного ряда к выборке в виде таблицы данных так называемым методом «гусеницы» или «змейки» [Данилов, Жиглявский, 1997]. Для этого вводится переобозначение переменных: прогнозируемые переменные обозначаются $y_j(t) = z_j(t)$, $j = 1, \dots, n$, а значения ряда на предыстории обозначаются $x_j(t) = z_j(t-1)$, $x_{j+n}(t) = z_j(t-2)$, ..., $x_{j+n(d-1)}(t) = z_j(t-d)$, $j = \overline{1, n}$. После этого могут использоваться алгоритмы построения решающих функций на основе таблиц данных. Для случая разнотипного пространства независимых переменных известны соответствующие методы распознавания образов и регрессионного анализа в классе логических решающих функций [Лбов, Старцева, 1999].

Однако подобное сведение существенно (в d раз) увеличивает размерность пространства, в котором строится решающая функция.

В предлагаемом алгоритме разбиение строится непосредственно в пространстве Z , что позволяет избежать увеличения размерности и строить решения при относительно коротких реализациях

Для нахождения приближенного к оптимальному в соответствии с критерием K разбиения λ применим алгоритм направленного поиска (LRP), строящий решение в виде дерева [Лбов, Старцева, 1999] или в виде непересекающихся многомерных интервалов. Под интервалом понимается произвольное множество соседних значений для переменной с упорядоченными значениями и произвольное подмножество значений, если переменная с неупорядоченными значениями. Многомерный интервал представляет собой декартово произведение интервалов по переменным.

Тестовый пример

Для иллюстрации работоспособности изложенного алгоритма, решим модельную задачу.

Пусть двумерный временной ряд задается случайным процессом с тремя состояниями. Области пространства, соответствующие состояниям, и вероятности переходов приведены на рис. 1. В каждой из областей при условии попадания в нее распределение равномерно.

Реализованный алгоритм построения разбиения на основе критерия информативности продемонстрировал правильное нахождение закономерностей, заложенных во временной ряд. Построенное им дерево решений изображено в правой части рисунка.

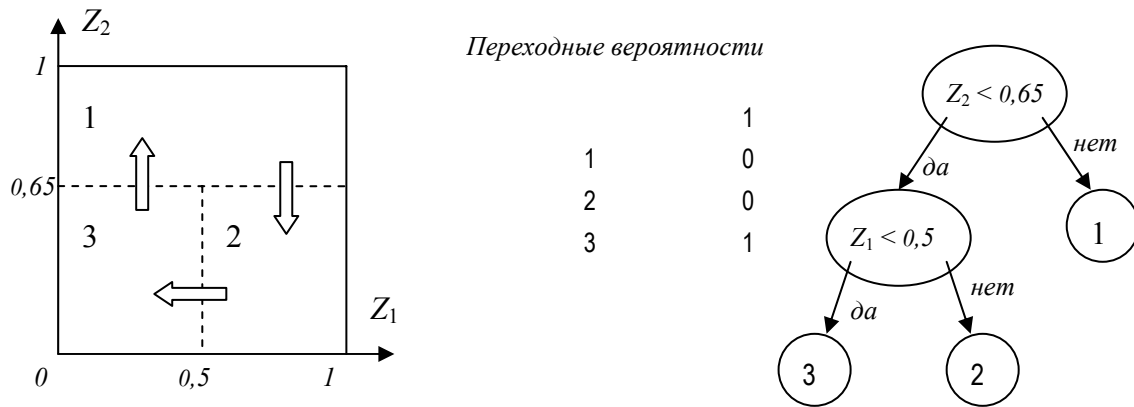


Рис. 1. Для тестового примера изображены соответствующие состояниям области в пространстве значений переменных, переходные вероятности и дерево решений, построенное предложенным алгоритмом.

Решение прикладной задачи

Предложенный алгоритм был испытан на задаче анализа временного ряда, представленного метеорологическими данными.

Ряд включал за период с 1915 по 2000 гг. среднегодовые значения следующих величин:

- Z_1 – температура воздуха;
- Z_2 – объем осадков;
- Z_3 – объем водостока.

Траектория данного ряда в проекции на плоскость Z_2, Z_3 приведена на рис. 2.

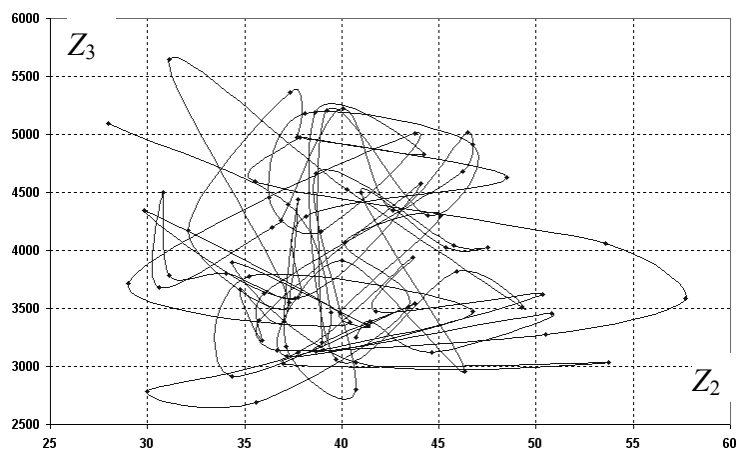


Рис. 2. Траектория временного ряда в пространстве переменных Z_2 и Z_3 .

Для выявления закономерностей данного ряда с помощью описанного алгоритма было построено дерево решений с 6-ю конечными вершинами, которые определяют состояния случайного процесса. Глубина предыстории $d = 1$.

Ниже полученное решение представлено в виде списка областей.

- 1: $Z_3 < 3960 \ \& \ Z_2 < 37.8 \ \& \ Z_3 < 3163$
- 2: $Z_3 < 3960 \ \& \ Z_2 < 37.8 \ \& \ Z_3 \geq 3163 \ \& \ Z_1 < -0.46$
- 3: $Z_3 < 3960 \ \& \ Z_2 < 37.8 \ \& \ Z_3 \geq 3163 \ \& \ Z_1 \geq -0.46$
- 4: $Z_3 < 3960 \ \& \ Z_2 \geq 37.8$
- 5: $Z_3 \geq 3960 \ \& \ Z_2 < 37.8$
- 6: $Z_3 \geq 3960 \ \& \ Z_2 \geq 37.8$

безусловн..		1	2	3	4	5	6
0.08	1	0.29	0	0	0.71	0	0
0.12	2	0.1	0.4	0	0	0.4	0.1
0.06	3	0.2	0.2	0	0.6	0	0
0.3	4	0.12	0	0.08	0.54	0.04	0.19
0.15	5	0	0.23	0.08	0.08	0.38	0.23
0.29	6	0	0.08	0.08	0.12	0.08	0.64

В таблице приведены оценки безусловных вероятностей нахождения в соответствующих состояниях (первый столбец), а также переходных вероятностей. Значение критерия для полученного решения $K = 0,8$.

Как можно заметить, переходные вероятности существенно отличаются друг от друга и от безусловных вероятностей. Это свидетельствует о том, что предложенный алгоритм может находить закономерности в многомерных временных рядах.

Заключение

Методы одновременного прогнозирования нескольких целевых переменных при решении задачи прогнозирования многомерного разнотипного временного ряда, по сравнению с методами построения решающей функции по каждой целевой переменной в отдельности, позволяют точнее учитывать взаимозависимости переменных. В изложенном в работе методе в качестве решающей функции выступает матрица переходов для состояний случайного процесса, соответствующих адаптивно подобранному разбиению пространства переменных. При этом разбиение строится непосредственно в пространстве переменных, описывающих временной ряд, что существенно снижает трудоемкость алгоритма. Критерием качества решения в таком методе является информативность матрицы переходов.

Применение предложенного метода к анализу метеорологических данных продемонстрировало эффективность метода и возможность получения закономерностей, представляющих интерес для дальнейшего содержательного анализа.

Литература

- [Лбов, Старцева, 1999] Г.С. Лбов, Н.Г. Старцева. Логические решающие функции и вопросы статистической устойчивости решений. Институт математики СО РАН, Новосибирск, 1999, 211 с.
- [Ростовцев, 1978] П.С. Ростовцев. Алгоритм построения типологий для больших массивов социально-экономической информации. // Модели агрегирования социально-экономической информации. Сборник научных трудов, изд. ИЭ и ОПП СО АН СССР, 1978.
- [Lbov, Nedel'ko, 2001] G.S. Lbov, V.M. Nedel'ko. A Maximum informativity criterion for the forecasting several variables of different types. // Computer data analysis and modeling. Robustness and computer intensive methods. Minsk, 2001, vol 2, 43–48.
- [Неделько, 2004] С. В. Неделько. Критерий информативности матрицы переходов и прогнозирование многомерного разнотипного временного ряда. // Искусственный интеллект, № 2, 2004, с. 145–149.
- [Данилов, Жиглявский, 1997] Д.Л. Данилов, А.А. Жиглявский. Главные компоненты временных рядов: метод «Гусеница». Санкт-Петербургский Государственный Университет, 1997.

Author's Information

Svetlana Valeryevna Nedel'ko – Institute of Mathematics SB RAS, Laboratory of Data Analysis, 630090, pr. Koptyuga, 4, Novosibirsk, Russia, e-mail: nedelko@math.nsc.ru

DETECTION OF LOGICAL-AND-PROBABILISTIC CORRELATION IN TIME SERIES¹

Tatyana Stupina

Abstract. An application of the heterogeneous variables system prediction method to solving the time series analysis problem with respect to the sample size is considered in this work. It is created a logical-and-probabilistic correlation from the logical decision function class. Two ways is considered. When the information about event is kept safe in the process, and when it is kept safe in depending process.

Keywords: the prediction of heterogeneous variables system, the adaptive method, multidimensional time series, logical decision function.

ACM Classification Keywords: G.3.

Introduction

The problem of detection correlations by data, which is presented by time series, is used in different intellectual analysis domains. We have the most difficult problem, when any prior information about process or object is absent. In addition to that several attendant problems are appeared. Firstly, it is necessary to define a class of decision functions (models). Secondary, we must work up a method of plotting optimal decision function, in other words to define optimality criterion by sample. In the third place, we must test our model on adequacy and effectiveness (capacity for general conclusion or statistical stability).

At present time there are many well-known scientific schools, what make researches to that line of investigation [Lukashin Y.P., 2003, Bezruchko B.P., Smirnov D.A., 2003, Lbov G.S., Starceva N.G., 1999]. However universal method is not exists. Several suppositions and priory information are used by every method. It says that problem is actual problem. The method is preferred if it uses lame suppositions to respect with decision function class and if it has capability to retarget model during learning on sample data. At present time so methods use the neural-network technological, the pyramidal-network, the wavelet analysis, the logical structures and others approaches. Such methods we can name as adaptive methods. The conception of adaptive has more comprehensive sense [Lukashin Y.P., 2003, Lbov G.S., Starceva N.G., 1999].

We will interpret concept of adaptive as consecutive selection of model's structure during process of learning on sample data in order to take effective prediction by time series. At the same time it is appeared additional problem – detection a time moment of changing model's structure (criterion of adaptive).

In this paper one is suggested two ways to joint analysis of several univariate time series by using MLRP-method. When the information about event is kept safe in the process, and when it is kept safe in depending process. That method was applied to prediction of multivariable heterogeneous time series [Stupina T.A., Lbov G.S., 2006]. The solving of practice problem from hydrological domain is presented here by MLRP-method. Model from the logical decision function class we will name as a logical-and-probabilistic correlation [Lbov G.S., Starceva N.G., 1999].

Problem Statement

Let us consider terminal time series $\{x(t), t \in T\}$, it is realization of some time-dependent random process $\eta(t)$. One is supposed that simultaneous distribution $p(\eta_1)$, $p(\eta_1, \eta_2)$, $p(\eta_1, \eta_2, \eta_3), \dots$, $p(\eta_1, \dots, \eta_T)$ is exist. The value set $D_{\eta(t)}$ of variables may be quantitative, nominal and ordinal type in a more case. Let the values of random process $\eta(t)$ are measured at consequent moments of the time with the gap $\Delta t = t_k - t_{k-1}$. Denote this set of moments as $T = \{t_1, \dots, t_k, \dots, t_N\}$, $N \ll \infty$.

¹ This work was financially supported by Lavrentiev's Grant № 7 of Youth Science Concours and RFBR 07-01-00331-a

Classical problem of prediction time series is consist in that we must take prediction at time moment $t = t_R$ on time period $t_{R+\tau}$ by analyzing prehistory $\{x(t_k)\}, k = 1, \dots, d$, with length d . As a rule the value τ is named as forestalling. Let us denote the set of every possible prehistory, that have length d , as a D_X , and the set of every possible all forestalling sets as a D_Y . Let us understand a prediction decision function as a f mapping of the D_X set on the D_Y set, i. e. $f: D_X \rightarrow D_Y$, $\dim D_X = d$, $\dim D_Y = \tau$. Model's construction f of prediction is defined by decision function class Φ .

If the simultaneous distribution is known than optimal decision function, constructing predicts to time $t + \tau$, is conditional average of distribution $E(\eta_{t+\tau} | \eta_{t-d}, \dots, \eta_t)$. In order to solve this problem it is necessary to restore conditional distribution. But that way is not practical because we have not enough size of sample in applied tasks. Therefore it is possible to offer a different depending on specified suggestions targets setting (concerning properties of random process) and the different methods (concerning decision function class) of their decision accordingly.

At present time it was developed many method for prediction depended on time random process (probabilistic characteristics of process are not changed on time). Its methods are based on constructing several models, which usually use some suggestion. For example, if we want to do long-time prediction than the best offer (concerning error variance value) is global model, if we want to do short-term prediction, than the best choice is a local model [Bezruchko B.P., Smirnov D.A., 2003]. Note that most models accomplish solitary prediction and as a rule it is at next time $t + \Delta t$ or at time moment $t + \tau \Delta t$, $\tau = 2, \dots, N - d$.

We propose model, that accomplishes prediction on all forestalling term τ , in other words, to time moments $t + k \Delta t$, $k = 1, \dots, \tau$. That prediction allows to take one decision function (structure of model) and to do simultaneously several predictions on future by one prehistory.

For that problem statement it is important to analyze several steps:

- The detection time moment of changing model's structure (criterion of adaptive);
- The optimization of prehistory length d ;
- The optimization of forestalling term τ .

In order to solve these items we will use class of logical decision function. We will consider two ways: a) when the information about event is kept safe in the process, and b) when it is kept safe in depending process. We will perform the primary ideas of these ways in following paragraphs.

Analysis univariate time series problem

Let we have univariate time series $\{x(t)\}$ of any random process $\eta(t)$. It is necessary to solve a problem of constructing function f by empirical data, that is presented as terminal points N for given prehistory length d and forestalling term τ . We will construct decision function from the logical decision function Φ_M by sample data, which is made from points of discrete time series. The procedure of building data table $v = \{v_x, v_y\}$ depends on problem statement and on data generally.

For example, it may be

- a) Shift of prehistory window step-by-step on time series,
- b) Shift of prehistory window to some position on time series,
- c) Building prehistory window from the series points, that is positioned on some distance.

Also we can consider some combination of items indicated above. The visual illustration of the univariate time series and principle of building sample table are presented on figure.1

Not lose commonality let us consider, that $\Delta t = 1$, then prehistory table is built as $v_x = \{x_{kj}\} = \{x(j+k-1)\}$, where $j = R - d + 1, \dots, R$, $d \leq R \leq N - \tau$, $k = 1, \dots, N - R - \tau$, and forestalling term table (future predictions) is built as a $v_y = \{x_{kj}\} = \{x(j+k)\}$, where $j = R + 1, \dots, R + \tau$, $k = 1, \dots, N - R - \tau$, for case (a) as above. With the

help of data table $v = \{v_x, v_y\}$ of the size $N - R - \tau$ we will construct sample decision function \bar{f} from the class Φ_M by the MLRP-method. So we have that choice of optimal length d^* of prehistory corresponds to the choice of informative characteristic subset. A choice of optimal forestalling term length τ^* will be correspond to definition of likely problem size (complexity) for a given sample size.

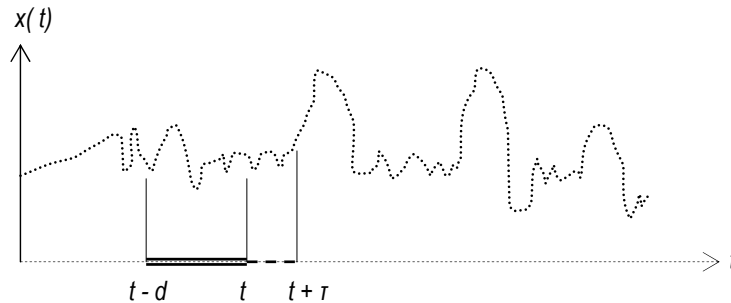


fig.1. Analysis univariate time series.

We define a time moment of changing model's structure (adaptive) as a time moment $t^* = t_{R-d}$ for which the condition $|F(\bar{f}) - F^*| \geq h$ is carried out, where the value F^* is threshold value of model quality, h is admissible value of deviation for established quality.

Below we will consider logical decision function class Φ_M and its properties. We will define criterion of quality $F(f)$ for decision function f .

The prediction with respect to other time series

This paragraph is devoted to detection of correlation between two univariate time series. That problem statement is well known and is commonly applied for solving practice problem [Bezruchko B.P., Smirnov D.A., 2003]. However the most methods indicate some power of correlation for the given time point.

The suggested method is founded on constructing function for that a definitional domain is assigned in domain of realizations of one time series $\{x(t)\}$, and a value domain (domain of prediction point $t + \tau \Delta t$, $\tau = 1, \dots, N - d$) is assigned in domain of realizations of other time series $\{y(t)\}$. It is supposed that one time series with respect to other process. The visual illustration of two dependent time series and the principle of building sample table are presented on figure.2.

The data table is constructed by principle like above. The power of correlation f is defined by quality value $F(f)$ a) on the learning sample and b) on the control sample. We will construct sample decision function \bar{f} (logical-and-probabilistic correlation) from the logical decision function class Φ_M by the MLRP-method like above.

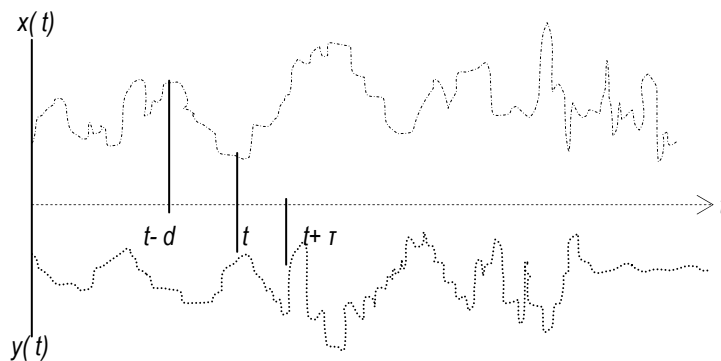


fig.2. Two dependent time series.

MLRP-method of creating logical-and-probabilistic model

From the beginning we consider a commonly probabilistic statement problem. Let the value (x,y) is a realization of a multidimensional random variable (X,Y) on a probability space $\langle \Omega, B, P \rangle$, where $\Omega = D_X \times D_Y$ is μ -measurable set (by Lebeg), B is the borel σ -algebra of subsets of Ω , P is the probability measure (we will define such as c , the strategy of nature) on B , D_X is heterogeneous domain of under review variable, $\dim D_X = n$, D_Y is heterogeneous domain of objective variable, $\dim D_Y = m$. The given variables can be arbitrary types (quantitative, ordinal, and nominal). For the pattern recognition problem, for example, the variable Y is nominal. Let us put Φ_0 is a given class of decision functions. Class Φ_0 is μ -measurable functions that puts some subset of the objective variable $E_Y \subseteq D_Y$ to each value of the under review variable $x \in D_X$, i.e. $\Phi_0 = \{f : D_X \rightarrow 2^{D_Y}\}$. For example the domain E_Y can contains the several patterns $\{\omega_1, \dots, \omega_k\}$ for pattern recognition problem.

The quality $F(c,f)$ of a decision function $f \in \Phi_0$ under a fixed strategy of nature c is determined as $F(c,f) = \int_{D_X} (P(E_Y(x)/x) - \mu(E_Y(x))) dP(x)$, where $E_Y(x) = f(x)$ is a value of decision functions in x , $P(y \in E_Y(x)/x)$ is a conditional probability of event $\{y \in E_Y\}$ under a fixed x , $\mu(E_Y(x))$ is measurable of subset E_Y . Note that if $\mu(E_Y(x))$ is probability measure, than criterion $F(c,f)$ is distance between distributions. If the specified probability coincides with equal distribution than such prediction does not give information on predicted variable (entropy is maximum). On the nominal-real space $\Omega = D_H \times D_\theta$ a measure μ is defined so as any $E \in B$, $E = \bigcup_{j=1}^{|E_H|} E_\theta^j \times \{z^j\}$, $\mu(E) = \sum_{j=1}^{|E_H|} \frac{\mu(E_\theta^j)}{|D_H| \mu(D_\theta)}$, were E_H is projection of set E on nominal space D_H , z^j - item of E_H , E_θ^j - set in D_θ corresponding to z^j , $\mu(E_\theta^j)$ - lebeg measure of set E_θ^j . For any subset of domains D_X or D_Y the measure μ is assigned similarly. Clearly, the prediction quality is higher for those E_Y whose measure is smaller (accuracy is higher) and the conditional probability $P(y \in E_Y(x)/x)$ (certainty) is larger. For a fixed strategy of nature c , we define an optimal decision function $f_0(x)$ such as $F(c,f_0) = \sup_{f \in \Phi_0} F(c,f)$, where Φ_0 is represented above class of decision functions.

In commonly when we solve this problem in practice the size of sample is small and type of variables may be different. In this case class of logical decision function Φ_M complexity M [Lbov G.S., Starceva N.G, 1999] is used. For the prediction problem of the heterogeneous system variables class Φ_M is defined as $\Phi_M = \{f \in \Phi_0 \mid f \sim \langle \alpha, r(a) \rangle, \alpha \in \Psi_M, r(\alpha) \in R_M\}$ (the mark ' \sim ' denotes the correspondence of pair $\langle \alpha, r(a) \rangle$ to symbol f), were Ψ_M is set of all possible partitioning $\alpha = \{E_X^1, \dots, E_X^M \mid E_X^t = \prod_{j=1}^n E_{X_j}^t, E_{X_j}^t \subseteq D_{X_j}, t = \overline{1, M}, \bigcup E_X^t = D_X\}$ of domain D_X on M noncrossing subsets, R_M is set all possible decisions $r(\alpha) = \{E_Y^1, \dots, E_Y^M \mid E_Y^t \in \mathfrak{S}_{D_Y}, t = \overline{1, M}\}$, \mathfrak{S}_{D_Y} - set of all possible m -measuring intervals. For that class the measure $\mu(E_Y(x)) = \frac{\mu(E_Y)}{\mu(D_Y)} = \prod_{j=1}^m \frac{\mu(E_{Y_j})}{\mu(D_{Y_j})}$ is the normalized measure of subset E_Y and it is introduced with taking into account the type of the variable. The measure $\mu(E_Y(x))$ is measure of interval, if we have a variable with ordered set of values and it is quantum of set, if we have a nominal variable (it is variable with finite non-ordering set of values and we have the pattern recognition problem). A complexity of Φ_M class is assigned as M if we have invariant prediction (decision is presented by form: if $x \in E_X^t$, than $y \in E_Y^t$), $M_\Phi = M$, and it is assembly (k_1, \dots, k_M) if we have multivariate, i.e. $E_Y^t = \bigcup_{i=1}^{k_t} E_Y^i$, $t = \overline{1, M}$ and $E_Y^i \cap E_Y^j = \emptyset$ for

$i \neq j$ (decision is presented by form: if $x \in E_x^t$, then $y \in E_y^1 \vee E_y^2 \vee \dots \vee E_y^{k_t}$). The class of logical decision function has universal property.

Statement. For any function $f \in \Phi^\circ$ and $\varepsilon > 0$ there are M and several logical decision function $f_M \in \Phi_M$ so that $|F(c, f) - F(c, f_M)| \leq \varepsilon$.

Others good properties of the logical decision function class are presented in work [Stupina T.A. 2006] for prediction system heterogeneous variables problem.

If the strategy of nature is unknown the sampling criterion $F(\bar{f})$ is used by method $Q(v_N)$ of constructing sample decision function \bar{f} , $\bar{F}(\bar{f}) = \sum_{t=1}^{M'} \bar{p}_x^t (\bar{p}_{y/x}^t - \bar{\mu}_y^t)$, where $\bar{p}_x^t = \frac{N(\hat{E}_x^t)}{N(D_x)} = \frac{N^t}{N}$, $\bar{p}_{y/x}^t = \frac{N(\hat{E}_y^t)}{N(\hat{E}_x^t)} = \frac{\hat{N}^t}{N^t}$, $\bar{\mu}_y = \mu(\hat{E}_y)$, $N^{(*)}$ is number of sample points, generating the set ** , $\bar{f} \sim \langle \alpha, r(\alpha) \rangle$, $\alpha = \{\hat{E}_x^1, \dots, \hat{E}_x^{M'}\} \in \Psi_{M'}$, $r(\alpha) = \{\hat{E}_y^1, \dots, \hat{E}_y^{M'}\} \in R_{M'}$. The optimal sample decision function is $\bar{f}^* = \arg \max_{\alpha \in \Psi_{M'}} \max_{r(\alpha) \in R_{M'}} \bar{F}(\bar{f})$. In order to solve this extreme problem we apply the algorithm MLRP of step-by-step increase attachments of decision trees. It do the branching of top point on that value criterion $\bar{F}(\bar{f})$ is maximum and the top point is divisible or $\bar{F}(\bar{f}) \geq F^*$. The top point is indivisible if 1) number of final top point is $M' = M^*$ or 2) $\hat{N}^t \leq N^*$. That criterion and parameters F^*, M^*, N^* assign method of constructing sample decision function.

In order to estimate the MLRP - method quality we did statistical modeling. The average of the criterion of sample decision function on samples of fixed size $m_F(c) = E_{v_N} F(c, \bar{f})$ is estimated for fixed nature strategy. Moreover we researched the averaging-out empirical functional quality $\varepsilon_N(c) = E_{v_N} F(c, \bar{f}) - E_{v_N} \bar{F}(\bar{f})$ for given strategy of nature with the purpose of estimating decision quality, and maximal removal of empirical functional quality average of distribution $\varepsilon_N^*(c) = \sup_{c: \bar{F}(\bar{f})=F_0} \varepsilon_N(c)$ for a given empirical quality value F_0 . It was taken for some

parametric nature strategy class, for given nature strategy complexity M , decision function complexity M' . The decision function is built by MLRP-method on sample of size N . Parameters n, m (dimensions of domains D_x and D_y) and number of fixed type variables were considered in problem statement on the whole. It's defined the complexity of nature strategy and complexity of decision function. The GenMLRP-algorithm was developed for modeling nature strategy parameters. Generation nature strategies were realized in accordance with definition, where parameters are established by random in the given interval. The properties of functional quality are presented in work [Stupina T.A., 2006] for uniform distribution on set D_y .

The MLRP-method was applied for prediction multivariate time series. Three random processes were simultaneously considered instead of one. Feature's systems (under review and predicted) were established. Procedure of building data table is offered in work [Stupina T.A., G. S. Lbov, 2006]. The example of solving practical problem is presented in next paragraph.

Application MLRP-method to prediction multivariate time series task

This paragraph is devoted to some of practical problem from hydrological area. It consists in the prediction of the columbine ($k = 1$), transmitting across Oby riverbed, the average monthly temperature ($k = 2$) and the atmospheric precipitates ($k = 3$) by like hydrometeor data (in the course of the 86 years) in control post of the city Kolpashevo. In order to construct decision function of prediction variable system (y_1, y_2, y_3) in April by variable system in the course November, December and January the average monthly data was worked up in the course of November ($i = 1$), December ($i = 2$), January ($i = 3$) and April ($j = 1$) in control post. The sample decision function was constructed by learning data $\{x_k(t_{k_i+12}), y_k(t_{k_j})\}$ of the size 76.

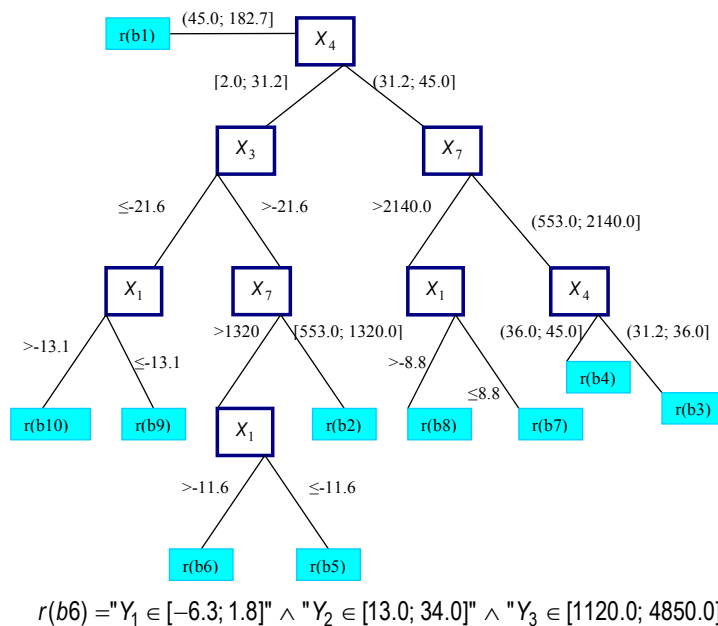


fig. 3. The decision tree of hydrological situations is presented.

Estimation of quality criterion (probability estimation of veritable decision by rule \bar{f}) was taken by control sample of the size 10 (it is last years of time series) and it was equal 0.8 that is satisfactory result.

Four important features were chosen from initial nine features for consecutive constructing decision tree (the decision function from logical decision function may be presented by dichotomous count). In compliance with construction decision function the average monthly temperature in November (x_1) and in January (x_2), the atmospheric precipitates in November (x_4), the columbine in November had the most influence on prediction quality. Visual illustration of some decision tree of hydrological situations and decision example on top $r(b6)$ is presented on figure.3 with complexity $M = 10$, $N^* = 3$.

Conclusion

In that paper two ways to solving analysis univariate time series problem was considered. It was founded on the MLRP-method constructing logical-and-probabilistic model for prediction heterogeneous variables system. The idea's approach and ways of realizations was formulated here.

The decision was constructed by MLRP-method from the logical decision function class. It allows taking optimal parameters as such prehistory length and forestalling term for univariate time series. Practical problem from hydrological area was decided by MLRP-method for prediction variable system. Simulation of the different type time series is matter of future researches.

We want to note that proposed approach to joint analyses of some time series can have more than enough applications. For example, we can solve problem of statistically important correlation detection between seismic processes arising on the most distant region. It allows us to understand and even perhaps to detect earthquake precursors.

Bibliography

- [Box Dj., Dgenkins G., 1974] Box Dj., Dgenkins G. Analysis time series. Prediction and direction. Publ. Moskow, Wold.– 1974 - 242 pp.
- [Lukashin Y.P., 2003] Lukashin Y.P Adaptive short time prediction methods for time series. Publ. Moskow, Finances and Statistic – 2003 – 416 pp.

- [Bezruchko B.P., Smirnov D.A., 2003] Bezruchko B.P., Smirnov D.A. The modern modeling by time series. <http://www.nonlinmod.sgu.ru/doc/review.pdf>.
- [Lbov G.S., Starceva N.G., 1999] Lbov G.S., Starceva N.G. Logical Decision Functions and Questions of Statistical Stability. Inst. of Mathematics, Novosibirsk.
- [Stupina T.A., Lbov G.S., 2006] T.A. Stupina. G. S. Lbov. Application of the multivariate prediction method to time series. International Journal ITHEA, Vol 13, No 3. - 2006 - pp.278-285.
- [Stupina T.A., 2005] Stupina T.A. Estimation of quality removal for prediction multivariate heterogeneous variable problem. Proceeding of the Russian conference "Mathematical methods of pattern recognition (MMPO-12)", Moscow, 20–26 November 2005 – pp. 209-212.
- [Stupina T.A., 2006] T.A. Stupina. Recognition of the Heterogeneous Multivariate Variable. Proceeding of the international conference, 2006 (KDS'2006), Varna (Bulgaria), Vol 1 – pp. 199-202.

Author's Information

Tatyana A. Stupina – Institute of Mathematics SBRAS, Koptuga 4 St, Novosibirsk, 630090, Russia;
e-mail: stupina@math.nsc.ru

ОДИН ПОДХОД К РАВНОВЕСИЯМ В ИГРАХ В УСЛОВИЯХ НЕОПРЕДЕЛЕННОСТИ

Сергей Мащенко

Abstract: There are proposed the necessary and sufficiency conditions of the virtual equilibria of games in the conditions of of uncertainties as in clean so in the mixed strategies.

Keywords: Nash's equilibria, a virtual equilibria, an uncertainty.

Введение

Игра в условиях неопределенности, несомненно, являются одной из интересных постановок задачи принятия решений в условиях конфликта. Рассмотрим игру $UG = (Y; X_i, u_i; i \in N)$ в условиях неопределенности, где $Y = \{y_1, y_2, \dots, y_m\}$ - для простоты конечное множество из m состояний природы; $N = \{1, 2, \dots, n\}$ - множество из n игроков; X_i - множество стратегий x_i игрока $i \in N$; $u_i(x, y)$ функция выигрыша игрока $i \in N$, которая определена на прямом произведении множества состояний природы Y и множества ситуаций игры $X = \prod_{i \in N} X_i$, т.е. $u_i : Y \times X \rightarrow R^1$. Каждый игрок стремится получить по возможности большее значения своей функции выигрыша.

В этой работе развивается один из наиболее привлекательных принципов оптимальности теории игр - равновесие Неша [1] для игры в условиях неопределенности. Следует напомнить, что для детерминированной игры в нормальной форме $(X_i, U_i; i \in N)$ ситуация $x^* = (x_i^*)_{i \in N}$ называется равновесием Неша, если каждый игрок, в отдельности, при выборе стратегии, отличной от x_i^* , получит выигрыш не лучше чем в ситуации равновесия.

Виртуальные равновесия

Поскольку для игры в условиях неопределенности выигрыши игроков зависят от состояния природы и определяются неоднозначно, будем для каждого игрока $i \in N$ рассматривать вектор выигрышей

$U_i(x) = (u_i(x, y))_{y \in Y}$ по множеству состояний природы Y . Такая не совсем обычная форма представления вектора связана с тем, что результаты этой работы легко обобщить на случай бесконечных множеств состояний природы. В этом случае вектор-функция $U_i(x) = (u_i(x, y))_{y \in Y}$ будет скалярной функцией $U_i(x, y)$ двух аргументов.

Определим понятие предпочтения. В этой работе будем придерживаться строгой и слабой аксиом Парето.

Интерпретация этих аксиом [2] в условиях неопределенности заключается в следующем. В соответствии с сильной аксиомой Парето говорят, что выигрыш игрока $i \in N$ в ситуации x' доминирует его выигрыш в ситуации x'' по множеству состояний природы $Y = \{y_1, y_2, \dots, y_m\}$ и обозначают это через $U_i(x') \succ U_i(x'')$, если $u_i(x', y) \geq u_i(x'', y)$ для $\forall y \in Y$ и хоть одно неравенство строгое, т.е. $\exists \bar{y} \in Y : u_i(x', \bar{y}) > u_i(x'', \bar{y})$. В соответствии со слабой аксиомой Парето говорят, что выигрыш игрока $i \in N$ в ситуации x' сильно доминирует его выигрыш в ситуации x'' по множеству состояний природы $Y = \{y_1, y_2, \dots, y_m\}$ и обозначают это через $U_i(x') \succ \succ U_i(x'')$, если $u_i(x', y) > u_i(x'', y)$ для $\forall y \in Y$.

Для формализации понятия равновесия игры в условиях неопределенности нам будут необходимы специальные отношения доминирования на множестве ситуаций игры, которые мы назовем отношением доминирования по Нешу и слабым отношением доминирования по Нешу. Будем говорить, что ситуация x игры UG доминирует по Нешу ситуацию x' ($x \succ^{NE} x'$), которая получается из ситуации x изменением, каким-то, но лишь одним игроком, своей стратегии, если: $\exists i \in N : U_i(x_i, x_{N \setminus i}) \succ U_i(x'_i, x_{N \setminus i})$. В случае слабого доминирования: $x \succ \succ^{NE} x' \Leftrightarrow \exists i \in N : U_i(x_i, x_{N \setminus i}) \succ \succ U_i(x'_i, x_{N \setminus i})$.

В соответствии с отношением доминирования (\succ^{NE}) на множестве ситуаций X определим множество сильных виртуальных равновесий (обозначим его через $SVE(UG)$). Оно будет состоять из ситуаций x^* , которые не доминируются по Нешу ни одной другой ситуацией, т.е. $SVE(UG) = \left\{ x^* \mid \neg \exists x \in X : x \succ^{NE} x^* \right\}$.

Аналогично определяется множество слабых виртуальных равновесий $WVE(UG) = \left\{ x^* \mid \neg \exists x \in X : x \succ \succ^{NE} x^* \right\}$.

Рассмотрим подробнее виртуальные равновесия. Обозначим через $SBR_i = \left\{ (x_i, x_{N \setminus i}) \mid \neg \exists \bar{x}_i \in X_i, U_i(\bar{x}_i, x_{N \setminus i}) \succ U_i(x_i, x_{N \setminus i}), x_{N \setminus i} \in X_{N \setminus i} \right\}$ множество ситуаций игры UG , состоящее из "недоминируемых по множеству состояний природы" ответов $x_i \in X_i$ игрока $i \in N$ на фиксированные наборы $x_{N \setminus i} = (x_j)_{j \in N \setminus \{i\}} \in X_{N \setminus i} = \prod_{j \in N \setminus \{i\}} X_j$ стратегий остальных игроков. Аналогично

$WBR_i = \left\{ (x_i, x_{N \setminus i}) \mid \neg \exists \bar{x}_i \in X_i : U_i(\bar{x}_i, x_{N \setminus i}) \succ \succ U_i(x_i, x_{N \setminus i}), x_{N \setminus i} \in X_{N \setminus i} \right\}$ - множество ситуаций, состоящее из "сильно недоминируемых по множеству состояний природы" ответов игрока $i \in N$ на фиксированные наборы стратегий остальных игроков. Тогда $SVE(UG) = \bigcap_{i \in N} SBR_i$ и $WVE(UG) = \bigcap_{i \in N} WBR_i$.

Рассмотрим пример виртуального равновесия для следующей игры двух лиц с двумя состояниями природы Y_1, Y_2 :

	A21	A22	A23
A11	(3,3) (1,1)	(4,1) (2,2)	(1,1) (3,3)
A12	(2,2) (1,3)	(1,3) (2,2)	(2,2) (3,1)
A13	(1,1) (3,3)	(2,2) (2,2)	(3,3) (1,1)

В этом примере: {A11, A12, A13} – множество стратегий первого игрока, {A21, A22, A23} - множество стратегий второго игрока; выигрыши игроков представлены в виде векторов, первая компонента вектора – выигрыш при состоянии природы y_1 , вторая компонента - выигрыш при состоянии природы y_2 ; векторы выигрышей первого игрока представленные в верхних левых углах ячеек таблицы, а второго – в нижних правых углах. Интересная особенность примера - отсутствие равновесий Неша в этой игре при любом фиксированном состоянии природы. По определению виртуального равновесия получим:

$$SVE = SBR_1 \cap SBR_2 = \{(A11,A21), (A11,A22), (A12,A22), (A13,A22), (A13,A23)\} \cap \{(A11,A23), (A12,A21), (A12,A22), (A12,A23), (A13,A21)\} = \{(A12,A22)\}.$$

Условия виртуального равновесия

Рассмотрим условия виртуального равновесия, что не основываются на специальных свойствах множеств стратегий и функций выигрыша игроков. Сначала рассмотрим условие слабого виртуального равновесия.

Обозначим через $S = \sum_{i \in N} \sup_{x \in X, y \in Y} u_i(x, y)$ верхнюю границу суммы функций выигрыша всех игроков и будем не ограничивая общности считать функции выигрыша игроков $u_i(x, y) > 0, \forall x \in X, \forall y \in Y, i \in N$.

Теорема 1. Ситуация x^* будет слабым виртуальным равновесием тогда и только тогда, когда существуют векторы параметров $\mu_i \in M_i = \left\{ \mu_i = (\mu_i(y))_{y \in Y} \mid \sum_{y \in Y} \mu_i(y) = S; \mu_i(y) \geq 0, y \in Y \right\}, i \in N$ такие, что ситуация x^* будет равновесием Неша в параметрической игре $G(\mu) = (X_i, \min_{y \in Y} (u_j(x, y) - \mu_i(y)); i \in N)$.

Доказательство. Докажем достаточность. Пусть x^* - равновесие Неша в параметрической игре $G(\bar{\mu})$ при некоторых значениях векторов параметров $\bar{\mu}_i \in M_i, i \in N$. Отсюда следует, что для $\forall i \in N$: $\min_{y \in Y} (u_j(x_i, x_{N \setminus i}^*, y) - \bar{\mu}_i(y)) \leq \min_{y \in Y} (u_j(x^*, y) - \bar{\mu}_i(y)) \leq u_i(x^*, y) - \bar{\mu}_i(y), \forall y \in Y, \forall x_i \in X_i$ поэтому $\forall i \in N, \forall x_i \in X_i$ существует $\bar{y} \in Y$, что $u_i(x_i, x_{N \setminus i}^*, \bar{y}) \leq u_i(x^*, \bar{y})$. Следовательно $\neg \exists x = (x_i, x_{N \setminus i}^*) \in X : x \succ_{NE} x^*$. Отсюда $x^* \in WVE(UG)$.

Докажем необходимость. Для этого возьмем векторы $\bar{\mu}_i = (\bar{\mu}_i(y))_{y \in Y}, i \in N$ с компонентами $\bar{\mu}_i(y) = u_j(x^*, y) * S / \sum_{y \in Y} u_j(x^*, y), i \in N$. Несложно убедиться, что $\bar{\mu}_i \in M_i, i \in N$. Из того, что x^* - слабое виртуальное равновесие следует, что $\neg \exists x \in X : x \succ_{NE} x^*$, т.е. $\forall i \in N, \forall x_i \in X_i, \exists \bar{y} \in Y : u_i(x_i, x_{N \setminus i}^*, \bar{y}) \leq u_i(x^*, \bar{y})$, а значит $u_i(x_i, x_{N \setminus i}^*, \bar{y}) - \bar{\mu}_i(\bar{y}) \leq u_i(x^*, \bar{y}) - \bar{\mu}_i(\bar{y})$. Поскольку для $\forall i \in N$ значение

$u_i(x^*, \bar{y}) - \bar{\mu}_i(\bar{y}) = 1 - S / \sum_{y \in Y} u_i(x^*, y), \forall \bar{y} \in Y$, то для $\forall x_i \in X_i$ справедливо

$\min_{y \in Y} (u_j(x_i, x_{N \setminus i}^*, y) - \bar{\mu}_i(y)) \leq 1 - S / \sum_{y \in N} u_i(x^*, y) = \min_{y \in Y} (u_i(x^*, y) - \bar{\mu}_i(y))$. Поэтому $x^* \in NE(G(\bar{\mu}))$.

Теорема доказана.

Рассмотрим условия сильного векторного равновесия.

Теорема 2. Ситуация $x^* \in X$ будет сильным виртуальным равновесием тогда и только тогда, когда

$$\sum_{y \in Y} u_j(x^*, y) = \max \left\{ \sum_{y \in Y} u_i(x_i, x_{N \setminus i}^*, y) \mid u_i(x_i, x_{N \setminus i}^*, y) \geq u_j(x^*, y), y \in Y; x_i \in X_i \right\}, \forall i \in N.$$

Доказательство. Пусть $x^* \in SVE(UG)$. Тогда $\neg \exists x \in X : x \succ^{NE} x^*$. Построим усечения множеств стратегий игроков: $X_i^* = \{x_i \in X_i \mid u_i(x_i, x_{N \setminus i}^*, y) \geq u_j(x^*, y), y \in Y\}$, $i \in N$. Заметим, что

$X_i^* \neq \emptyset, \forall i \in N$. Предположим противное, что x^* не удовлетворяет условиям теоремы. Тогда $\exists i \in N, \exists \bar{x}_i \in X_i^* : \sum_{y \in Y} u_i(\bar{x}_i, x_{N \setminus i}^*, y) > \sum_{y \in Y} u_j(x^*, y)$. Отсюда следует $u_i(\bar{x}_i, x_{N \setminus i}^*, y) \geq u_j(x^*, y), \forall y \in Y$ и

$\exists \bar{y} \in Y : u_i(\bar{x}_i, x_{N \setminus i}^*, \bar{y}) > u_j(x^*, \bar{y})$, следовательно $\bar{x} \succ^{NE} x^*$. Поэтому $x^* \notin SVE(UG)$. Получили противоречие.

Пусть ситуация x^* удовлетворяет условиям теоремы. Предположим противное, что $x^* \notin SVE(UG)$. Тогда

$\exists i \in N$, для которого $\exists \bar{x} = (\bar{x}_i, x_{N \setminus i}^*) \in X$, что $\bar{x} \succ^{NE} x^*$, т.е. $u_i(\bar{x}_i, x_{N \setminus i}^*, y) \geq u_j(x^*, y), \forall y \in Y$ и $\exists \bar{y} \in Y : u_i(\bar{x}_i, x_{N \setminus i}^*, \bar{y}) > u_j(x^*, \bar{y})$. Суммируя эти неравенства по $y \in Y$, получим: $\sum_{y \in Y} u_j(\bar{x}_i, x_{N \setminus i}^*, y) > \sum_{y \in Y} u_j(x^*, y)$. А поскольку $\bar{x}_i \in X_i^*$, $i \in N$, то приходим к противоречию. Теорема доказана.

Виртуальные равновесия в смешанных стратегиях

Как равновесия Неша в классических играх, так и виртуальные равновесия в играх в условиях неопределенности, к сожалению, не всегда существуют даже в случае конечных множеств стратегий игроков. Одним из известных [1] радикальных приемов найти им “замену” в классических играх является изменение условий игры. Игрокам предлагается принимать решение опосредовано, используя специально построенный тотализатор, который реализует случайный выбор стратегий по определенным распределениям вероятностей на множестве стратегий. Эти распределения вероятностей называют смешанными стратегиями.

Рассмотрим смешанное расширение игры в условиях неопределенности с целью построения виртуальных равновесий. Для простоты будем считать множества стратегий игроков конечными.

Будем говорить, что игра $MUG = \langle Y; M_i, \bar{\mu}_i; i \in N \rangle$ является смешанным расширением игры UG , если: множество смешанных стратегий i -го игрока

$$M_i = \{ \mu_i = \mu_i(x_i)_{x_i \in X_i} \mid \mu_i(x_i) \geq 0, x_i \in X_i; \sum_{x_i \in X_i} \mu_i(x_i) = 1 \};$$

- смешанные стратегии: $\mu_i = \mu_i(x_i)_{x_i \in X_i}$ (дискретные распределения вероятностей на множестве стратегий i -го игрока, которые мы будем представлять векторами, состоящими из $|X_i|$ компонент);

- смешанные ситуации игры: $\mu = (\mu_i)_{i \in N}$ - элементы множества смешанных ситуаций игры $M = \prod_{i \in N} M_i$;
- вектор-функция математического ожидания выигрыша i -го игрока - $\bar{U}_i(\mu) = (\bar{u}_i(\mu, y))_{y \in Y}$, где $\bar{u}_i(\mu, y) = \sum_{x \in X} u_i^j(x, y) \prod_{k \in N} \mu_k(x_k)$ - математическое ожидание выигрыша i -го игрока при состоянии природы $y \in Y$.

Для упрощения изложения иногда будем выделять в $\bar{u}_i(\mu, y)$ смешанную стратегию i -го игрока. Для этого сделаем следующие обозначения:

$$\begin{aligned} \bar{u}_i(\mu, y) &= \sum_{x \in X} u_i(x, y) \prod_{k \in N} \mu_k(x_k) = \sum_{x_i \in X_i} \mu_i(x_i) \left(\sum_{x_{N \setminus i} \in X_{N \setminus i}} u_i(x_i, x_{N \setminus i}, y) \prod_{k \in N \setminus i} \mu_k(x_k) \right) = \\ &= \sum_{x_i \in X_i} v_i(x_i, \mu_{N \setminus i}, y) \mu_i(x_i) = \langle b_i(\mu_{N \setminus i}, y), \mu_i \rangle, y \in Y, i \in N. \end{aligned} \quad (1)$$

Отметим, что множество стратегий M_i игрока $i \in N$ выпукло и компактно (по определению - это $(|X_i| - 1)$ -мерный единичный симплекс), а функции выигрыша - линейны по каждой переменной μ_i , в отдельности.

Найти виртуальные равновесия в смешанных стратегиях по определению достаточно сложно. Возникает вопрос, а можно ли свести эту задачу к известной? Этот вопрос положительно решается для нахождения как слабых, так и сильных виртуальных равновесий.

Пусть $\langle \alpha_i, \bar{U}_i(\mu) \rangle$ - линейная свертка вектор-функции математического ожидания выигрыша i -го игрока с вектором коэффициентов $\alpha_i = (\alpha_i(y))_{y \in Y}$, которые мы будем считать неизвестными параметрами.

Рассмотрим однокритериальную параметрическую игру $MUG(\alpha) = (Y; M_i, \langle \alpha_i, \bar{U}_i(\mu) \rangle; i \in N)$.

Обозначим множество значений параметров α через

$$A_{\geq 0} = \{ \alpha = (\alpha_i)_{i \in N} = (\alpha_i(y))_{i \in N, y \in Y} : \alpha_i(y) \geq 0, y \in Y; \sum_{y \in Y} \alpha_i(y) = 1, i \in N \}.$$

Теорема 3. Множество слабых виртуальных равновесий смешанного расширения игры G совпадает с семейством по параметру $\alpha \in A_{\geq 0}$ равновесий Неша смешанного расширения параметрической игры

$$MUG(\alpha), \text{ то есть } WVE(MUG) = \bigcup_{\alpha \in A_{\geq 0}} NE(MUG(\alpha)).$$

Доказательство. Покажем сначала включение $\bigcup_{\alpha \in A_{\geq 0}} NE(MUG(\alpha)) \subseteq WVE(MUG)$. Зафиксируем

некоторое значение $\alpha \in A_{\geq 0}$. Пусть $\mu^* \in NE(MUG(\alpha))$. По определению равновесия Неша, с учетом обозначений (1), имеем:

$$\forall \mu_i \in M_i, \forall i \in N: \sum_{y \in Y} \alpha_i(y) \langle b_i(\mu_{N \setminus i}^*, y), \mu_i^* \rangle \geq \sum_{y \in Y} \alpha_i(y) \langle b_i(\mu_{N \setminus i}^*, y), \mu_i \rangle. \quad (2)$$

Предположим противное. Пусть, $\mu^* \notin WVE(MUG)$ тогда $\exists \mu \in M: \mu \succ^{NE} \mu^*$, т.е. $\exists i \in N: \langle b_i(\mu_{N \setminus i}^*, y), \mu_i \rangle > \langle b_i(\mu_{N \setminus i}^*, y), \mu_i^* \rangle, \forall y \in Y$. Просуммируем эти неравенства с коэффициентами

$\alpha_i(y) \geq 0: \sum_{y \in Y} \alpha_i(y) = 1$, получим: $\exists i \in N: \sum_{y \in Y} \alpha_i(y) \langle b_i(\mu_{N^i}^*, y), \mu_i \rangle > \sum_{y \in Y} \alpha_i(y) \langle b_i(\mu_{N^i}^*, y), \mu_i^* \rangle$, что противоречит условию (2). Следовательно $\mu^* \in WVE(MUG)$.

Докажем включение $WVE(MUG) \subseteq \bigcup_{\alpha \in A_{\geq 0}} NE(MUG(\alpha))$. Пусть $\mu^*(x) \in WVE(MUG)$, следовательно, по

определению слабого виртуального равновесия $\neg \exists \mu \in M: \mu \succ^{NE} \mu^*$, т.е.

$\forall i \in N, \neg \exists \mu_i \in M_i: \bar{U}_i(\mu_i, \mu_{N^i}^*) \succ^{NE} \bar{U}_i(\mu^*)$. Другими словами, с учетом обозначений (1), система неравенств: $\langle b_i(\mu_{N^i}^*, y), \mu_i \rangle > \langle b_i(\mu_{N^i}^*, y), \mu_i^* \rangle, y \in Y$, не будет иметь решений при условии $\mu_i \in M_i$.

Поскольку левые части неравенств линейны по μ_i , а поэтому является вогнутыми функциями, то по теореме Фана-Гликсберга-Гоффмана [4] существует такой вектор $\alpha_i^* = (\alpha_i^*(y))_{y \in Y} \geq 0$, что:

$$\sum_{y \in Y} \alpha_i^*(y) \langle b_i(\mu_{N^i}^*, y), \mu_i \rangle \leq \sum_{y \in Y} \alpha_i^*(y) \langle b_i(y)(\mu_{N^i}^*, y), \mu_i^* \rangle.$$

А это в свою очередь, с учетом обозначений (1) и по определению равновесия Неша означает, что $\mu^* \in NE(MUG(\alpha^*)) \subseteq \bigcup_{\alpha \in A_{\geq 0}} NE(MUG(\alpha))$. Теорема доказана.

Обозначим через $A_{>0} = \{\alpha = (\alpha_i)_{i \in N} = (\alpha_i(y))_{i \in N, y \in Y_i}: \alpha_i(y) > 0, y \in Y; \sum_{y \in Y} \alpha_i(y) = 1, i \in N\}$.

Теорема 4. Множество сильных виртуальных равновесий смешанного расширения игры UG совпадает с семейством по параметру $\alpha \in A_{>0}$ равновесий Неша смешанного расширения параметрической игры $MUG(\alpha)$, т.е. $SVE(MUG) = \bigcup_{\alpha \in A_{>0}} NE(MUG(\alpha))$.

Доказательство. Покажем сначала включение $\bigcup_{\alpha \in A_{>0}} NE(MUG(\alpha)) \subseteq SVE(MUG)$. Пусть

$\mu^* \in NE(MUG(\alpha))$, при некотором $\alpha \in A_{>0}$. По определению равновесия Неша, с учетом обозначений (1), имеем:

$$\forall \mu_i \in M_i, \forall i \in N, \sum_{y \in Y} \alpha_i(y) \langle b_i(\mu_{N^i}^*, y), \mu_i \rangle \geq \sum_{y \in Y} \alpha_i(y) \langle b_i(\mu_{N^i}^*, y), \mu_i^* \rangle. \quad (3)$$

Предположим противное. Пусть $\mu^* \notin SVE(MUG)$, тогда $\exists \mu \in M: \mu \succ^{NE} \mu^*$, т.е.

$$\exists i \in N: \begin{cases} \langle b_i(\mu_{N^i}^*, y), \mu_i \rangle \geq \langle b_i(\mu_{N^i}^*, y), \mu_i^* \rangle, \forall y \in Y, \\ \exists \bar{y} \in Y: \langle b_i(\mu_{N^i}^*, \bar{y}), \mu_i \rangle > \langle b_i(\mu_{N^i}^*, \bar{y}), \mu_i^* \rangle. \end{cases}$$

Просуммируем эти неравенства с коэффициентами $\alpha_i(y) > 0$, получим:

$\sum_{y \in Y} \alpha_i(y) \langle b_i(\mu_{N^i}^*, y), \mu_i \rangle > \sum_{y \in Y} \alpha_i(y) \langle b_i(\mu_{N^i}^*, y), \mu_i^* \rangle$, что противоречит условию (3). Следовательно

$\mu^* \in SVE(MUG)$.

Докажем включение $SVE(MUG) \subseteq \bigcup_{\alpha \in A_{>0}} NE(MUG(\alpha))$. Пусть $\mu^* \in SVE(MUG)$, поэтому по определению

сильного виртуального равновесия $\neg \exists \mu \in M : \mu \succ^{NE} \mu^*$, т.е. $\forall i \in N, \neg \exists \mu_i \in M_i : \bar{U}_i(\mu_i, \mu_{N \setminus i}^*) > \bar{U}_i(\mu^*)$.

Другими словами, с учетом обозначений (1), для $\forall i \in N$, система линейных неравенств:

$$\begin{cases} \sum_{x_i \in X_i} v_i(x_i, \mu_{N \setminus i}^*, y) \mu_i(x_i) \geq \sum_{x_i \in X_i} v_i(x_i, \mu_{N \setminus i}^*, y) \mu_i^*(x_i), \forall y \in Y, \\ \exists \bar{y} \in Y : \sum_{x_i \in X_i} v_i(x_i, \mu_{N \setminus i}^*, \bar{y}) \mu_i(x_i) > \sum_{x_i \in X_i} v_i(x_i, \mu_{N \setminus i}^*, \bar{y}) \mu_i^*(x_i), \end{cases}$$

при условии $\sum_{x_i \in X_i} \mu_i(x_i) = 1; \mu_i(x_i) \geq 0, x_i \in X_i$, будет несовместной.

Пусть $X_i^* = \{x_i \mid \mu_i^*(x_i) = 0\}$ - множество чистых стратегий i -го игрока, вероятность выбора которых по смешанной стратегии μ^*_i равняется нулю. Тогда при условии

$$\sum_{x_i \in X_i} \mu_i(x_i) = 0; \mu_i(x_i) \geq 0, x_i \in X_i^* \quad (4)$$

будет несовместной система линейных неравенств:

$$\begin{cases} \sum_{x_i \in X_i} v_i(x_i, \mu_{N \setminus i}^*, y) \mu_i(x_i) \geq 0, \forall y \in Y, \\ \exists \bar{y} : \sum_{x_i \in X_i} v_i(x_i, \mu_{N \setminus i}^*, \bar{y}) \mu_i(x_i) > 0, \end{cases} \quad (5)$$

Действительно, если допустить противное, что смешанная стратегия μ_i является ее решением, то легко удостовериться подстановкой, что смешанная стратегия $\tilde{\mu}_i = \mu_i^* + \varepsilon \mu_i, \varepsilon > 0$, будет образовывать ситуацию $(\tilde{\mu}_i, \mu_{N \setminus i}^*) \succ^{NE} \mu^*$, что противоречит условию $\mu^* \in SVE(MUG)$.

Из несовместности систем линейных неравенств (4),(5) по теореме Таккера об альтернативе для систем линейных неравенств [5] следует существование векторов $\alpha_i = (\alpha_i(y))_{y \in Y} > 0; \lambda_i = \lambda_i(x_i)_{x_i \in X_i} \geq 0$ и произвольного числа λ_i^0 таких, что выполняются равенства:

$$\sum_{y \in Y} \alpha_i(y) v_i(x_i, \mu_{N \setminus i}^*, y) = \lambda_i^0 - \lambda_i(x_i), x_i \in X_i^* \quad (6)$$

$$\sum_{y \in Y} \alpha_i(y) v_i(x_i, \mu_{N \setminus i}^*) = \lambda_i^0, x_i \in X_i \setminus X_i^* \quad (7)$$

$$\lambda_i^0 (1 - \sum_{x_i \in X_i} \mu_i^*(x_i)) + \sum_{x_i \in X_i} \lambda_i(x_i) \mu_i^*(x_i) = 0. \quad (8)$$

Пусть μ_i - произвольная смешанная стратегия i -го игрока. Запишем произведение каждого равенства (6),(7) с $\mu_i(x_i) - \mu_i^*(x_i)$, возьмем их сумму по $x_i \in X_i$ и с помощью (8) оценим правую часть полученного равенства:

$$\begin{aligned} \sum_{x_i \in X_i} (\mu_i(x_i) - \mu_i^*(x_i)) \left(\sum_{y \in Y} \alpha_i(y) v_i(x_i, \mu_{N \setminus i}^*, y) \right) &= \sum_{x_i \in X_i^*} (\mu_i(x_i) - \mu_i^*(x_i)) (\lambda_i^0 - \lambda_i(x_i)) + \\ \sum_{x_i \in X_i \setminus X_i^*} (\mu_i(x_i) - \mu_i^*(x_i)) \lambda_i^0 &= \sum_{x_i \in X_i} (\mu_i(x_i) - \mu_i^*(x_i)) \lambda_i^0 - \sum_{x_i \in X_i^*} (\mu_i(x_i) - \mu_i^*(x_i)) \lambda_i(x_i) = \end{aligned}$$

$$\begin{aligned}
&= \lambda_i^0 - \sum_{x_i \in X_i} \mu_i^*(x_i) \lambda_i^0 + \sum_{x_i \in X_i^*} \mu_i^*(x_i) \lambda_i(x_i) - \lambda_i^0 + \sum_{x_i \in X_i} \mu_i(x_i) \lambda_i^0 - \sum_{x_i \in X_i^*} \mu_i(x_i) \lambda_i(x_i) = \\
&= \sum_{x_i \in X_i} (1 - \mu_i^*(x_i)) \lambda_i^0 + \sum_{x_i \in X_i^*} \mu_i^*(x_i) \lambda_i(x_i) - \sum_{x_i \in X_i} (1 - \mu_i(x_i)) \lambda_i^0 - \sum_{x_i \in X_i^*} \mu_i(x_i) \lambda_i(x_i) = \\
&= - \sum_{x_i \in X_i} (1 - \mu_i(x_i)) \lambda_i^0 - \sum_{x_i \in X_i^*} \mu_i(x_i) \lambda_i(x_i) = - \sum_{x_i \in X_i^*} \mu_i(x_i) \lambda_i(x_i) \leq 0.
\end{aligned}$$

Последние соотношения следуют из условия: $\sum_{x_i \in X_i} \mu_i(x_i) = 1$; $\mu_i(x_i) \geq 0$, $x_i \in X_i$. Поэтому с учетом обозначений (1), имеем: $\forall i \in N$, $\sum_{y \in Y} \alpha_i(y) \langle b_i(y)(\mu_{N \setminus i}^*), \mu_i^* \rangle \geq \sum_{y \in Y} \alpha_i(y) \langle b_i(y)(\mu_{N \setminus i}^*), \mu_i \rangle$, $\forall \mu_i \in M_i$.

Отсюда по определению равновесия Неша [1], имеем: $\mu^* \in NE(MUG(\alpha)) \subseteq \bigcup_{\alpha \in A_{>0}} NE(MUG(\alpha))$.

Доказанные выше теоремы, дают возможность сформулировать следующее важное следствие.

Следствие (о существовании виртуальных равновесий в смешанных стратегиях). Если множества $X_i, i \in N$ "чистых стратегий" конечны, то в смешанном расширенными играх в условиях неопределенности множество слабых виртуальных равновесий включает в себя множество сильных виртуальных равновесий и они являются непустыми подмножествами множества смешанных ситуаций игры MUG , т.е. $WVE(MUG) \supseteq SVE(MUG) \neq \emptyset$.

Доказательство. Включение $WVE(MUG) \supseteq SVE(MUG)$ элементарно следует из определений множеств слабых и сильных виртуальных равновесий. Из этого включения следует, что для доказательства теоремы достаточно рассмотреть множество $SVE(MUG)$. Поскольку по теореме 4,

$$SVE(MUG) = \bigcup_{\alpha \in A_{>0}} NE(MUG(\alpha)),$$

то рассмотрим игру $GM(\alpha) = \left\langle M_i, \sum_{y \in Y} \alpha_i(y) \bar{u}_i(\mu, y); i \in N \right\rangle$

при любых: $\alpha_i(y) > 0, y \in Y, i \in N$. Поскольку множества стратегий игроков M_i выпуклы и компактны (по определению, они являются $(|X_i| - 1)$ -мерные единичные симплексы), а функции выигрыша игроков линейны по каждой отдельной вероятности $\mu_i(x_i)$, $x_i \in X_i$, и непрерывны на множестве смешанных ситуаций игры $M = \prod_{i \in N} M_i$ (линейная свертка линейных по каждой отдельной вероятности функций), то игра $MUG(\alpha)$ удовлетворяет условиям теоремы Неша [1], а, следовательно, для нее существует не пустое компактное множество равновесий Неша. Отсюда следует, что множество $SVE(MUG) \neq \emptyset$.

Заключение

В заключение следует отметить, что приведенные выше условия сильного виртуального равновесия существенно проигрывают в плане конструктивности условиям слабого векторного равновесия, поскольку при построении множества сильных виртуальных равновесий надо заранее знать значения выигрышей игроков. В случае же слабого векторного равновесия, проверка условий сводится к игровой задаче, что позволяет более гибко использовать теорему 2 для решения прикладных задач. С другой стороны, сильная аксиома Парето, на которой базируется понятие сильного виртуального равновесия, имеет более широкий спектр применения на практике, чем слабая аксиома Парето, что сужает круг задач, в которых, может быть использована концепция слабого виртуального равновесия.

Проведенное исследование смешанного расширения игры в условиях неопределенности дает конструктивный подход, как для нахождения, так и для более глубокого изучения свойств виртуальных равновесий.

Ссылки

[1] Мулен Э. Теория игр с примерами из математической экономики. –М.: Мир, 1985. [2] Подиновский В.В., Ногин В.Д. Парето-оптимальные решения многокритериальных задач. –М.: Наука, 1982. [3] Mangasarian O.J. Nonlinear programming. New York-London: McGraw-Hill, 1969. [4] Рокафелар Р. Выпуклый анализ. М: Мир, 1973.

Информация об авторе

Мащенко Сергей Олегович – Киевский национальный университет имени Тараса Шевченко, Доцент; Проспект академика Глушкова, 6, Киев – 207, Украина; e-mail: msomail@yandex.ru

ГИПЕРСЛУЧАЙНЫЕ ЯВЛЕНИЯ: ОПРЕДЕЛЕНИЯ И ОПИСАНИЕ

Игорь Горбань

Аннотация: Дан обзор новой ветви теории вероятностей и математической статистики – теории гиперслучайных явлений. Изложены математические основы теории и методы моделирования физических объектов с помощью гиперслучайных моделей, учитывающих непредсказуемый характер изменения свойств объектов и статистических условий их наблюдения.

Ключевые слова: неопределенность, случайное явление, гиперслучайное явление, состоятельность, точность.

Введение

Одной из основных проблем познания является адекватное описание реального мира. Физические и математические модели, используемые для описания различных явлений, постоянно совершенствуются. С получением новых данных и развитием представлений о мире старые модели отходят на второй план, заменяются новыми, более совершенными.

Если до конца средневековья мир виделся неизменным и описывался преимущественно детерминированными моделями, то в настоящее время он рассматривается как динамично меняющаяся структура. Главным средством математического описания мира стали модели, использующие в качестве абстрактных математических объектов случайные явления – случайные события, величины, процессы и поля.

Исчерпывающей характеристикой любого случайного явления является функция распределения вероятности, определяемая для строго постоянных условий наблюдения. Обеспечить на практике полную стабильность условий нельзя, а, следовательно, нельзя задать или оценить абсолютно точно функцию распределения. Это накладывает ограничения на применение классических случайных моделей с фиксированным законом распределения. Корректное их использование оказывается возможным лишь при пренебрежимо малых изменениях условий наблюдения.

Если условия меняются существенно и непредсказуемо, в модель обычно вводят дополнительные элементы неопределенности (см., например, [1 – 8]). Такой паллиатив, не смотря на некоторую искусственность, дает неплохие результаты, в особенности при решении не очень сложных задач.

Стремление найти простые средства учета неопределенности условий наблюдения играла, по всей видимости, не последнюю роль при формировании ряда относительно новых научных направлений, таких как теория нечетких множеств [9], теория нейронных сетей [10], теория хаотических динамических систем [11, 12], теория интервальных данных [13, 14] и др.

Постоянно продолжающийся поиск универсальных и эффективных путей адекватного математического описания явлений недавно привел к новому классу моделей, в которых в качестве абстрактных математических объектов выступают, так называемые, гиперслучайные явления [15].

Под гиперслучайными явлениями подразумеваются семейство случайных событий, величин, функций или полей, зависящие от параметра $g \in G$, который рассматривается как независимая переменная и ассоциируется с условиями наблюдения (или условиями формирования) рассматриваемых объектов.

Математически случайные события описываются с помощью вероятностного пространства [16], задаваемого триадой $(\Omega, \mathfrak{F}, P)$, где Ω – пространство элементарных событий $\omega \in \Omega$, \mathfrak{F} – борелевское поле (σ – алгебра подмножеств событий) и P – вероятностная мера подмножеств событий.

При менее строгом, но более наглядном статистическом определении (по Р. фон Мизесу [17]), вероятность $P(A)$ случайного события A представляется как предел частоты $p_N(A)$ его появления при проведении опытов в одинаковых условиях и устремлении количества опытов N к бесконечности: $P(A) = \lim_{N \rightarrow \infty} p_N(A)$. При небольших значениях N частота $p_N(A)$ может колебаться, однако по мере увеличения N постепенно стабилизируется и при $N \rightarrow \infty$ стремится к определенному пределу $P(A)$.

Гиперслучайные явления можно описать с помощью тетрады $(\Omega, \mathfrak{F}, G, P_g)$ [15], где Ω – пространство элементарных событий $\omega \in \Omega$, \mathfrak{F} – борелевское поле, G – множество условий $g \in G$, P_g – вероятностная мера подмножеств событий, зависящая от условия g . Таким образом, вероятностная мера задается для всех подмножеств событий и всех возможных условий $g \in G$. Мера же для условий $g \in G$ остается не определенной.

Используя менее строгий статистический подход, гиперслучайное событие A можно трактовать как событие, частота появления которого $p_N(A)$ при увеличении числа опытов N не стабилизируется и при $N \rightarrow \infty$ не имеет предела.

Целью статьи является обзор концептуальных подходов, методов и результатов теории гиперслучайных явлений, полученных в работах [15, 18 – 23] и обобщенных в монографии [24].

Гиперслучайные события и величины

Гиперслучайное событие A борелевского поля \mathfrak{F} нельзя описать какой-то вероятностью, однако событие A/g в условиях $g \in G$ может быть представлено вероятностью $P_g(A) = P(A/g)$. Эта вероятность меняется при изменении условий. Диапазон значений вероятности описывается с помощью верхней $P_s(A)$ и нижней $P_l(A)$ границ:

$$P_s(A) = \sup_{g \in G} P(A/g), \quad P_l(A) = \inf_{g \in G} P(A/g).$$

В фиксированных условиях ($g = \text{const}$) эти границы совпадают и гиперслучайная величина вырождается в случайную, характеризуемую вероятностью $P(A) = P_s(A) = P_l(A)$.

Границы $P_S(A)$, $P_I(A)$ являются полумерами. Для них получены выражения, подобные тем, которые описывают правила сложения, умножения, а также теорему Байеса и другие теоремы теории вероятностей.

Для описания скалярной гиперслучайной величины X предложен ряд характеристик, аналогичных вероятностным характеристикам случайной величины. Основные из них – верхняя $F_S(x)$ и нижняя $F_I(x)$ границы функции распределения, а также плотности вероятности верхней $f_S(x)$ и нижней $f_I(x)$ границ, определяемые следующим образом:

$$F_S(x) = \sup_{g \in G} P\{X \leq x / g\}, \quad F_I(x) = \inf_{g \in G} P\{X \leq x / g\},$$

$$f_S(x) = \frac{dF_S(x)}{dx}, \quad f_I(x) = \frac{dF_I(x)}{dx},$$

где $P\{X \leq x / g\}$ – вероятность выполнения неравенства $X \leq x$ в условиях g .

Было установлено, что границы функции распределения и плотности вероятности границ гиперслучайной величины обладают теми же свойствами, что и соответствующие характеристики случайной величины. Кроме того, всегда $F_S(x) \geq F_I(x)$.

Среди границ функции распределения расположена зона неопределенности шириной $\Delta F(x) = F_S(x) - F_I(x)$ (рис. 1). Для случайной величины X величина $\Delta F(x)$ равна нулю для всех x . Когда для всех x верхняя граница $F_S(x)$ стремится к единице, а нижняя граница $F_I(x)$ – к нулю, зона неопределенности стремится к бесконечности. В этом случае гиперслучайная величина вырождается в хаотическую.

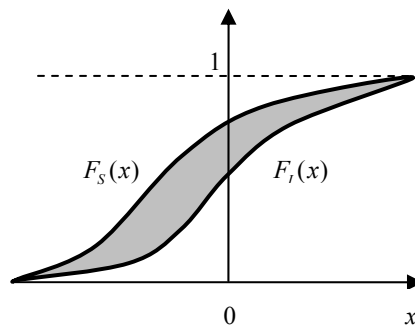


Рисунок 1. — Границы функции распределения и зона неопределенности (затемненная область).

Для описания гиперслучайных величин можно использовать характеристики, аналогичные тем, которые используются при описании случайных величин. К ним относятся центральные и нецентральные моменты границ, определяемые для гиперслучайной величины X на основе математического ожидания $M_S[\varphi(X)]$, $M_I[\varphi(X)]$ соответствующих границ функции $\varphi(X)$:

$$M_S[\varphi(X)] = \int_{-\infty}^{\infty} \varphi(x) f_S(x) dx, \quad M_I[\varphi(X)] = \int_{-\infty}^{\infty} \varphi(x) f_I(x) dx.$$

Математические ожидания границ $m_{Sx} = M_S[X]$, $m_{Ix} = M_I[X]$, дисперсии границ $D_{Sx} = M_S[(X - m_{Sx})^2]$, $D_{Ix} = M_I[(X - m_{Ix})^2]$, нецентральные моменты границ

$m_{Sx^v} = M_S [X^v]$, $m_{Ix^v} = M_I [X^v]$, центральные моменты границ $\mu_{Sx^v} = M_S [(X - m_{Sx})^v]$, $\mu_{Ix^v} = M_I [(X - m_{Ix})^v]$, где v – степень момента.

Для описания гиперслучайных величин можно использовать характеристики и другого типа: границы центральных и нецентральных моментов. Они определяются на основе границ математического ожидания функции $\varphi(X)$:

$$M_S[\varphi(X/g)] = \sup_{g \in G} \int_{-\infty}^{\infty} \varphi(x/g) f(x/g) dx, \quad M_I[\varphi(X/g)] = \inf_{g \in G} \int_{-\infty}^{\infty} \varphi(x/g) f(x/g) dx,$$

где $f(x/g)$ – плотность вероятности в условиях g .

Для гиперслучайной величины X границами математического ожидания являются величины $m_{Sx} = M_S [X]$, $m_{Ix} = M_I [X]$, границами дисперсии – $D_{Sx} = M_S [(X - m_{x/g})^2]$, $D_{Ix} = M_I [(X - m_{x/g})^2]$, где $m_{x/g}$ – математическое ожидание случайной величины X/g . Границы нецентральных моментов описываются выражениями $m_{Sx^v} = M_S [X^v]$, $m_{Ix^v} = M_I [X^v]$, а границы центральных моментов – выражениями $\mu_{Sx^v} = M_S [(X - m_{x/g})^v]$, $\mu_{Ix^v} = M_I [(X - m_{x/g})^v]$.

В общем случае, операторы $M_S[\cdot]$, $M_I[\cdot]$ отличаются от операторов $M_S[\cdot]$, $M_I[\cdot]$. Отличаются и моменты границ от границ моментов, хотя в некоторых отдельных случаях они выражаются друг через друга, например, когда функции распределения $F(x/g)$ для разных g не имеют точек пересечения [19]. Если при отсутствии такого пересечения условная дисперсия $D_{x/g}$ возрастает с ростом условного математического ожидания $m_{x/g}$ (тип распределения «а»), то имеют место следующие равенства: $m_{Sx} = m_{Ix}$, $m_{Ix} = m_{Sx}$, $D_{Sx} = D_{Ix}$, $D_{Ix} = D_{Sx}$; если же условная дисперсия $D_{x/g}$ уменьшается с ростом условного математического ожидания $m_{x/g}$ (тип распределения «б»), то – равенства $m_{Sx} = m_{Ix}$, $m_{Ix} = m_{Sx}$, $D_{Sx} = D_{Sx}$, $D_{Ix} = D_{Ix}$.

Результаты обобщены на комплексные \dot{X} и векторные \vec{X} гиперслучайные величины [15, 18], а также на вещественные $X(t)$, комплексные $\dot{X}(t)$ и векторные $\vec{X}(t)$ гиперслучайные функции [18, 19].

Гиперслучайные функции

Скалярный гиперслучайный процесс $X(t)$ может быть представлен как множество случайных процессов $X(t)/g$, которые определены для множества условий $g \in G$. Процесс описывается верхней $F_S(\vec{x}; \vec{t})$ и нижней $F_I(\vec{x}; \vec{t})$ границами многомерных функций распределения, многомерными плотностями вероятностей границ $f_S(\vec{x}; \vec{t})$ и $f_I(\vec{x}; \vec{t})$, моментами границ $m_{Sx^{\vec{v}}}(\vec{t})$, $m_{Ix^{\vec{v}}}(\vec{t})$, $\mu_{Sx^{\vec{v}}}(\vec{t})$, $\mu_{Ix^{\vec{v}}}(\vec{t})$ и границами моментов $m_{Sx^{\vec{v}}}(\vec{t})$, $m_{Ix^{\vec{v}}}(\vec{t})$, $\mu_{Sx^{\vec{v}}}(\vec{t})$, $\mu_{Ix^{\vec{v}}}(\vec{t})$, где $\vec{v} = (v_1, \dots, v_L)$ – степень вектора моментов, L – мерность распределения. Выражения, которые описывают такие характеристики, подобны выражениям, описывающим гиперслучайные величины:

$$F_S(\vec{x}; \vec{t}) = \sup_{g \in G} P\{X(t_1) \leq x_1, \dots, X(t_L) \leq x_L / g\}, \quad F_I(\vec{x}; \vec{t}) = \inf_{g \in G} P\{X(t_1) \leq x_1, \dots, X(t_L) \leq x_L / g\},$$

$$f_S(\vec{x}; \vec{t}) = \frac{\partial^L F_S(\vec{x}; \vec{t})}{\partial x_1 \dots \partial x_L}, \quad f_I(\vec{x}; \vec{t}) = \frac{\partial^L F_I(\vec{x}; \vec{t})}{\partial x_1 \dots \partial x_L},$$

$$m_{Sx}(t) = M_S[X(t)] = \int_{-\infty}^{\infty} xf_S(x;t)dx, \quad m_{Ix}(t) = M_I[X(t)] = \int_{-\infty}^{\infty} xf_I(x;t)dx,$$

$$m_{Sxv_1 \dots v_L}(t_1, \dots, t_L) = M_S[X^{v_1}(t_1) \dots X^{v_L}(t_L)] = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_1^{v_1} \dots x_L^{v_L} f_S(x_1, \dots, x_L; t_1, \dots, t_L) dx_1 \dots dx_L,$$

$$m_{Ixv_1 \dots v_L}(t_1, \dots, t_L) = M_I[X^{v_1}(t_1) \dots X^{v_L}(t_L)] = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_1^{v_1} \dots x_L^{v_L} f_I(x_1, \dots, x_L; t_1, \dots, t_L) dx_1 \dots dx_L,$$

$$\mu_{Sxv_1 \dots v_L}(t_1, \dots, t_L) = M_S[(X(t_1) - m_{Sx}(t_1))^{v_1} \dots (X(t_L) - m_{Sx}(t_L))^{v_L}],$$

$$\mu_{Ixv_1 \dots v_L}(t_1, \dots, t_L) = M_I[(X(t_1) - m_{Ix}(t_1))^{v_1} \dots (X(t_L) - m_{Ix}(t_L))^{v_L}].$$

$$m_{Sx}(t) = M_S[X(t)], \quad m_{Ix}(t) = M_I[X(t)],$$

$$m_{Sxv_1 \dots v_L}(t_1, \dots, t_L) = M_S[X^{v_1}(t_1) \dots X^{v_L}(t_L)], \quad m_{Ixv_1 \dots v_L}(t_1, \dots, t_L) = M_I[X^{v_1}(t_1) \dots X^{v_L}(t_L)],$$

$$\mu_{Sxv_1 \dots v_L}(t_1, \dots, t_L) = M_S[(X(t_1) - m_{x/g}(t_1))^{v_1} \dots (X(t_L) - m_{x/g}(t_L))^{v_L}],$$

$$\mu_{Ixv_1 \dots v_L}(t_1, \dots, t_L) = M_I[(X(t_1) - m_{x/g}(t_1))^{v_1} \dots (X(t_L) - m_{x/g}(t_L))^{v_L}].$$

Корреляционная и ковариационная функции границ имеют следующий вид:

$$K_{Sx}(t_1, t_2) = M_S[X(t_1)X(t_2)], \quad K_{Ix}(t_1, t_2) = M_I[X(t_1)X(t_2)],$$

$$R_{Sx}(t_1, t_2) = M_S[(X(t_1) - m_{Sx}(t_1))(X(t_2) - m_{Sx}(t_2))],$$

$$R_{Ix}(t_1, t_2) = M_I[(X(t_1) - m_{Ix}(t_1))(X(t_2) - m_{Ix}(t_2))],$$

а границы корреляционной и ковариационной функций – вид:

$$K_{Sx}(t_1, t_2) = M_S[X(t_1)X(t_2)], \quad K_{Ix}(t_1, t_2) = M_I[X(t_1)X(t_2)],$$

$$R_{Sx}(t_1, t_2) = M_S[(X(t_1) - m_{x/g}(t_1))(X(t_2) - m_{x/g}(t_2))],$$

$$R_{Ix}(t_1, t_2) = M_I[(X(t_1) - m_{x/g}(t_1))(X(t_2) - m_{x/g}(t_2))].$$

Стационарные и эргодические гиперслучайные функции

Некоторые гиперслучайные функции обладают свойствами стационарности и эргодичности [20]. Функция $X(t)$ названа стационарной гиперслучайной функцией, у которой математическое ожидание границ не зависит от времени, а корреляционные функции границ зависят лишь от временного интервала $\tau = t_2 - t_1$: $K_{Sx}(t_1, t_2) = K_{Sx}(\tau)$, $K_{Ix}(t_1, t_2) = K_{Ix}(\tau)$.

Стационарной гиперслучайной функцией для всех условий названа функция $X(t)$, математическое ожидание $m_{x/g}(t) = \int_{-\infty}^{\infty} xf(x;t/g)dx$ которой не зависит от времени t ($m_{x/g}(t) = m_{x/g}$), а корреляционная функция

$$K_{x/g}(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f(x_1, x_2; t_1, t_2 / g) dx_1 dx_2$$

зависит лишь от интервала τ и условий g : $K_{x/g}(t_1, t_2) = K_{x/g}(\tau)$.

Корреляционные функции границ $K_{Sx}(\tau)$, $K_{Ix}(\tau)$ определяются спектральными плотностями мощности границ $S_{Sxx}(f)$, $S_{Ixx}(f)$, связанные между собой следующими выражениями:

$$S_{Sxx}(f) = \int_{-\infty}^{\infty} K_{Sx}(\tau) \exp(-j2\pi f\tau) d\tau, \quad S_{Ixx}(f) = \int_{-\infty}^{\infty} K_{Ix}(\tau) \exp(-j2\pi f\tau) d\tau,$$

$$K_{Sx}(\tau) = \int_{-\infty}^{\infty} S_{Sxx}(f) \exp(j2\pi f\tau) df, \quad K_{Ix}(\tau) = \int_{-\infty}^{\infty} S_{Ixx}(f) \exp(j2\pi f\tau) df,$$

где f – частота.

Границы спектральной плотности мощности определяются следующим образом:

$$S_{Sxx}(f) = \sup_{g \in G} S_{Sxx/g}(f), \quad S_{Ixx}(f) = \inf_{g \in G} S_{Ixx/g}(f),$$

где $S_{Sxx/g}(f)$ – спектральная плотность мощности в условиях g :

$$S_{Sxx/g}(f) = \int_{-\infty}^{\infty} K_{Sx/g}(\tau) \exp(-j2\pi f\tau) d\tau, \quad K_{Sx/g}(\tau) = \int_{-\infty}^{\infty} S_{Sxx/g}(f) \exp(j2\pi f\tau) df.$$

Для двух стационарно связанных между собой функций $X(t)$, $Y(t)$ корреляционные функции границ определяются таким образом:

$$K_{Sxy}(\tau) = \int_{-\infty}^{\infty} \dot{S}_{Sxy}(f) \exp(j2\pi f\tau) df, \quad K_{Ixy}(\tau) = \int_{-\infty}^{\infty} \dot{S}_{Ixy}(f) \exp(j2\pi f\tau) df,$$

где $\dot{S}_{Sxy}(f)$, $\dot{S}_{Ixy}(f)$ – спектральные мощности границ:

$$\dot{S}_{Sxy}(f) = \int_{-\infty}^{\infty} K_{Sxy}(\tau) \exp(-j2\pi f\tau) d\tau, \quad \dot{S}_{Ixy}(f) = \int_{-\infty}^{\infty} K_{Ixy}(\tau) \exp(-j2\pi f\tau) d\tau.$$

Границы спектральной плотности мощности определяются выражениями:

$$\dot{S}_{Sxy}(f) = \sup_{g \in G} \dot{S}_{Sxy/g}(f), \quad \dot{S}_{Ixy}(f) = \inf_{g \in G} \dot{S}_{Ixy/g}(f),$$

где $\dot{S}_{Sxy/g}(f)$ – спектральные мощности для условий g :

$$\dot{S}_{Sxy/g}(f) = \int_{-\infty}^{\infty} K_{Sxy/g}(\tau) \exp(-j2\pi f\tau) d\tau, \quad K_{Sxy/g}(\tau) = \int_{-\infty}^{\infty} \dot{S}_{Sxy/g}(f) \exp(j2\pi f\tau) df.$$

Были установлены свойства этих характеристик и введен ряд новых понятий, в частности гиперслучайный белый шум [20].

Гиперслучайная функция может быть представлена как множество непересекающихся случайных функций длительностью T , определенных на интервалах $T_g = [Tg, T(g+1))$, в пределах которых условия не меняются ($g = 0, \pm 1, \pm 2, \dots$).

Пусть $X_g(t)$ – часть функции $X(t)$, соответствующая интервалу T_g и приведенная к интервалу $[-T/2, T/2)$:

$$X_g(t - T(g + 0,5)) = \begin{cases} X(t), & \text{если } t \in T_g, \\ 0, & \text{если } t \notin T_g. \end{cases}$$

Функция $X_g(t)$ в фиксированных условиях $g = 0, \pm 1, \pm 2, \dots$ является случайной функцией, определенная на интервале $t \in [-T/2, T/2)$. Множество таких функций в неопределенных условиях является гиперслучайной функцией $Y(t) = \{X_g(t), g = 0, \pm 1, \dots\}$. Гиперслучайной функцией является также функция $\varphi(Y(t_1), \dots, Y(t_L))$, где $t_1, \dots, t_L \in [-T/2, T/2)$.

Гиперслучайная функция $X(t)$, стационарная для всех условий, для которой $\lim_{T \rightarrow \infty} \bar{m}_\varphi(T) = m_\varphi$, названа эргодической. В этом выражении $\bar{m}_\varphi(T)$ – среднее по t :

$$\bar{m}_\varphi(T) = \bar{M}_T[\varphi(Y(t_1), \dots, Y(t_L))] = \frac{1}{T} \int_{-T/2}^{T/2} \varphi(Y(t_1+t), \dots, Y(t_L+t)) dt,$$

а $m_\varphi = M[\varphi(Y(t_1), \dots, Y(t_L))]$ – математическое ожидание.

Гиперслучайная эргодическая функция $X(t)$ может быть представлена следующим рядом:

$$X(t) = \lim_{T \rightarrow \infty} \sum_g X_g(t - T(g + 0, 5)).$$

Когда $T \rightarrow \infty$, границы математического ожидания и границы корреляционной и ковариационной функций описываются следующим образом:

$$\begin{aligned} \bar{m}_{sx_T} &= \sup_{g \in G} \frac{1}{T} \int_{T_g} x_g(t) dt, & \bar{m}_{ix_T} &= \inf_{g \in G} \frac{1}{T} \int_{T_g} x_g(t) dt, \\ \bar{K}_{sx_T}(\tau) &= \sup_{g \in G} \frac{1}{T} \int_{T_g} x_g(t+\tau) x_g(t) dt, & \bar{K}_{ix_T}(\tau) &= \inf_{g \in G} \frac{1}{T} \int_{T_g} x_g(t+\tau) x_g(t) dt, \\ \bar{R}_{sx_T}(\tau) &= \sup_g \frac{1}{T} \int_{T_g} [x_g(t+\tau) - \bar{m}_{x_T/g}] [x_g(t) - \bar{m}_{x_T/g}] dt, \\ \bar{R}_{ix_T}(\tau) &= \inf_g \frac{1}{T} \int_{T_g} [x_g(t+\tau) - \bar{m}_{x_T/g}] [x_g(t) - \bar{m}_{x_T/g}] dt, \\ & \text{где } \bar{m}_{x_T/g} = \frac{1}{T} \int_{T_g} x_g(t) dt. \end{aligned}$$

Гиперслучайные модели

Разработанный математический аппарат позволяет эффективно моделировать разнообразные реальные физические объекты и оценки в условиях непредсказуемого изменения свойств и статистических условий наблюдения.

Предложены [22 – 24] различные модели измерения: детерминировано-гиперслучайные, случайно-гиперслучайные и гиперслучайно-гиперслучайные, в которых объекты измерения (физические величины и процессы) представляются соответственно детерминированными, случайными и гиперслучайными моделями, а их оценки – гиперслучайными моделями.

В рамках детерминировано-гиперслучайной модели измерения векторного параметра $\vec{\theta}$ точность точечной оценки $\vec{\Theta}^*$ в фиксированных условиях g может быть описана величиной среднего квадрата абсолютной погрешности $\Delta_g^2 = M[|\vec{\Theta}^* - \vec{\theta}|^2 / \vec{\theta}, g]$, где M – оператор математического ожидания. Для

характеристики точности оценки без привязки к конкретным условиям можно говорить об интервале, в котором находится величина Δ_g^2 . Границами этого интервала могут быть $\Delta_{\min}^2 = \min[\Delta_S^2, \Delta_I^2]$ и $\Delta_{\max}^2 = \max[\Delta_S^2, \Delta_I^2]$, где $\Delta_S^2 = M_S[|\vec{\Theta}^* - \vec{\theta}|^2 / \vec{\theta}]$, $\Delta_I^2 = M_I[|\vec{\Theta}^* - \vec{\theta}|^2 / \vec{\theta}]$ – средние относительно границ квадраты абсолютной погрешности Δ_S^2 , Δ_I^2 .

Точность точечной оценки можно характеризовать также границами среднего квадрата погрешности:

$$\Delta_S^2 = \sup_{g \in G} M[|\vec{\Theta}^* - \vec{\theta}|^2 / \vec{\theta}, g], \quad \Delta_I^2 = \inf_{g \in G} M[|\vec{\Theta}^* - \vec{\theta}|^2 / \vec{\theta}, g].$$

В скалярном случае границы среднего квадрата ошибки Δ_S^2 , Δ_I^2 и Δ_g^2 , Δ_i^2 можно представить соответственно как $\Delta_S^2 = \sigma_S^2 + \varepsilon_{S0}^2$, $\Delta_I^2 = \sigma_I^2 + \varepsilon_{I0}^2$ и $\Delta_g^2 = \sup_{g \in G} [\sigma_g^2 + \varepsilon_{0/g}^2]$, $\Delta_i^2 = \inf_{g \in G} [\sigma_g^2 + \varepsilon_{0/g}^2]$, где $\sigma_S^2 = M_S[(\Theta^* - m_S)^2 / \theta]$, $\sigma_I^2 = M_I[(\Theta^* - m_I)^2 / \theta]$ – дисперсии границ оценки, $\sigma_g^2 = M[(\Theta^* - m_{\theta^*/g})^2 / \theta, g]$ – условная дисперсия оценки, $\varepsilon_{S0} = (m_S - \theta)$, $\varepsilon_{I0} = (m_I - \theta)$ – смещения (систематические погрешности) относительно соответственно верхней и нижней границ функции распределения оценки и $\varepsilon_{0/g} = (m_{\theta^*/g} - \theta)$ – смещение (систематическая погрешность) в условиях g (рис. 2).

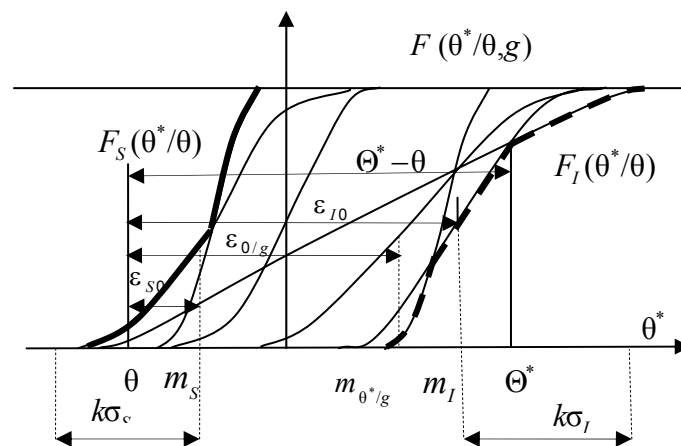


Рисунок 2 – Вер условных функций распределения $F(\theta^*/\theta, g)$ (тонкие кривые) для различных условий g , верхняя $F_S(\theta^*/\theta)$ (жирная непрерывная кривая) и нижняя $F_I(\theta^*/\theta)$ (жирная пунктирная кривая) границы функции распределения.

Для характеристики абсолютной погрешности оценки $\Delta = \Theta^* - \theta$ скалярного параметра θ можно использовать интервал $[\varepsilon_{S0} - k\sigma_S, \varepsilon_{I0} + k\sigma_I]$, определяемый систематическими погрешностями ε_{S0} , ε_{I0} , средноквадратичными отклонениями границ σ_S , σ_I и некоторой константой k , а для характеристики оценки Θ^* – интервал $[m_S - k\sigma_S, m_I + k\sigma_I]$ (см. рис. 2). Если условные функции распределения случайных величин $\Theta^* / \theta, g$ не пересекаются и с ростом условных математических ожиданий оценки условные дисперсии оценки возрастают (тип распределения «а») или уменьшаются (тип распределения

«б»), последний интервал определяется границами математического ожидания оценки m_s , m_i и границами среднеквадратичного отклонения оценки σ_s , σ_i . Для распределения типа «а» он равняется $[m_i - k\sigma_i, m_s + k\sigma_s]$, а для распределения типа «б» – $[m_i - k\sigma_s, m_s + k\sigma_i]$.

Гиперслучайная оценка $\vec{\Theta}^*$ фиксированного параметра $\vec{\theta}$ названа состоятельной, если при всех условиях $g \in G$ она сходится по вероятности к этому параметру:

$$\lim_{N \rightarrow \infty} P\{\vec{\Theta}^* - \vec{\theta} > \varepsilon / \vec{\theta}, g\} = 0 \quad \forall g \in G, \text{ где } N - \text{объем выборки для каждого условия } g, \varepsilon > 0 -$$

как угодно малое число. Состоятельность оценки означает, что она удовлетворяет закону больших чисел.

Необходимое условие состоятельности гиперслучайной оценки – вырождение ее в случайную величину при $N \rightarrow \infty$. Отсюда следует, что оценки, сохраняющие гиперслучайный характер при $N \rightarrow \infty$, являются несостоятельными.

В работе [22] была выдвинута гипотеза о том, что все реальные физические массовые явления происходят в постоянно меняющихся статистических условиях, а потому все физические явления, обычно рассматриваемые как случайные, в действительности носят не случайный, а гиперслучайный характер. Эта особенность имеет место не только при конечном, а и при бесконечно большом интервале наблюдения. Из этого следует, что оценки реальных величин и функций являются несостоятельными. Принимая гипотезу гиперслучайности, следует признать, что, достичь бесконечно большой точности измерений принципиально невозможно ни при каких условиях.

Границы среднего квадрата абсолютной погрешности Δ_s^2 , Δ_i^2 оценки Θ^* параметра θ , сформированной по выборке \vec{X} размером N , и границы $D_s[\Theta^*/\theta]$, $D_i[\Theta^*/\theta]$ условной дисперсии $D[\Theta^*/\theta, g]$ этой оценки описываются неравенствами [22]

$$\Delta_s^2 \geq D_s[\Theta^*/\theta] \geq \sup_{g \in G} \left[\left(1 + \frac{\partial \varepsilon_{0/g}}{\partial \theta} \right)^2 J_{N/g}^{-1} \right], \quad \Delta_i^2 \geq D_i[\Theta^*/\theta] \geq \inf_{g \in G} \left[\left(1 + \frac{\partial \varepsilon_{0/g}}{\partial \theta} \right)^2 J_{N/g}^{-1} \right],$$

где $J_{N/g}$ – информация по Фишеру для случайной величины $\Theta^*/\theta, g$:

$$J_{N/g} = M \left[\left(\frac{\partial \ln f_N(\vec{X}/\theta, g)}{\partial \theta} \right)^2 \right] = -M \left[\frac{\partial^2 \ln f_N(\vec{X}/\theta, g)}{\partial \theta^2} \right],$$

где $f_N(\vec{x}/\theta, g)$ – плотность вероятности выборки $\vec{X}/\theta, g$.

Средние же квадраты абсолютной погрешности Δ_s^2 , Δ_i^2 относительно границ и дисперсии границ $D_s[\Theta^*/\theta]$, $D_i[\Theta^*/\theta]$ определяются неравенствами

$$\Delta_s^2 \geq D_s[\Theta^*/\theta] \geq \frac{\left(1 + \frac{\partial \varepsilon_{s0}}{\partial \theta} \right)^2}{M_s \left[\left(\frac{\partial \ln f_{SN}(\vec{X}/\theta)}{\partial \theta} \right)^2 \right]}, \quad \Delta_i^2 \geq D_i[\Theta^*/\theta] \geq \frac{\left(1 + \frac{\partial \varepsilon_{i0}}{\partial \theta} \right)^2}{M_i \left[\left(\frac{\partial \ln f_{IN}(\vec{X}/\theta)}{\partial \theta} \right)^2 \right]},$$

где $f_{SN}(\vec{x}/\theta)$, $f_{IN}(\vec{x}/\theta)$ – плотности вероятностей верхней и нижней границ \vec{X}/θ .

Выводы

1. Гиперслучайное явление – это математическая структура, описывающая совокупность случайных событий, величин или функций. Каждый элемент структуры ассоциируется с фиксированными статистическими условиями его наблюдения или формирования. Вероятностная мера определяется для каждого элемента структуры (фиксированных условий), однако для всей структуры, мера остается неопределенной.
2. Гиперслучайная величина или функция может быть описана верхней и нижней границами функции распределения. Между этими границами расположена зона неопределенности. Случайное явление – вырожденное гиперслучайное явление. Для случайной величины или функции границы функции распределения совпадают, а зона неопределенности отсутствует. Полный хаос – тоже вырожденное гиперслучайное явление. В этом случае зона неопределенности максимальна.
3. Основными характеристиками гиперслучайных величин и функций, наряду с границами функции распределения, являются различные характеристики этих границ, в частности плотность вероятности границ, математические ожидания границ, дисперсии границ и др., а также границы характеристик, в частности границы плотности вероятности, границы математического ожидания, границы дисперсии и др.
4. Приведены результаты исследований оценок гиперслучайных величин и функций. Отмечено, что все реальные статистические условия наблюдения постоянно меняются. Поэтому все реальные физические массовые явления, обычно рассматриваемые как случайные, в действительности носят не случайный, а гиперслучайный характер. Эта особенность имеет место не только при конечном, а и при бесконечно большом интервале наблюдения. Из этого следует, что все оценки реальных величин и функций являются несостоятельными, а это означает, что достичь бесконечно большой точности измерений принципиально невозможно ни при каких условиях.

Литература

1. Леман Е. Проверка статистических гипотез. Пер. с англ./Пер. Ю.В. Прохорова. – М.: Наука, 1971. – 375 с.
2. Королюк В.С. и др. Справочник по теории вероятностей и математическая статистика. – М.: Наука, 1985. – 637с.
3. Левин Б.Р. Теоретические основы статистической радиотехники. Т. 3. М., Сов. радио, 1976. – 285 с.
4. Под ред. Бакута П. А. Теория обнаружения сигналов. М., «Радио и связь», 1984. – С.440.
5. Ван Трис Г. Теория обнаружения, оценок и модуляции. – М.: Сов. радио, 1972. – Т. 1. – 743 с.; 1975. – Т. 2. – 343 с.; 1977. – Т. 3. – 662 с.
6. Хьюбер П. Робастность в статистике. М., Мир, 1984. – 303с.
7. Кнопов П.С., Голодников А.Н., Пепеляев В.А. Оценивание параметров надежности при наличии неполной первичной информации – Компьютерная математика, №1, 2003. – С. 15 – 23.
8. Кравцов Ю.А. Случайность, детерминированность, предсказуемость. Успехи физических наук, Т. 158, вып. 1, 1989. С. 93 – 122.
9. L.A. Zadeh and J. Kasprzyk (Eds.), "Fuzzy logic for the management of uncertainty," John Wiley & Sons, New York, 1992.
10. M.T. Hagan, H.B. Demuth, and M.H. Beale, "Neural network design," Boston, MA: PWS Publishing, 1996.
11. R.M. Crownover, "Introduction to fractals and chaos," Jones and Bartlett Pub., Inc., Boston – London, 1995.
12. Гринченко В.Т., Мацапура В.Т., Снарский А.А.. Введение в нелинейную динамику. К.: Наукова думка, 2005. – 263 с.
13. Вошинин, Бочков А. Ф., Сотиров Г.Р. Метод анализа данных при интервальной нестатистической ошибке // Заводская лаборатория. – 1990. – Т. 56, №7. – С. 76 – 81.
14. Алефельд Г., Херцбергер Ю. Введение в интервальные вычисления. – М.: Мир, 1987. – 356 с.
15. Горбань И.И. Гиперслучайные явления и их описание // Акустичний вісник. – 2005. – т.8, № 1 – 2. – С. 16 – 27.
16. Колмогоров А. Н. Основные понятия теории вероятностей, ОНТИ, 1936. – 175 с.

17. R. von Mises, "Mathematical theory of probability and statistics," Edited and complemented by H. Geiringer. N.Y. and London, Acad. Press, 1964.
18. Горбань И.И. Гиперслучайные функции и их описание // Радиоэлектроника. – 2006. – № 1. – С. 3 – 15.
19. Горбань И.И. Методы описания гиперслучайных величин и функций // Акустичний вісник. – 2005. – т. 8, № 3. – С. 24 – 33.
20. Горбань И.И. Стационарные и эргодические гиперслучайные функции. // Радиоэлектроника. – 2006. – № 6. – С. 54 – 70.
21. Горбань И.И. Оценки характеристик гиперслучайных величин // Математические машины и системы. – 2006. – № 1. – С. 40 – 48.
22. Горбань И.И. Точечный и интервальный методы оценки параметров гиперслучайных величин. // Математические машины и системы. – 2006. – № 2. – С. 3 – 14.
23. Горбань И.И. Представление физических явлений гиперслучайными моделями. // Математические машины и системы. – 2007. – № 1. – С. 34 – 41.
24. Горбань И.И. Теория гиперслучайных явлений. – К.: Аверс, 2007. – 180 с.

Сведения об авторе

Горбань Игорь Ильич – заместитель генерального директора по научной работе ГП „УкрНИУЦ”, доктор технических наук, профессор, Киев, Украина. E-mail: gorban@ukrmdnc.org.ua.

АНАЛИЗ МОДЕЛИ ЛЕОНТЬЕВА ПРИ НЕЧЁТКО ЗАДАННЫХ ПАРАМЕТРАХ МЕТОДОМ БАЗИСНЫХ МАТРИЦ

Алексей Волошин, Владимир Кудин, Григорий Кудин

Аннотация. Рассматривается модель Леонтьева (МЛ) с нечетко заданными параметрами. МЛ интерпретируется как задачу прогноза «затраты-выпуск» на основе известной статистической информации при нечётко заданных значениях элементов технологической матрицы, вектора ограничений и границ переменных.

Ключевые слова: модель Леонтьева, количественный и качественный анализ, нечёткое множество, базисная матрица, функция принадлежности.

Введение

Математический аппарат нечётких множеств является формой задания неопределённых параметров, значения которых неизвестны до момента принятия решения. Одним из механизмов устранения неопределённости в задании параметров при построении модели является наличие в контуре принятия решения экспертов – лиц, принимающих решение (ЛПР). ЛПР качественно определяют структуру модели, указывают механизм устранения неопределённостей при ее формировании [Орловский, 1981]. Существенным усложнением МЛ [Леонтьев, 1972], [Гасс., 1961] является включение ограничений на значения переменных [Орловский, 1981]. Одной из особенностей МЛ является то, что она содержит в себе математические проблемы анализа линейных систем таких, как системы линейных алгебраических уравнений с квадратной невырожденной матрицей ограничений (СЛАУ), линейных алгебраических неравенств с соответствующей матрицей ограничений (СЛАН), а также и задач линейного программирования (ЗЛП) [Волошин, 1993], [Войналович, 1987, 1988], [Кудин, 2002]. Осуществление качественного анализа модели [Орловский, 1981] подразумевает включение и количественного анализа

непротиворечивости ее структурных элементов [Волошин, 1993, 2004], [Войналович, 1987, 1988], [Кудин, 2002]. Можно выделить такие основные стадии анализа:

- проверка математической и «машинной» невырожденности матрицы ограничений, определение ее ранга;
- направленная «коррекция» ранга матрицы ограничений путем изменения отдельных её элементов (при необходимости);
- выявление совместных свойств МЛ и дополнительных ограничений на переменные – (определение разрешимости);
- определение свойств МЛ для многогранного множества ограничений на переменные, проведение, при необходимости, направленных изменений;
- нахождение решений (в случае разрешимости);
- анализ свойств полученных решений.

Постановка задачи

Введем в рассмотрение варианты МЛ, к которым можно прийти в результате эквивалентных преобразований канонической модели:

1. СЛАР вида

$$Au = C, \quad (1)$$

2. СЛАН вида

$$Au \leq C, \quad (2)$$

3. Модели СЛАН (2) могут исследоваться при наличии целевой функций вида

$$\max_{u \in R^m} Bu, \quad (3)$$

как задача анализа модели линейного программирования (2)-(3), в которой $A = \{a_{ij}\}_{i=1, \dots, m}^{j=1, \dots, m}$ невырожденная квадратная матрица размерности $(m \times m)$, $a_j = (a_{j1}, a_{j2}, \dots, a_{jm})$, $j \in J = I = \{1, 2, \dots, m\}$ – строки матрицы A , $u = (u_1, u_2, \dots, u_m)^T$ – вектор переменных, $B = (b_1, b_2, \dots, b_m)$, $C = (c_1, c_2, \dots, c_n)^T$ – вектора градиента целевой функции и ограничений модели, $a_j u \leq c_j$, $j \in J$ полупространство, которое определено гиперплоскостью $a_j u = c_j$, $j \in J$.

Считаем, что элементы ограничения $a_l u \leq c_l$, $l \in J$, нормаль которого занимает, k -ю могут претерпевать изменения согласно соотношений $a_l(t) u \leq c_l(t)$, $l \in J$, где $a_l(t) = (a_{l1}(t), a_{l2}(t), \dots, a_{lm}(t))$, $l \in J$ (элемент модели a_{lr} становится равным $a_{lr}(t)$, $r \in I$, а c_l будет $c_l(t)$). Все определенные функции зависимы от аргумента $t \in (-\infty, +\infty)$ из класса C^2 . Такие изменения в элементах модели можно интерпретировать, как влияние изменений цен в интервале $t \in (t_H, t_B)$ на значение технологических элементов модели (2) по k -у ресурсу.

Предполагается, что в системе имеется $P = \{1, 2, \dots, p\}$ экспертов. Каждый эксперт формирует свою функцию принадлежности $\mu_p(t)$, $p \in P$. Эти функции являются кусочно-линейными, для которых эксперт устанавливает уровни значений $\lambda^{(p)}$, $p \in P$. Это означает, что при $1 \geq \mu_p(t) \geq \lambda^{(p)}$ определяется интервал изменения значений T_p , $p \in P$, где $T_p = [t_{p(H)}^{(-)}, t_{p(B)}^{(+)}] \subseteq (-\infty, +\infty)$, $p \in P$ [Орловский, 1981]. Результирующий интервал изменения переменной $T = (t_H, t_B)$ согласованный по P экспертам может определяться, например, как $T = \bigcap_{p=1}^P T_p$ (или $T = \bigcup_{p=1}^P T_p$). Модель (1) исследуется в пространстве E^m . При наличии в контуре принятия решения экспертов (ЛПР) фаза качественного анализа моделей (1),(2)-(3)

определяет последующую задачу, как задачу количественного анализа - исследования при указанных уровнях $\lambda^{(p)}$, $p = \{1, 2, \dots, p\}$ указанных экспертами влияние изменения элементов ограничения согласно соотношению $a_l(t)u \leq c_l(t)$, $l \in J$ при $t \in T$ на ранее выбранное оптимальное решение.

В работе предложено развитие методологии последовательного анализа [Волошин, 1987] и метода базисных матриц (МБМ) [Кудин, 2002] для проведения количественного анализа влияния функциональных изменений в МЛ на ее свойства такие, как невырожденность матрицы ограничений, оптимальные решения исходной задачи (2)-(3) при изменении элементов модели (k -го ограничения) в виде $a_l(t)u \leq c_l(t)$, $l \in J$, $t \in T$.

Основные положения метода базисных матриц (МБМ)

В предлагаемом МБМ введены в рассмотрение строчные базисные матрицы [Войналович, 1987, 1988], [Кудин, 2002]. Базисные матрицы в ходе итераций решения задачи последовательно изменяются вводом-выводом из нее строк-нормалей ограничений. В общем случае в исследуемой модели количество ограничений превышает количество переменных вида (2), а в данном случае в МЛ $m = n$:

Определение 1. Матрицу $A_{\bar{\sigma}}$, составленную из m линейно независимых нормалей ограничений (2), будем называть базисной, а решение соответствующей ей системы уравнений $A_{\bar{\sigma}} u_0^T = C^0$ базисным. Две базисные матрицы отличающиеся одной строкой будем называть смежными.

Пусть: β_{ij} , $i, j \in I = \{1, 2, \dots, m\}$ - элементы базисной подматрицы $A_{\bar{\sigma}}$, e_{ri} - элементы матрицы $A_{\bar{\sigma}}^{-1}$, обратной к $A_{\bar{\sigma}}$; $e_k = (A_{\bar{\sigma}}^{-1})_k$ - столбец обратной матрицы. Решение $u_0 = (u_{01}, u_{02}, \dots, u_{0m})$ системы уравнений $A_{\bar{\sigma}} u^T = c^0$, где, в общем случае, c^0 - подвектор C , компоненты которого состоят из правых частей ограничений (2), нормали которых образуют базисную матрицу $A_{\bar{\sigma}}$; $\alpha_r = (\alpha_{r1}, \alpha_{r2}, \dots, \alpha_{rm})$ - вектор разложения нормали ограничения $a_r u_1 \leq c_r$ за строками базисной матрицы $A_{\bar{\sigma}}$, $\alpha_0 = (\alpha_{01}, \alpha_{02}, \dots, \alpha_{0m})$ - вектор разложения градиента целевой функции (3) по строкам базисной матрицы $A_{\bar{\sigma}}$, $\Delta_r = a_r u_0^T - c_r$ - невязка r -го ограничения (2) в вершине u_0 ; $J_{\bar{\sigma}}, J_H$, $J = J_{\bar{\sigma}} \cup J_H$ - множества индексов базисных и небазисных ограничений (2). В работе [Войналович, 1987] приведены формулы связи базисного решения, коэффициентов разложения нормалей ограничений и целевой функции (3), коэффициентов обратной матрицы, невязок ограничений и значений целевой функции при переходе к базисной матрице $\bar{A}_{\bar{\sigma}}$, которая образуется из матрицы $\bar{A}_{\bar{\sigma}}$ заменой ее строки a_k на a_l , которая не входит в базисную матрицу $A_{\bar{\sigma}}$. В новой базисной матрице $\bar{A}_{\bar{\sigma}}$ введенные величины будем называть элементами метода базисных матриц и будем обозначать черточкой сверху, т.е. $\bar{\beta}_{ij}$, $\bar{\alpha}_r$, $\bar{\Delta}_k$, \bar{e}_{ri} , $\bar{\alpha}_0$. Пусть $a_{i1}, a_{i2}, \dots, a_{im}$ - нормали ограничений, $a_j u^T \leq c_j$, $j \in J_{\bar{\sigma}}$, где $J_{\bar{\sigma}} = \{i_1, i_2, \dots, i_m\}$ - индексы ограничений, нормали которых образуют строки базисной матрицы $A_{\bar{\sigma}}$, a_l - нормаль ограничения $a_l u \leq c_l$, $\alpha_l = (\alpha_{l1}, \alpha_{l2}, \dots, \alpha_{lm})$ - коэффициенты разложения вектора a_l по строкам базисной матрицы $A_{\bar{\sigma}}$.

Лемма 1. (Критерий линейной независимости системы векторов). Необходимым и достаточным условием линейной независимости строк матрицы модели $a_{i_1}, a_{i_2}, \dots, a_{i_{k-1}}, a_l, a_{i_{k+1}}, \dots, a_{i_m}$, образованных заменой строки a_{i_k} , которая занимает k -ю строку в базисной матрице $A_{\bar{\sigma}}$, строкой a_l , является выполнение условия $\alpha_{lk} \neq 0$.

Теорема 1. (О связи между смежными базисными матрицами). Между коэффициентами разложения нормалей ограничений (2) и целевой функции (3) за строками базисной матрицы, элементами обратных матриц, базисными решениями, невязками ограничений (2) и значениями целевой функции для двух смежных базисных матриц имеют место соотношения

$$\bar{\alpha}_{rk} = \frac{\alpha_{rk}}{\alpha_{lk}}, \quad \bar{\alpha}_{ri} = \alpha_{ri} - \frac{\alpha_{rk}}{\alpha_{lk}} \alpha_{li}, \quad r = \overline{0, n}; \quad i = \overline{1, m}; \quad i \neq k; \quad (4)$$

$$\bar{e}_{rk} = \frac{e_{rk}}{\alpha_{lk}}, \quad \bar{e}_{ri} = e_{ri} - \frac{e_{rk}}{\alpha_{lk}} \alpha_{li}, \quad r = \overline{1, m}; \quad i = \overline{1, m}; \quad i \neq k; \quad (5)$$

$$\bar{u}_{0j} = u_{0j} - \frac{e_{jk}}{\alpha_{lk}} \Delta_l, \quad j = \overline{1, m}, \quad (6)$$

$$\bar{\Delta}_k = -\frac{\Delta_l}{\alpha_{lk}}, \quad \bar{\Delta}_r = \Delta_r - \frac{\alpha_{rk}}{\alpha_{lk}} \Delta_l, \quad r = \overline{1, n}; \quad r \neq k; \quad (7)$$

$$\bar{B}u_0 = Bu_0 - \frac{\alpha_{0k}}{\alpha_{lk}} \Delta_l, \quad (8)$$

причем условием того, что матрица остаётся базисной при замещении вектором a_i k -й строки базисной матрицы A_σ , есть выполнение условия $\alpha_{lk} \neq 0$, условием допустимости опорного базисного решения есть $\alpha_{lk} < 0$, роста значений целевой функции $\alpha_{0k} < 0$.

Доказательство леммы 1 и теоремы 1 основывается на теоретических положениях, изложенных в [Войналович, 1987, 1988], [Кудин, 2002].

Соотношения (4)-(8) будут основополагающими при построении алгоритма поиска не только оптимального решения, но проведения анализа свойств МЛ методом базисных матриц.

Определение 2. Допустимое базисное решение u_0 оптимальное, если $Bu_0 \geq Bu$ для всех u , которые удовлетворяют (2).

Теорема 2. Для оптимальности базисного решения u_0 необходимо и достаточно неотрицательности коэффициентов разложения вектора нормали целевой функции (3) по строкам базисной матрицы A_σ , т.е. $\alpha_{0k} \geq 0$ для всех $k = \overline{1, m}$, причем задача (2),(3) с квадратной невырожденной матрицей ограничений имеет единственное решение тогда и только тогда, когда $\alpha_{0i} > 0$, $i = \overline{1, m}$, а необходимым и достаточным условием не единственности решения задачи есть $\exists i \in I$ таких, что $\alpha_{0i} = 0$, причём множество решений имеет ребра неограниченности.

Справедливость критерия оптимальности вытекает из формулы (8) теоремы 1.

Следствие 1. При тех значениях переменной функции, что $\hat{\alpha}_{lk}(t) = a_l(t)(A_b^{-1})_k \neq 0$, $t \in [t_0, t_k]$ сохраняется неизменной величина ранга при изменении ограничения k МЛ на ограничение $a_l(t)u \leq c_l(t)$, $l \in J$ при $t \in T$. Значения t при которых $\hat{\alpha}_{lk}(t) = a_l(t)(A_b^{-1})_k = 0$, $t \in [t_0, t_k]$, $l \notin J_\sigma$ уменьшают ранг системы (матрица ограничений модели (2) становится вырожденной).

Справедливость следствия вытекает из леммы 1.

Следствие 2. Достаточным условием сохранения оптимальности решением при изменении k -го ограничения модели в виде $a_l(t)u \leq c_l(t)$, $l \in J_\sigma$ нормаль которого занимает k -ю строку в базисной

матрице является выполнение соотношений: $a_l(t)(A_\sigma^{-1})_k > 0$, $\frac{\alpha_{0k}}{\alpha_{lk}(t)} \geq 0$, $\frac{\alpha_{0i}}{\alpha_{0k}} \geq \frac{\alpha_{li}(t)}{\alpha_{lk}(t)}$.

$t \in T$, $i \neq k$, $i \in I$. Условием единственности решения (2), (3) являются $\frac{\alpha_{0k}}{\alpha_{lk}(t)} > 0$, $t \in T$,

$\frac{\alpha_{0i}}{\alpha_{0k}} > \frac{\alpha_{li}(t)}{\alpha_{lk}(t)}$, $t \in T$, $i \neq k$, $i \in I$.

Доказательство. Согласно леммы 1 и следствия 1 условием невырожденности матрицы модели (2) является выполнение при всех $t \in T$ соотношения $\alpha_{ik}(t) = a_i(t)(A_\sigma^{-1})_k \neq 0$. Из формулы (4) и теоремы 2 следует, что при операции замещения k -й строки базисной матрицы для выполнения условий оптимальности необходимо выполнение $\bar{\alpha}_{0k} = \frac{\alpha_{0k}}{\alpha_{ik}(t)} \geq 0$, $\bar{\alpha}_{0i} = \alpha_{0i} - \frac{\alpha_{0k}}{\alpha_{ik}(t)}\alpha_{li}(t) \geq 0, i \neq k, i \in I, t \in T$, поскольку условие оптимальности $\alpha_{0i} \geq 0, i \in I$. Отсюда вытекает выполнение условия $\alpha_{ik}(t) = a_i(t)(A_\sigma^{-1})_k \geq 0$ и $\frac{\alpha_{0i}}{\alpha_{0k}} \geq \frac{\alpha_{li}(t)}{\alpha_{ik}(t)}, t \in T$ для сохранения оптимальности решения. Отсюда непосредственно, с учетом теоремы 2, следует справедливость условий единственности (не единственности) решений (2)-(3).

Выводы

Применение симплексной идеологии на основе метода базисных матриц для качественного и количественного анализа модели Леонтьева даёт возможности:

- исследовать свойства решений систем линейных уравнений и неравенств при изменениях в векторах ограничений;
- проводить анализ свойств модели Леонтьева при изменении значений отдельных ее элементов;
- использовать решение исходной модели Леонтьева при анализе возмущенной модели;
- контролировать или направлено изменять ранга системы;
- находить решение квадратной системы уравнений за фиксированное количество шагов;
- строить начальные решения на основе тривиальных базисных матриц, что исключает трудоёмкие начальные вычисления;
- применять схему анализа для задач, которые предусматривают многошаговость или многократность расчетов при изменениях в компонентах модели.

Литература

- [Леонтьев, 1972] Леонтьев В.В., Форд Д. Межотраслевой анализ воздействия структуры экономики на окружающую среду // Экономика и математические методы.- 1972.-Т.VII.-Вып.3.-С.370-400.
- [Гасс, 1961] Гасс С. Линейное программирование. Физматгиз,-1961.
- [Волошин, 1987] Волошин А.Ф. Метод локализации области оптимума в задачах математического программирования // Докл. АН СССР. - 1987. -293, N 3.- С. 549-553.
- [Волошин, 2004] Волошин А.Ф. Метод анализа стичных балансовых эколого=экономических моделей большой размерности // Научные записки КНУ им Т. Шевченко. Том 7.– Киев, 2004. - С. 43-55.
- [Орловский, 1981] Орловский С.А Принятие решения при нечёткой исходной информации.- М.: Наука,-1981,- 206с.
- [Волошин,1993] Волошин А.Ф. Войналович В.М., Кудин В.И. Предоптимизационные и оптимизационные схемы сокращения размерности задачи линейного программирования // Автоматика,N4, 1993.
- [Войналович, 1987] Волкович В.Л., Войналович В.М., Кудин В.И. Релаксационная схема строчного симплекс метода // Автоматика .- 1987. -N4.-С. 79-86.
- [Войналович, 1988] Волкович В.Л., Войналович В.М., Кудин В.И. Релаксационная схема двойственного строчного симплекс метода // Автоматика.-1988. -N 1,-С.39-46.
- [Кудин, 2002] Кудин В.И. Применение метода базисных матриц при исследовании свойств линейной системы // Вестник Киевского университета. Серия физ.-мат. науки. - 2002.-2., С. 56-61.

Авторы

Алексей Ф. Волошин – Киевский национальный университет им. Т. Шевченко, профессор, факультет кибернетики; пр. Акад. Глушкова ,2, корп. 6, Киев, Украина; e-mail: ovoloshin@unicyb.kiev.ua

Владимир И. Кудин – Киевский национальный университет им. Т. Шевченко, старший научный сотрудник, факультет кибернетики; пр. Акад. Глушкова ,2, корп. 6, Киев, Украина; e-mail: V.I.Kudin@mail.ru

Григорий И. Кудин – Киевский национальный университет им. Т. Шевченко, старший научный сотрудник, факультет кибернетики; пр. Акад. Глушкова ,2, корп. 6, Киев, Украина; e-mail: kuding@mail.univ.kiev.ua

ВЕКТОРНЫЕ КОМБИНАТОРНЫЕ ЗАДАЧИ В ПРОСТРАНСТВЕ СОЧЕТАНИЙ С ДРОБНО-ЛИНЕЙНЫМИ ФУНКЦИЯМИ КРИТЕРИЕВ

Наталья Семенова, Людмила Колечкина, Алла Нагорная

Аннотация: Рассматривается векторная задача дискретной оптимизации с дробно – линейными функциями критериев на допустимом множестве, которое имеет комбинаторные свойства сочетаний. Исследуются структурные свойства области допустимых решений и множеств Парето – оптимальных (эффективных), строго эффективных, слабо эффективных решений. Установлена взаимосвязь между задачами векторной оптимизации на комбинаторном множестве сочетаний и на непрерывном допустимом множестве. Предлагается один из возможных подходов для решения многокритериальной комбинаторной задачи с дробно-линейными целевыми функциями на множестве сочетаний.

Ключевые слова: векторная оптимизация, дискретная оптимизация, дробно-линейные функции, множество сочетаний.

Введение

Большее число задач управления, планирования, проектирования и другие решаются с применением комбинаторных оптимизационных моделей. Это обстоятельство привело к появлению работ, которые посвящены исследованию комбинаторных задач оптимизации [1,2,4]. Наряду с этим большой интерес представляют работы, которые посвящены изучению различных аспектов решения многокритериальных задач оптимизации, в частности дискретных [1-6]. Интерес к исследованию проблем решения многокритериальных моделей дискретной оптимизации обусловлен их широким применением для решения важных задач экономики, проектирования сложных систем, принятия решений в условиях неопределенности и других. Задачи векторной оптимизации возникают в ситуациях, когда нужно оптимизировать более чем одну целевую функцию. Дробно – линейные функции цели, являющиеся отношением двух линейных форм, имеют широкий диапазон применений в задачах, где требуется оптимизировать некоторые относительные показатели качества, такие как себестоимость, рентабельность, производительность, трудоемкость и другие. Модели, использующие такие критерии, отражают тенденции постоянного снижения уровня себестоимости в расчете на единицу продукции и повышения качественных показателей производства при увеличении масштабов производства. Представляет интерес рассмотрение таких задач как задач комбинаторной оптимизации на допустимом множестве сочетаний.

Как известно, большинство комбинаторных оптимизационных задач могут быть сведены к задачам целочисленного программирования, но это не всегда оправдано, поскольку при этом теряется возможность учета комбинаторных свойств задачи [1]. В монографиях [2,4] показано, что выпуклой оболочкой множества сочетаний является многогранник сочетаний Π , множество вершин которого $vert\Pi$ равно множеству S_A сочетаний, то есть $vert\Pi = vert(convS_A)$. Данное свойство многогранника сочетаний позволяет свести решение задачи исходной задачи к решению задачи на непрерывном допустимом множестве. Поэтому достаточно важным и актуальным является исследование и построение методов решения многокритериальных задач на основе исследования выпуклых оболочек таких множеств с учетом комбинаторных свойств области допустимых решений, в частности сочетаний.

В настоящее время существует большое количество методов решения многокритериальных задач, но ни один из них в существующем виде не применим к комбинаторным задачам на сочетаниях, поэтому нам представляется достаточно важным рассмотреть возможные подходы к решению многокритериальных задач на комбинаторных множествах сочетаний.

В данной работе исследуются структурные свойства области допустимых решений и множеств Парето – оптимальных (эффективных), строго эффективных, слабо эффективных решений. Установлена взаимосвязь между задачами векторной оптимизации на комбинаторном множестве сочетаний и на непрерывном допустимом множестве. Для задач комбинаторного типа на сочетаниях предложен один из возможных подходов к их решению.

Постановка задачи. Основные определения

Для постановки задачи используем понятие мультимножества A , которое определяется основанием $S(A)$ и кратностью элементов $k(a)$.

Пусть задано мультимножество $A = \{a_1, a_2, \dots, a_g\}$, а $S(A) = \{e_1, e_2, \dots, e_n\}$ его основание, где $e_j \in R_1$ $\forall j \in N_n = \{1, 2, \dots, n\}$, и кратность элементов $k(a_j) = r_j, j \in J_n, r_1 + r_2 + \dots + r_n = g$.

Возьмем произвольное $k \in N_g$. Упорядоченной k -выборкой из мультимножества A называется набор

$$a = (a_{i_1}, a_{i_2}, \dots, a_{i_k}), \quad (1)$$

где $a_{i_j} \in A \forall i_j \in N_k, \forall j \in N_k, i_s \neq i_t$, если $s \neq t \forall s \in N_k, \forall t \in N_k$.

Определение. [2] Множество $P(A)$, элементами которого есть k -выборки вида (1) из мультимножества A , называется евклидовым комбинаторным множеством, если для произвольных его элементов $a' = (a'_1, a'_2, \dots, a'_k), a'' = (a''_1, a''_2, \dots, a''_k)$ выполняются условия: $(a' \neq a'') \Leftrightarrow (\exists j \in J_k : a'_j \neq a''_j)$, то есть множество $P(A)$ имеет такое свойство: два элемента множества $P(A)$ отличны друг от друга, если они независимо от других отличий различаются порядком размещения символов, которые их образуют.

Пусть мультимножество A является множеством.

Множество всех k -выборок из множества $S(A)$ вида $(a_{i_1}, a_{i_2}, \dots, a_{i_k})$ называют евклидовым множеством k -сочетаний без повторения из n различных действительных чисел, если выполняется условие $e_{i_1} < e_{i_2} < \dots < e_{i_k}$. Обозначим это множество сочетаний $S_n^k(A)$.

Совокупность всех k -выборок вида (1) из мультимножества A называется общим множеством сочетаний $S_{gn}^k(A)$, если выполняется условие $a_{i_1} \leq a_{i_2} \leq \dots \leq a_{i_k}$.

Интерес к евклидовым комбинаторным множествам связан с возможностью рассматривать их как точки пространства R^k , то есть с возможностью погружения их в арифметическое евклидовое пространство R^k . Пусть $P(A)$ – комбинаторное множество, а элемент $P(A)$ описывается формулой (1). Отображение

$f: P(A) \rightarrow P_f(A) \subset R^k$ называется погружением множества $P(A)$ в арифметическое евклидово пространство, если f ставит множество $P(A)$ во взаимно однозначное соответствие множеству $P_f(A) \subset R^k$ по правилу: для $a = (a_{i_1}, \dots, a_{i_k}) \in P(A)$, $x = f(a)$, $x = (x_1, \dots, x_k) \in P_f(A)$, имеем $x_j = a_{i_j} \quad \forall j \in N_k$.

Рассмотрим многокритериальные оптимизационные задачи на комбинаторных множествах сочетаний следующего вида:

$$Z_l(F, X): \max \{F(x) \mid x \in X\},$$

где $F(x)$ – векторный критерий, $F(x) = (f_1(x), f_2(x), \dots, f_l(x))$,

$$f_i(x) = \frac{\langle c^i, x \rangle + c_0^i}{\langle d^i, x \rangle + d_0^i}, \quad i \in N_l, \quad (2)$$

$X = \Pi \cap D$, где $\Pi = \text{conv } S_{gn}^k(A)$, $D = \{x \in R^k \mid Ax \leq b\}$, D – выпуклый многогранник в R^k , $A \in R^{m \times k}$, $b \in R^m$.

Под решением задачи $Z_l(F, X)$ будем понимать задачу поиска элементов одного из следующих множеств: $P(F, X)$ – множества Парето-оптимальных (эффективных решений), $Sl(F, X)$ – оптимальных по Слейтеру (слабо эффективных) решений, $Sm(F, X)$ – оптимальных по Смейлу (строго эффективных) решений. Согласно [3, 6] для любого $x \in X$ истинны утверждения:

$$x \in Sl(F, X) \Leftrightarrow \{y \in X \mid F(y) > F(x)\} = \emptyset, \quad (3)$$

$$x \in P(F, X) \Leftrightarrow \{y \in X \mid F(y) \geq F(x), F(y) \neq F(x)\} = \emptyset, \quad (4)$$

$$x \in Sm(F, X) \Leftrightarrow \{y \in X \mid y \neq x, F(y) \geq F(x)\} = \emptyset. \quad (5)$$

Очевидно, что

$$Sm(F, X) \subset P(F, X) \subset Sl(F, X). \quad (6)$$

Из конечности допустимой области X следует не пустота множества $P(F, X)$ и его внешняя устойчивость, которая означает, что $\forall y \in X \exists x \in P(F, X) : F(x) \geq F(y)$.

Известно [2], что выпуклой оболочкой точек евклидового комбинаторного множества сочетаний с повторениями есть многогранник Π сочетаний с повторениями, который описывается следующей системой неравенств:

$$\begin{cases} e_1 \leq x^1, \\ x^i \leq x^{i+1}, \quad \forall i \in J_{k-1}, \\ x^k \leq e_n, \end{cases} \quad (7)$$

где e_i – элементы основания $S(A)$ мультимножества A .

Учитывая комбинаторные свойства многогранника сочетаний, можно сформулировать критерий его вершины, который целесообразно использовать при построении метода решения задачи.

Критерий вершины. [2] Точки $x^i = (x_1^i, x_2^i, \dots, x_k^i) \in R^k$, где $x_j^i = e_1 \quad \forall j \in J_{k-i+1}$,

$x_{j+k-i+1}^i = e_n \quad \forall j \in J_{i-1}, \forall i \in J_{k+1}$ и только они могут быть вершинами многогранника сочетаний с повторениями.

Свойства множеств Парето оптимальных (эффективных), строго эффективных, слабо эффективных решений

Теорема 1. Элементы множеств $Sm(F, X)$ – строго эффективных, $P(F, X)$ – Парето оптимальных, $Sl(F, X)$ – слабо эффективных решений многокритериальных задач на комбинаторных множествах сочетаний находятся в вершинах многогранника сочетаний Π .

Доказательство. Учитывая соотношения (6) между введенными множествами эффективных решений и тот факт, что множество допустимых решений X является подмножеством множества $S_{gn}^k(A)$ сочетаний, справедливы следующие включения: $Sm(F, X) \subset P(F, X) \subset Sl(F, X) \subset S_{gn}^k(A)$.

В соответствии с теоремой 2.4. [2] множество сочетаний $S_{gn}^k(A)$ совпадает с множеством вершин общего многогранника сочетаний $vert\Pi_{gn}^k(A)$, то есть $S_{gn}^k(A) = vert\Pi_{gn}^k(A)$. Таким образом, справедливы включения $Sm(F, X) \subset P(F, X) \subset Sl(F, X) \subset vert\Pi_{gn}^k(A)$. Теорема доказана.

Данное свойство многогранника сочетаний позволяет свести решение задачи $Z_l(F, X)$ к решению задачи $Z_l(F, G)$ на непрерывном допустимом множестве $G = \Pi \cap D$.

Для любой точки $y \in vert\Pi$ данной задачи справедливые достаточные условия оптимальности всех указанных видов эффективных решений, полученные в работе [3].

Справедливая следующая теорема.

Теорема 2. $Sl(F, X) = Sl(F, G) \cap vert\Pi$.

Для строго квазивыпуклых функций критериев в работе [5] доказана следующая теорема.

Теорема 3. Если функции $f_i(x)$, $i \in N_l$, векторного критерия являются строго квазивыпуклыми и полунепрерывными снизу на линейных отрезках X , тогда множество $Sl(F, X)$ слабо эффективных решений задачи является объединением эффективных множеств $P(F, X)$ подзадач $Z_I(F, X)$, $I \subset N_l$, $I \neq \emptyset$, т.е. $Sl(F, X) = \bigcup \{P_I(F, X) : I \subset N_l, |I| \leq k + 1\}$.

Как известно, дробно – линейные функции являются строго квазивыпуклыми на выпуклом множестве, поэтому для множества $Sl(F, X)$ - слабо эффективных решений задачи $Z_l(F, X)$ справедлива эта теорема.

Пусть $x_0 \in X$ – некоторое эффективное (Парето-оптимальное) решение. Введем следующие обозначения:

$$P = (p_1, p_2, \dots, p_l) \in R^{k \times l}, Q = (q_1, q_2, \dots, q_l) \in R^{k \times l}, p_i = (\langle d_i, x^0 \rangle c_i - \langle c_i, x^0 \rangle d_i) \in R^k, i \in L,$$

$q_i = (d_i^0 c_i - c_i^0 d_i) \in R^k, i \in L, y \in R^l, \alpha > 0, \alpha \in R^l$ – некоторый постоянный вектор, и рассмотрим задачу линейного программирования следующего вида:

$$(ЗЛП): \max \{ \langle \alpha, y \rangle \mid Q^T x^0 - (P + Q)^T x + y = 0, x \in \Pi \cap D \}.$$

Для задачи $Z_l(F, X)$ справедлива следующая теорема, устанавливающая взаимосвязь ее решений с решениями задачи (ЗЛП).

Теорема 4. Точка $x_0 \in X$ – является эффективным решением задачи $Z_l(F, X)$ тогда и только тогда, когда (ЗЛП) имеет оптимальное решение (x^*, y^*) при $y^* = 0$.

Для решения многокритериальных задач существует ряд методов, которые позволяют находить элементы указанных множеств решений. Но задача $Z_l(F, X)$ усложняется наличием критериев, которые представляют собой дробно-линейные функции вида (2), что составляет первую особенность выше сформулированной задачи. Вторая особенность задачи $Z_l(F, X)$ - это наличие комбинаторных свойств области допустимых решений. Поэтому при решении данной задачи следует учесть все ее особенности.

На основании доказанных теорем продолжая и развивая работы [2-5] предлагается один из возможных подходов к решению рассмотренной задачи, который заключается в выполнении следующих этапов:

1. на начальном этапе рассматривается некоторая задача с дробно-линейной целевой функцией на множестве сочетаний;
2. для отдельно взятой задачи поочередно делается переход от дробно-линейной целевой функции к линейной функции цели;
3. рассматривается многокритериальная задача с линейными целевыми функциями, но на измененном множестве допустимых решений;
4. осуществляется переход от многокритериальной задачи с линейными целевыми функциями к однокритериальной с линейной целевой функцией.

Рассмотрим более детально второй шаг предлагаемого подхода. Как известно, для решения задач с дробно-линейной функцией цели существует большое количество методов, которые условно делятся на методы линеаризации, параметрические методы, модификации симплекс-методов, среди которых известны есть метод Чарнса и Купера, алгоритм Гилмори и Гомори и другие. Но ни один из них не учитывает комбинаторные условия области допустимых решений.

Для решения задачи $Z(F, X)$ осуществим переход к задачам с линейными функциями цели. Обозначим:

$$y_0 = \frac{1}{\langle d^i, x \rangle + d_0}, \quad z_i = x_i y_0, \quad \forall i \in J_k. \quad (8)$$

Соотношение (8) задает отображение для множества области допустимых решений $\psi(X) = X^t \subset R^{k+1}$.

Если $\langle d^i, x \rangle + d_0 > 0$, то $y_0 > 0$, $x_i \geq 0, \forall i \in J_k$, а $\psi(x) = z = (y_0, z_1, \dots, z_k)$.

При таких условиях задача с дробно-линейной целевой функцией сводится к задаче с линейной целевой функцией

$$f'(x^*) = f(z^*) = \langle c^i, z \rangle, \quad (9)$$

при условии

$$X' = \Pi' \cap D'. \quad (10)$$

Следует отметить, что, если $z = (y_0, z_1, \dots, z_k)$ – решение задачи (9), (10) с линейной функцией цели, то она удовлетворяет условию (8), а $x = (x_1^*, \dots, x_k^*)$, удовлетворяет ограничениям задачи $Z_l(F, X)$.

Таким образом, комбинаторная задача с дробно-линейной функцией цели на общем множестве сочетаний может быть сведена к задаче с линейной функцией цели, при этом количество переменных увеличивается на одну, а к системе ограничений прибавляется еще одно равенство $\langle d^i, z \rangle + d_0 y_0^* = 1$.

Выводы

Рассмотрена постановка многокритериальной задачи с учетом комбинаторных свойств области допустимых решений и дробно-линейными функциями критериев. Установлена взаимосвязь между задачей с линейной целевой функцией и векторной задачей с дробно-линейными функциями критериев, определенной на комбинаторном допустимом множестве сочетаний. Это дает возможность применять

классические методы оптимизации к решению векторных комбинаторных задач на множестве сочетаний и на этой основе развивать новые оригинальные методы решения, используя свойства комбинаторных множеств и их выпуклых оболочек.

На основании доказанных теорем, продолжая исследования и развивая результаты работ [2 - 5], предложен подход к решению задачи $Z_i(F, X)$, основанный сведении поиска решения исходной задачи к решению серии скалярных (однокритериальных) задач, проверке оптимальности полученных решений. Методы решения однокритериальных задач основаны на идеях декомпозиции, отсекающих плоскостей Келли, релаксации.

Библиография

- [1] Сергиенко И.В., Каспицкая М.Ф. Модели и методы решения на ЭВМ комбинаторных задач оптимизации. – Киев. – Наукова думка, 1981.- 287 с.
- [2] Стоян Ю.Г., Ємець О.О. Теорія і методи евклідової комбінаторної оптимізації. – Київ: ІСДО, 1993.–188 с.
- [3] Лебедева Т.Т., Семенова Н.В., Сергієнко Т.І. Умови оптимальності та розв'язуваності в задачах лінійної векторної оптимізації з опуклою допустимою множиною // Доповіді НАНУ. – 2003. – №10. – С. 80-85.
- [4] Ємець О.О., Колечкіна Л.М. Задачі комбінаторної оптимізації з дробово – лінійними цільовими функціями. – Київ.- Наукова думка, 2005.- 118 с.
- [5] C. Malivert, N. Boissard. Structure of Efficient Sets for Strictly Quasi convex Objectives // Journal of Convex Analysis – 1994.Vol. 1. – №2. – С.143-150.
- [6] Лебедева Т.Т., Семенова Н.В., Сергиенко Т.И. Устойчивость векторных задач целочисленной оптимизации: взаимосвязь с устойчивостью множеств оптимальных и неоптимальных решений // Кибернетика и сист. анализ. – 2005.– №4.– С. 90-100.

Информация об авторах

Наталія Владимировна Семенова – Інститут кібернетики ім. В.М.Глушкова НАН України, старший науковий співробітник, 03680 МСП Києв 187, проспект академіка Глушкова, 40, Україна; e-mail: nvsemenova@meta.ua

Людмила Николаевна Колечкина - Інститут кібернетики ім. В.М.Глушкова НАН України, докторант, 03680 МСП Києв 187, проспект академіка Глушкова, 40, Україна; e-mail: ludapl@ukr.net

Алла Николаевна Назорная – Полтавський університет потребителської кооперації України, асистент кафедри інформаційно-випчислительних систем, Полтава, вулиця Ковалюка, 3; e-mail: vpn2006@rambler.ru

АППРОКСИМАЦИЯ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ КРИВЫМИ БЕЗЬЕ

Виталий Вишневский, Владимир Калмыков, Татьяна Романенко

Abstract: Very often the experimental data are the realization of the process, fully determined by some unknown function, being distorted by hindrances. Treatment and experimental data analysis are substantially facilitated, if these data to represent as analytical expression. The experimental data processing algorithm and the example of using this algorithm for spectrographic analysis of oncologic preparations of blood is represented in this article.

Ключевые слова: график, экспериментальные данные, кривые Безье

Введение

Очень часто экспериментальные данные являются искаженной помехами реализацией определенного процесса, вполне определяемого некоторой неизвестной функцией вида $y = f(x)$. В большинстве случаев экспериментальные данные такого рода представляют в виде графиков. Графики являются, видимо, самым простым и давно используемым в самых различных областях человеческой деятельности средством когнитивного представления экспериментальных данных, которые позволяют наглядно оценить качественные особенности процесса, несмотря на помехи, погрешности измерений. Графики, отображающие один и тот же процесс, характеристику некоего объекта могут существенно отличаться друг от друга масштабами, количеством использованных измерений, уровнем помех, и пр. В то же время особенности формы кривой графика характеризуют параметры отображаемого объекта или процесса.

Автоматическая (автоматизированная) обработка таких графиков предполагает сопоставление их форм с целью идентификации, чтобы установить, характеризуют ли различные графики одни и те же или различные процессы или объекты. Решение этой задачи средствами нейросетевого или статистического распознавания зрительных образов осложняется тем, что каждый график может отличаться от других масштабами, уровнем помех, количеством измерений. В то же время известен метод [1] нейросетевого распознавания контуров изображений, представленных кривыми Безье. Именно использование кривых Безье позволило обеспечить инвариантность описаний контуров изображений относительно положения изображений в поле зрения, масштаба, поворота. В таком случае сами исходные экспериментальные данные – графики могут быть заменены другими, аналитическими кривыми, хорошо совпадающими с формой кривой исходного графика, и, в то же время, инвариантными относительно изменения масштаба, количества измерений, уровня помех. Такие кривые могут быть заданы функциями $x(t)$, $y(t)$, которые представляют собой полиномы n -ной степени от параметра t :

$$\begin{aligned} x(t) &= a_0 t^n + a_1 t^{n-1} + \dots + a_{n-1} t + a_n, \\ y(t) &= b_0 t^n + b_1 t^{n-1} + \dots + b_{n-1} t + b_n. \end{aligned} \quad (1)$$

Выбор n – степени полиномов зависит от сложности аппроксимируемых кривых.

Реализация такого подхода позволила бы отобразить все существенные особенности формы графика, исключив влияние помех. Кроме того, вместо описаний графиков в пространстве сигналов можно рассматривать аппроксимирующие их полиномы в пространстве их коэффициентов. Существенным преимуществом такого описания является его инвариантность относительно количества измеренных сигналов в каждом графике.

Однако, если прямая задача получения графика по заданным коэффициентам полиномов является тривиальной, то обратная задача – определение коэффициентов аппроксимирующих полиномов по точкам графика – не имеет простого решения [2]. Одно из решений данной задачи рассматривается в настоящей работе.

Формальная постановка задачи

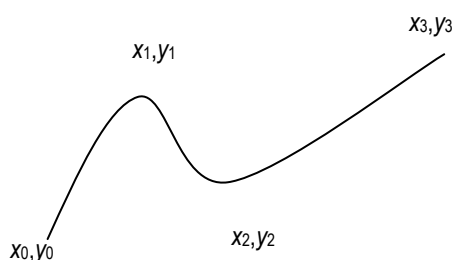


Рис. 1. Кривая Безье

График может быть представлен ломаной линией, точки излома которой образованы последовательностью N полученных экспериментально значений некоторой функции $y(x)$: $y_N(x) = \{y(x_0), y(x_1), y(x_2), \dots, y(x_N)\}$, при заданных значениях $x_0, x_1, x_2, \dots, x_N$. Рассмотрим кривую Безье в качестве аппроксимирующей кривой как наиболее часто используемый в практических приложениях, параметрически заданный полином третьей степени.

Под кривой Безье понимаем дугу плоской кривой третьего порядка, заданную в параметрическом виде таким образом [3]:

$$Bz(t) = \sum_{j=0}^3 Br_j(t) \cdot Q_j \quad (2)$$

где $Br_j(t)$ – базовые полиномы Бернштейна третьей степени

Q_j – коэффициенты кривой

t – действительный параметр, который изменяется в интервале $[0, 1]$.

Полиномы Бернштейна определяются следующими формулами:

$$Br_j(t) = C_3^j \cdot (1-t)^{3-j} t^j, \quad 0 \leq j \leq 3 \quad (3)$$

где C_3^j , $0 \leq j \leq 3$ – биномиальные коэффициенты

$$C_3^j = \frac{3!}{j!(3-j)!} \quad (4)$$

Таким образом, кривая Безье (рис. 1) может быть задана двумя полиномами третьей степени, зависящими от параметра t :

$$\begin{aligned} X(t) &= x_0 \cdot (1-t)^3 + x_1 \cdot (1-t)^2 \cdot t + x_2 \cdot (1-t) \cdot t^2 + x_3 \cdot t^3 \\ Y(t) &= y_0 \cdot (1-t)^3 + y_1 \cdot (1-t)^2 \cdot t + y_2 \cdot (1-t) \cdot t^2 + y_3 \cdot t^3 \end{aligned} \quad (5)$$

Представление экспериментальных данных в виде кривой Безье имеет ряд преимуществ. Описывая достаточно широкий класс кривых, кривые Безье используют относительно небольшое количество коэффициентов – 8. Важным преимуществом с точки зрения пользователя является наглядность. Коэффициенты (x_0, y_0) , (x_3, y_3) являются координатами исходной и конечной точек дуги аппроксимирующей кривой, соответственно, а коэффициенты (x_1, y_1) , (x_2, y_2) – координатами точек, которые управляют ее формой (так называемых "усов"). Изменяя коэффициенты или перемещая соответствующие им точки на экране монитора (как начальную и конечную, так и управляющие), пользователь может в интерактивном режиме редактировать форму аппроксимирующей кривой, улучшая результаты автоматической аппроксимации.

Таким образом, каждый график может быть представлен вектором $\mathbf{v} = \{x_0, y_0, x_1, y_1, x_2, y_2, x_3, y_3\}$, компонентами которого являются коэффициенты уравнений кривой Безье. Описание графика в виде вектора \mathbf{v} , всегда имеет одно и то же количество компонент независимо от числа измеренных экспериментально точек, что упрощает решение различных задач, в частности, задач распознавания графиков.

Как ранее было отмечено, определение коэффициентов кривой Безье, наиболее точно воспроизводящей тот или иной график, не является тривиальной задачей. В работе [2] для определения коэффициентов кривой Безье, аппроксимирующей последовательность экспериментальных точек, использован итерационный способ минимизации суммы квадратов расстояний от каждой точки графика до аппроксимирующей кривой. Такая постановка обусловлена тем, что рассматриваются произвольные графики, а не только те, которые отображают некоторую неизвестную функцию. Допускается, что график представляет собой совокупность несвязанных, взаимно независимых точек. Для вычисления расстояния от каждой точки графика до аппроксимирующей кривой должна быть найдена соответствующая ближайшая точка на аппроксимирующей кривой, что само по себе при параметрическом задании кривой является итерационной задачей. При значительном количестве точек на графике и заранее не определенном количестве итераций такой способ решения задачи привел к существенным вычислительным трудностям.

В данной работе предлагается принять в качестве меры сходства данного экспериментального графика с некоторой аппроксимирующей кривой площадь $S(\mathbf{v}, y(x))$ фигуры, ограниченной контуром, образованным упомянутыми кривой, определяемой коэффициентами \mathbf{v} , и экспериментальной кривой, образованной точками графика, отображающего некоторую неизвестную функцию $y(x)$. Тогда поиск кривой Безье с коэффициентами \mathbf{v}_{opt} , которая наилучшим образом аппроксимирует заданный экспериментальный

график, можно рассматривать как определение коэффициентов кривой Безье, которые минимизируют значение $S(\mathbf{v}, y_{\mathcal{D}}(x))$:

$$\mathbf{v}_{opt} = \underset{\mathbf{v}}{\operatorname{argmin}} S(\mathbf{v}, y_{\mathcal{D}}(x)) \quad (6)$$

Алгоритм

Как видно из формулы 6, основой алгоритма поиска оптимального значения коэффициентов \mathbf{v}_{opt} является вычисление площади, ограниченной контуром $S(\mathbf{v}, y_{\mathcal{D}}(x))$. Непосредственное вычисление $S(\mathbf{v}, y_{\mathcal{D}}(x))$ не представляется возможным из-за того, что данный контур не является односвязным.

График $y_{\mathcal{D}}(x)$ и кривая Безье с коэффициентами \mathbf{v} неоднократно пересекаются. Вычисление координат точек пересечения затруднено параметрическим заданием кривой Безье. Наиболее простым решением представляется замена кривой Безье ломаной линией V_t , точки излома которой (x_t, y_t) принадлежат кривой Безье и вычислены для значений $t = 0, \delta, 2\delta, 3\delta, \dots, 1$. Значение δ выбрано достаточно малым, чтобы обеспечить достаточно точное для практических вычислений отображение ломаной линией V_t исходной кривой Безье. Результатом принятого упрощения является возможность вычислить для каждого значения $y(x)$ соответствующее значение $V_t(n)$. Тогда площадь $S(\mathbf{v}, y_{\mathcal{D}}(x))$ может быть вычислена как

$$S(\mathbf{v}, y_{\mathcal{D}}(x)) = \sum_{n=0}^N |V_t(x_n) - y_{\mathcal{D}}(x_n)| \quad (7)$$

Значение \mathbf{v}_{opt} , соответствующее минимальному значению $S(\mathbf{v}, D(n))$, может быть получено путем использования метода градиентного спуска.

Экспериментальная проверка алгоритма

Предложенный алгоритм был экспериментально проверен более чем на 1100 реальных графиках спектрограмм медицинских препаратов крови, полученных по методу Онкотест [4]. Метод ранней диагностики онкологических заболеваний позволяет определить наличие или отсутствие онкологического заболевания у пациента по виду спектрограмм медицинских препаратов крови (рис.2).

В настоящее время проводятся исследования для определения, при условии наличия онкологического заболевания, локализации заболевания. На рис.3 показаны спектрограммы препарата крови больных онкологическим заболеванием одинаковой локализации (желудок).

Несмотря на очевидные различия спектрограмм по уровням сигналов, помех, формы кривых имеют несомненное сходство. Специфичным свойством приведенных графиков является форма соответствующих им кривых, которая может соответствовать диагнозу.

Решение этой задачи предполагает дополнительные исследования, которые должны основываться на обработке большого количества экспериментальных данных. В свою очередь, обработка больших массивов данных невозможна без решения задач их компьютерной обработки, в числе которых обязательно должны быть решены задачи автоматической или автоматизированной классификации спектрограмм по их форме с использованием методов распознавания образов.

Непосредственное использование статистических методов распознавания для автоматизированной постановки диагноза в данном случае не представляется возможным из-за большой изменчивости значений масштаба (десятки раз), и, также, конфигурации функции реализующей один и тот же образ. Поэтому представилось целесообразным для каждого графика найти некоторую подходящую кривую Безье, полиномиальные коэффициенты которой, после приведения к одному масштабу, могут быть использованы как параметры образа, инвариантные к изменениям масштаба, количества измерений, уровня помех. Примеры спектрограмм и аппроксимирующие их кривые Безье приведены на рис. 5,6.

Результат достигался в большинстве случаев за 15 - 40 итераций (в особенно сложных случаях - до 300). Как видно из рисунков, аппроксимирующие кривые Безье дают удовлетворительное приближение.

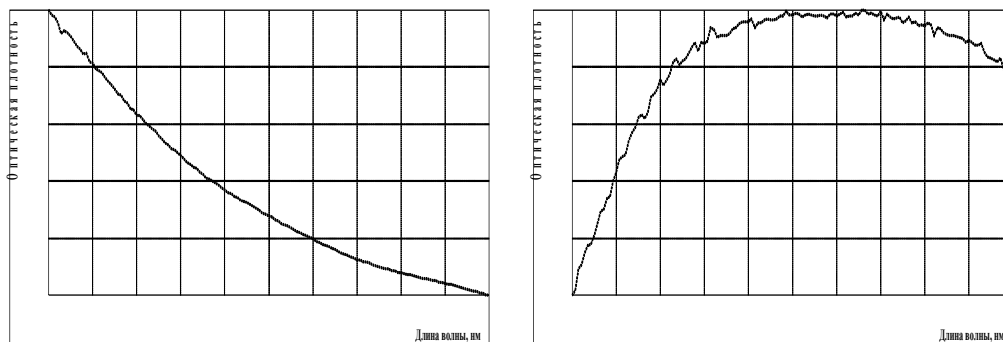


Рис.2. Примеры спектрограмм препарата сыворотки крови: слева – спектрограмма здорового, справа – больного человека.

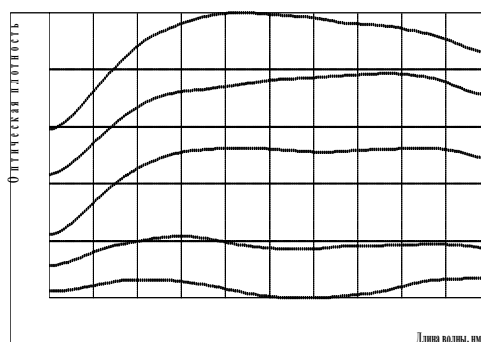


Рис.3 Примеры спектрограмм препарата крови больных онкологическим заболеванием одинаковой локализации (желудок).

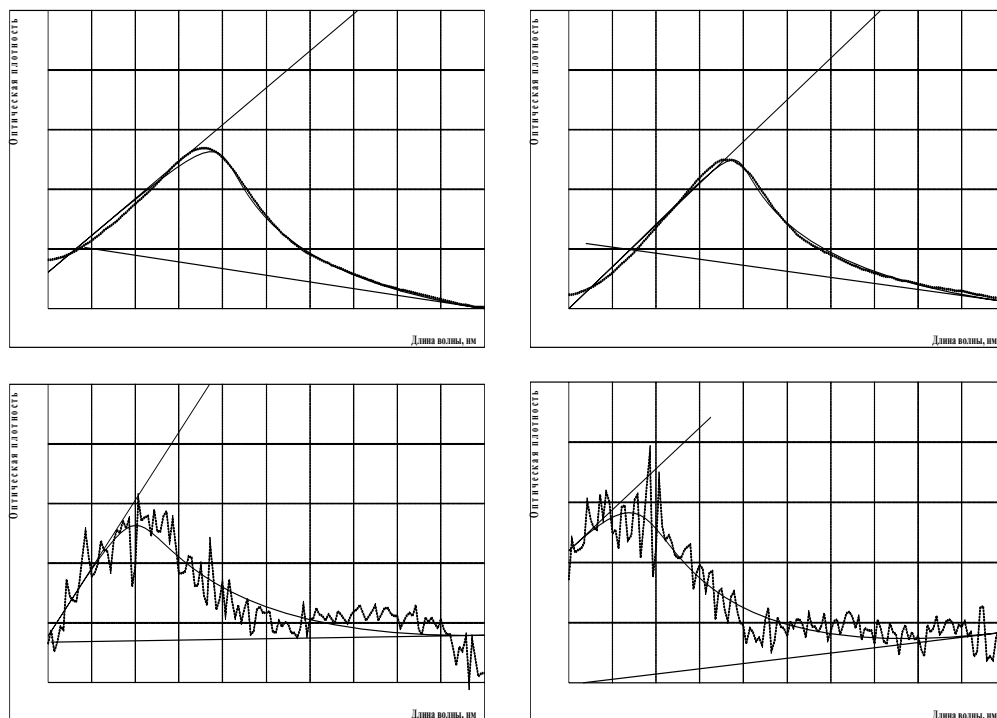


Рис.4 Спектрограммы препаратов крови и их аппроксимация кривыми Безье; локализация – печень

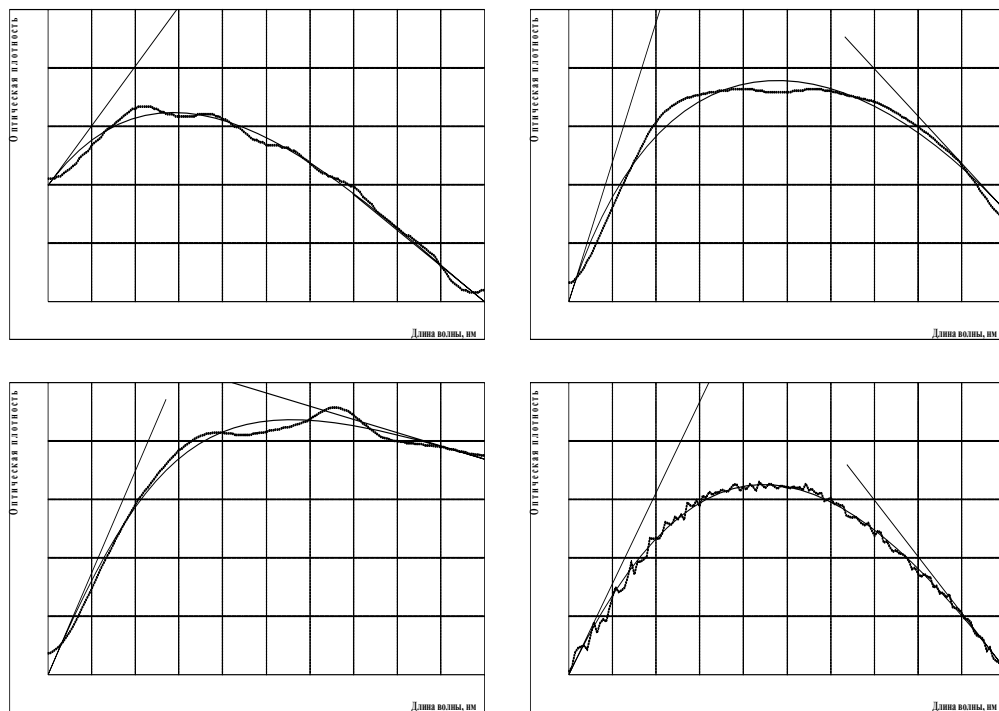


Рис.5 Спектрограммы препаратов крови и их аппроксимация кривыми Безье; локализация – молочная железа

Заключение

Предложенный алгоритм позволяет найти параметры кривой Безье, которая аппроксимирует экспериментальный график, с целью автоматической классификации. Приведены примеры аппроксимации спектрограмм с целью выполнения автоматизированной диагностики. Это даст в будущем возможность автоматизировать процесс обследования населения с целью раннего выявления злокачественных опухолей.

Литература

- [1] Кодирование контуров, представленных кривыми Безье, в задачах нейросетевой классификации / Куссиль М.Э. // Математические машины и системы. – 2004. – № 3. – С.17-30.
- [2] Вишневский В.В., Рысцов И.К., Волжева М.В Итерационный алгоритм построения кривой Безье по заданными точками //Математические машины и системы. - 2004 №4 - С. 108 – 116.
- [3] Денискин Ю.И. Особенности аппроксимации обводов параметрическими полиномами в форме Бернштейна //Прикладная геометрия. - 1999. - вып. 2 № 2
- [4] Вишневский В.В., Владимиров В.А., Романенко Т.Н. Программно-аппаратный комплекс "Онкотест-WM-01" // Тезисы доклада II Международной школы-семинара "Телемедицина: опыт и перспективы".

Информация об авторах

Виталий Вишневский – заведующий отделением, старший научный сотрудник, кандидат технических наук, e-mail: vit@immsp.kiev.ua

Владимир Калмыков – старший научный сотрудник, кандидат технических наук, e-mail: kvq@immsp.kiev.ua

Татьяна Романенко – младший научный сотрудник, e-mail: romanenko@immsp.kiev.ua

Институт проблем математических машин и систем, просп. акад. Глушкова 42, 03680, Киев 187, Украина.

D.2.2. Decision Tools and Techniques

APPLICATION OF CASE-BASED REASONING FOR INTELLIGENT DECISION SUPPORT SYSTEMS

Alexandr Eremeev, Pavel Varshavsky

Abstract: *The method of case-based reasoning for a solution of problems of real-time diagnostics and forecasting in intelligent decision support systems (IDSS) is considered. Special attention is drawn to case library structure for real-time IDSS (RT IDSS). This work was supported by RFBR.*

Keywords: *Intelligent decision support systems, analogous reasoning, case-based reasoning.*

ACM Classification Keywords: *H.4.2 [Information systems applications]: Types of systems – Decision support; I.2.6 [Artificial intelligence]: Learning – analogies.*

Introduction

The problem of human reasoning simulating (so called “common sense” reasoning) in artificial intelligence (AI) systems and especially in IDSS is very actual nowadays [1]. That is why special attention is turned to case-based and analogous reasoning methods and models. The analogy and precedents (cases) can be used in various applications of AI and for solving various problems [2-6], e.g., for diagnostics and forecasting or for machine learning. AI experts model case-based reasoning by computers in order to develop more flexible models of search for solutions and learning.

In this paper, we consider method of case-based reasoning for a solution of problems of real-time diagnostics and forecasting in RT IDSS [4]. These systems are usually characterized by strict constraints on the duration of the search for the solution. One should note that, when involving models of case-based and analogous reasoning in RT IDSS, it is necessary to take into account a number of the following requirements to systems of this kind [1]:

- The necessity of obtaining a solution under time constraints defined by real controlled process;
- The necessity of taking into account time in describing the problem situation and in the course of the search for a solution;
- The impossibility of obtaining all objective information related to a decision and, in accordance with this, the use of subjective expert information;
- Multiple variants of a search, the necessity to apply methods of plausible (fuzzy) search for solutions with active participation of a decision making person (DMP);
- Nondeterminism, the possibility of correction and introduction of additional information in the knowledge base of the system.

The methods of case-based reasoning may be applied in units of analysis of the problem situation, search for solutions, learning, adaptation and modification, modeling and forecasting. The use of the respective methods in IDSS broadens the possibilities of IDSS and increases the efficiency of making decisions in various problem (abnormal) situations.

Case-based reasoning

Case-based reasoning, like analogous reasoning, is based on analogy; however, there are certain differences in their implementation [4, 7]. In the most encyclopedias, a precedent (from Latin, *precedentis*) is defined as a case that took place earlier and is an example or justification for subsequent events of this kind. To create a precedent means to give grounds for similar cases in the future, and to establish a precedent is to find a similar case in the past.

As the practice shows, when a new problem situation arises, it is reasonable to use this method of case-based reasoning without drawing an analogy. This is caused by the fact that humans operate with these reasoning schemes at the first stages, when they encounter a new unknown problem.

Case-based reasoning is an approach that allows one to solve a new problem using or adapting a solution of a similar well-known problem. As a rule, case-based reasoning methods include four main stages that form a CBR-cycle, the structure of which is represented in Fig. 1 [8].

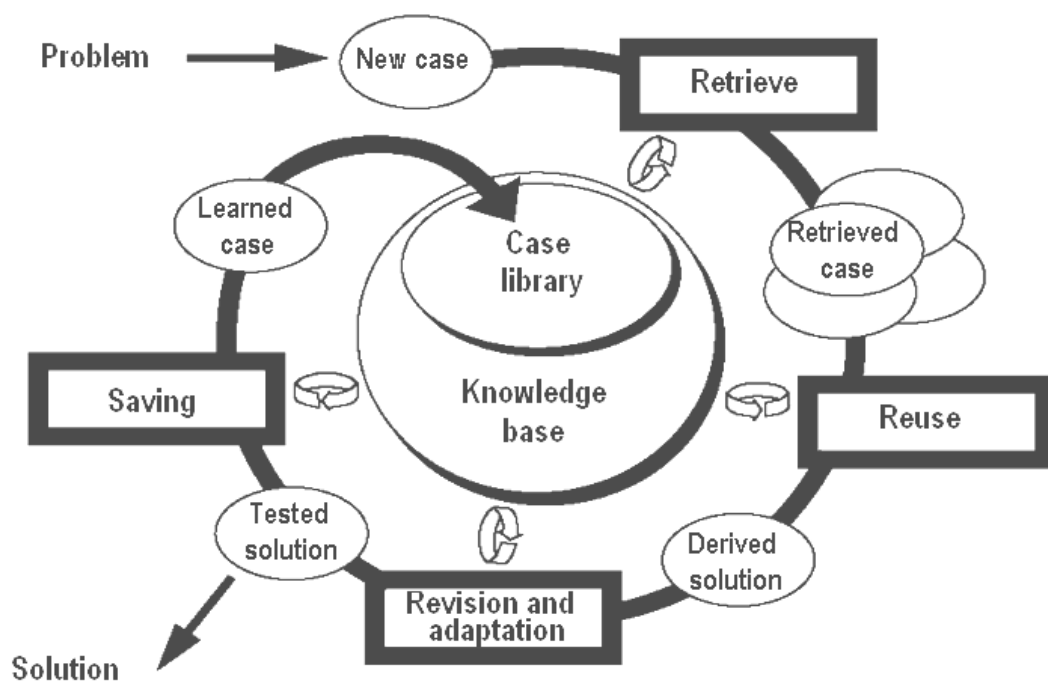


Fig. 1 The structure of CBR-cycle

The main stages are as follows:

- Retrieving the closest (most similar) case (or cases) for the situation from the case library;
- Using the retrieved case (precedent) for solving the current problem;
- If necessary, reconsidering and adaptation of the obtained result in accordance with the current problem;
- Saving the newly made solution as part of a new case.

It is necessary to take into account that a solution on the basis of cases may not attain the goal for the current situation, e.g., in the absence of a similar (analogous) case in the case library. This problem can be solved if one presupposes in the CBR-cycle the possibility to update the case library in the reasoning process (inference) [4, 7]. A more powerful (in detecting new facts or new information) method of reasoning by analogy is a means of updating case libraries. We also note that the elements of case-based reasoning may be used successfully in analogy-based reasoning methods; i.e., these methods successfully complement each other and their integration in IDSS is very promising.

Use of the mechanism of cases for RT IDSS consists in output of the decision to the operator (DMP) for the current situation on the basis of cases which is contained in system. As a rule, the last stage in a CBR-cycle is excluded and performed by the expert (DMP) because the case library should contain only reliable information confirmed by the expert. Reconsidering and adaptation of the taken decision is required seldom because the same object (subsystem) is considered.

The modified CBR-cycle for RT IDSS includes following stages:

- Retrieving the closest (most similar) case (or cases) for the situation from the case library;
- Using the retrieved case (precedent) for solving the current problem.

Case-based reasoning for IDSS consists in definition of similarity degree of the current situation with cases from case library. For definition of similarity degree, the nearest neighbor algorithm is used [4].

The nearest neighbor algorithm

The class of selection algorithms used by the most CBR products is called nearest neighbor. Let's explain the work of algorithm on a simple example. Consider an item that has two attributes: a **temperature** and a **liquid level**. Let us draw all the items at diagram. On the one (x-axis) there is the temperature, say, from **10 – 50°C**, while other (y-axis) contain a value of **liquid level** (for simplicity, let's make the range **1000 – 5000 mm**).

In the library, there are two cases:

- **C₁: t = 30°C, h = 3500 mm;**
- **C₂: t = 40°C, h = 1500 mm.**

For the current situation (**Target**): **t = 20°C, h = 3000 mm**.

If we plotted them on the chart, it might look like fig. 2.

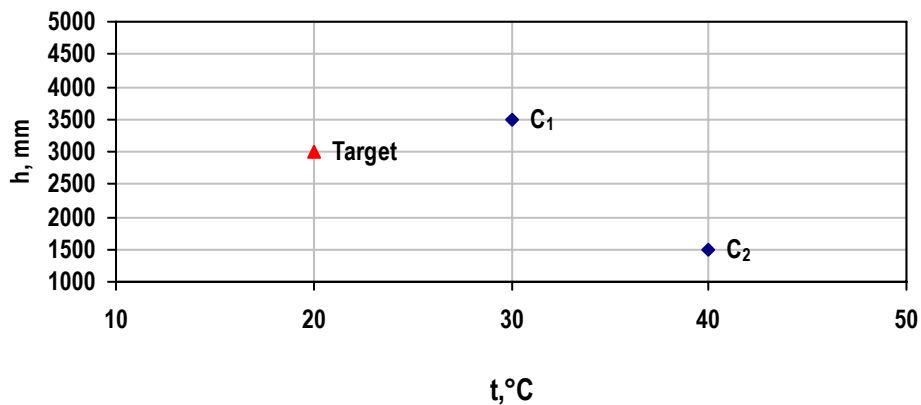


Fig. 2 Coordinate plane

Now we shall calculate the distance from the Target up to C₁ and C₂:

$$D_1 = \sqrt{(20 - 30)^2 + (3000 - 3500)^2} = 500,10;$$

$$D_2 = \sqrt{(20 - 40)^2 + (3000 - 1500)^2} = 1500,13.$$

The maximal distance D_{MAX} between points with coordinates (10,1000) and (50,5000) is similarly calculated ($D_{MAX} = 4000,20$).

Then values of similarity degree (SIM) of the current situation with two cases from case library are calculated:

- for C₁: $SIM_1 = (1 - D_1/D_{MAX}) = (1 - 500,10/4000,20) = 0,8750$ (87,50%);
- for C₂: $SIM_2 = (1 - D_2/D_{MAX}) = (1 - 1500,13/4000,20) = 0,6250$ (62,50%).

In case there are n ($n > 2$) parameters for the description of a situation and cases, a more complex variant is considered, and it is differed from the presented one only that n coordinates are used.

Further, we shall view the structure of case library for RT IDSS on the basis of nonclassical logics for monitoring and control of complex objects like power units.

The structure of case library for RT IDSS

The case library for RT IDSS should join in itself the cases concerning a particular subsystem of complex object, and also contain the information on each parameter which is used for the description of cases (parameter type and range).

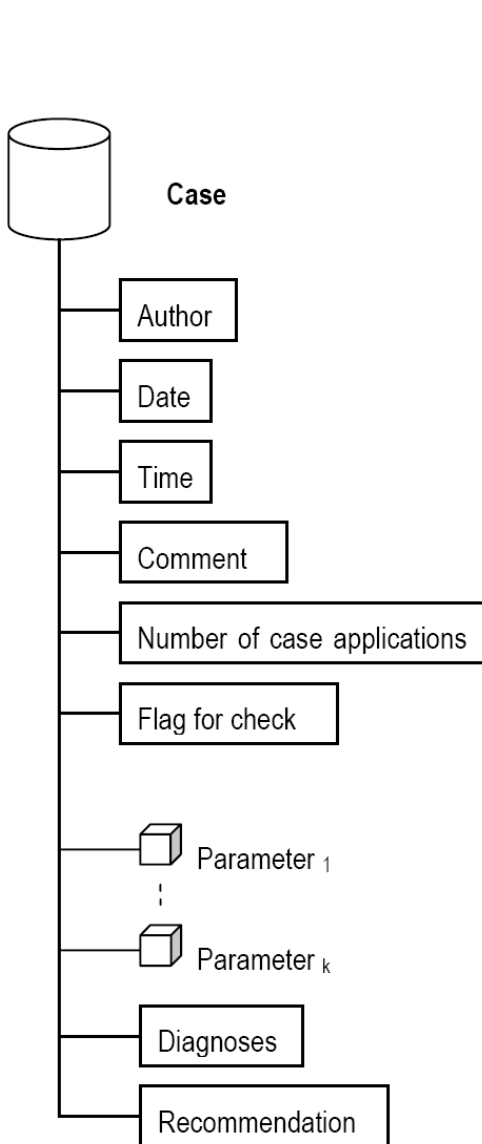


Fig. 3 Case structure

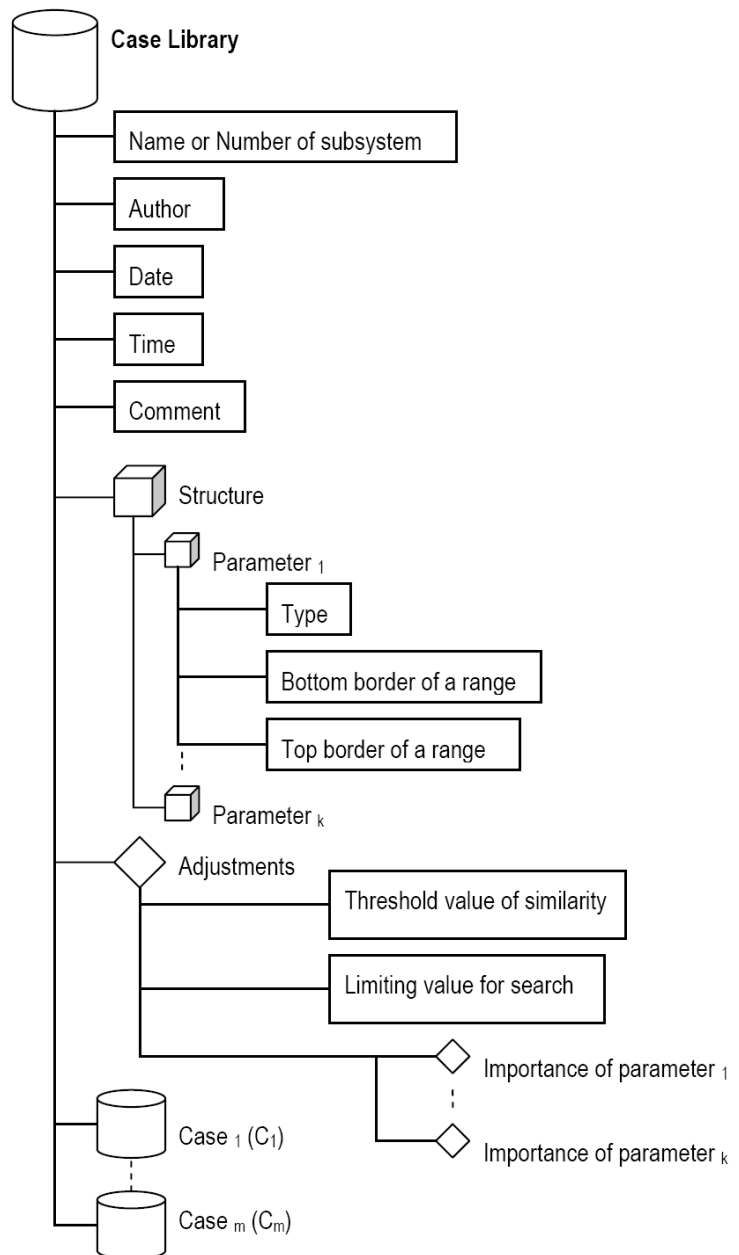


Fig. 4 The structure of case library for RT IDSS

Besides, the case library should include such adjustments, as:

- the significance of parameter;
- a threshold value of similarity;
- a value which limits quantity of considered cases.

It is necessary to emphasize, that the case library can be formed on the basis of:

- the experience, accumulated by the expert;
- analysis of the system archive;
- the analysis of emergencies;
- operative instructions;
- technological requirements.

The case library can be included in the structure of the knowledge base of RT IDSS or act as a separate component of the system.

Case structure is presented in figure 3, and the structure of case library for RT IDSS in figure 4.

Application of case-based reasoning for diagnostics of complex object states

As a complex object, we shall understand an object which has a complex architecture with various interrelations, with a lot of controllable and operated parameters and small time for acceptance of operating influences. As a rule, such complex objects as the power unit are subdivided into technological subsystems and can function in various modes (in regular, emergency, etc.).

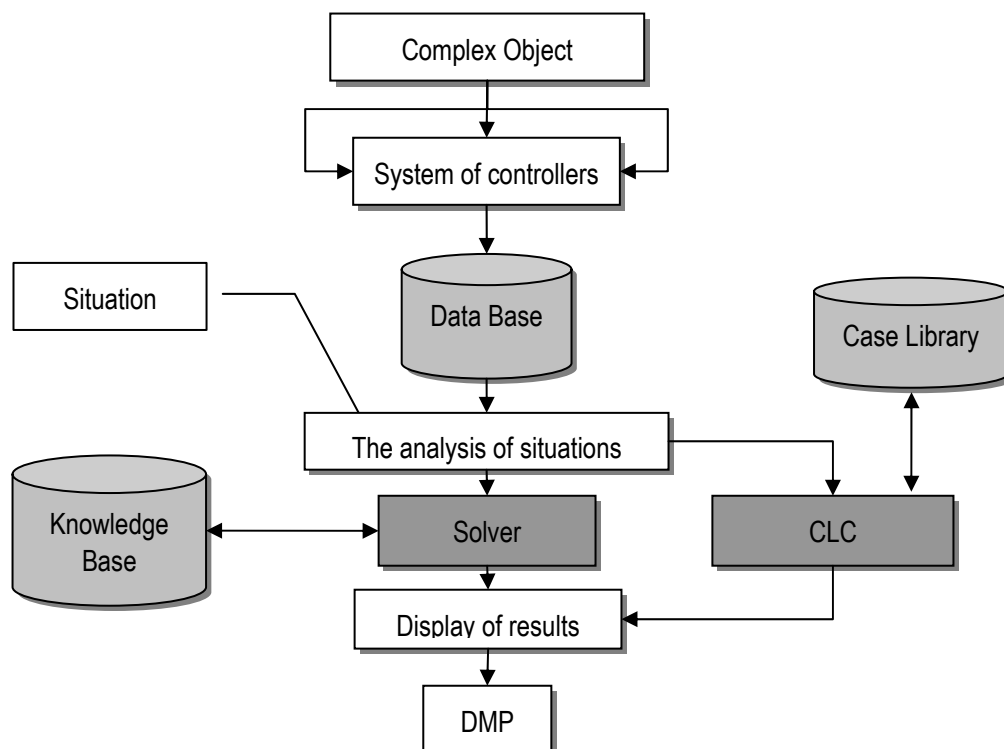


Fig. 5 The scheme of functioning for RT IDSS with use of CLC

For the description of such complex object and its subsystems, the set of parameters is used. The state of object is characterized by a set of concrete values of parameters.

In the operative mode reading of parameters values from sensors for all object is made by the system of controllers with an interval in 4 seconds. For this time interval, it is necessary to give out to the DMP (operator) the diagnosis and the recommendation on the developed situation.

Diagnosing and detection of operating influences is carried out on the basis of expert knowledge, technological requirements and operative instructions. The developed software (Case Libraries Constructor – CLC) can be applied to the decision of the specified problems.

Basic components of CLC are:

- module for storage and loadings case libraries and for data import;
- a subsystem of visualization for browsing the structure of case libraries;
- a subsystem of editing and adjustment of case libraries;
- a module of new cases check;
- a subsystem of case library testing and case-based reasoning.

CLC was implemented in Borland C++ Builder 6.0 for Windows NT/2000/XP and was registered in Russian Federal Service for intellectual property, patents and trademarks.

The implemented tool was applied in the prototype of a RT IDSS for monitoring and control of complex objects like power units on an example of a pressurizer in water-moderated reactor contour of the atomic power station (fig. 5).

Implementation of case libraries with use of CLC for systems of expert diagnosing is subdivided into the following main stages:

- Creation of case libraries for subsystems of complex object;
- Adjustment of the created case libraries;
- Addition of cases in case libraries;
- Check of the added cases;
- Testing of the filled case libraries with using case-based reasoning;
- Reservation of the created case libraries for their subsequent transfer to operative maintenance.

Conclusion

The method of case-based reasoning was considered from the aspect of its application in modern IDSS and RT IDSS, in particular, for a solution of problems of real-time diagnostics and forecasting.

The CBR-cycle is considered and its modification for application in RT IDSS is offered.

The nearest neighbor algorithm for definition of similarity degree of the current situation with cases from case library is described.

The structure of case library for RT IDSS is proposed.

The proposed method of case-based reasoning was implemented in Borland C++ Builder 6.0 for Windows NT/2000/XP. The implemented tool was registered in Russian Federal Service for intellectual property, patents and trademarks.

The main functional components of the implemented tool (Case Libraries Constructor – CLC) are specified.

The presented tool was applied in the prototype of a RT IDSS on the basis of nonclassical logics for monitoring and control of complex objects like power units.

The possibility of application of analogous reasoning in case-based reasoning is underlined. We also note that the elements of case-based reasoning may be used successfully in analogy-based reasoning methods; i.e., these methods successfully complement each other and their integration in IDSS is very promising [7].

Bibliography

1. Vagin V.N., Ereemeev A.P. Certain Basic Principles of Designing Real-Time Intelligent Decision Systems // Journal of Computer and Systems Sciences International, v. 40(6), 2001, pp. 953-961.
2. Kolodner J.L. Improving human decision making through case-based decision aiding // AI Magazine, 12(2), 1991, pp. 52–68.

3. Leake D.B. CBR in Context: The Present and Future // Case-Based Reasoning: Experiences, Lessons and Future Directions, AAAI Press / The MIT Press, 1996, pp. 3–31.
4. P.R. Varshavskii, A.P. Ereemeev Analogy-Based Search for Solutions in Intelligent Systems of Decision Support. Integrated models and flexible calculations in artificial intelligence. Journal of Computer and Systems Sciences International, Vol. 44, No. 1, 2005, pp. 90–101.
5. Ereemeev, P. Varshavsky Analogous Reasoning for Intelligent Decision Support Systems. Proceedings of the XI-th International Conference "Knowledge-Dialogue-Solution" – Varna, vol. 1, 2005, pp. 272-279.
6. A.P. Ereemeev, P.R. Varshavsky Reasoning by structural analogy in intelligent decision support systems // Proceedings of the Seventh Joint Conference on Knowledge-Based Software Engineering JCKBSE'06. IOS Press, 2006, pp. 303-306.
7. Alexander Ereemeev, Pavel Varshavsky Analogous Reasoning and Case-Based Reasoning for Intelligent Decision Support Systems // International Journal INFORMATION Theories & Applications (ITHEA) 2006, Vol.13 № 4, pp. 316-324.
8. Aamodt, E. Plaza "Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches". AI Communications, No. 7, 1994.

Authors' Information

A.P. Ereemeev – Applied Mathematics Department of the Moscow Power Engineering Institute (Technical University), Krasnokazarmennaya str., 14, Moscow, 111250, Russia; e-mail: eremeev@apmsun.mpei.ac.ru

P.R. Varshavsky – Applied Mathematics Department of the Moscow Power Engineering Institute (Technical University), Krasnokazarmennaya str., 14, Moscow, 111250, Russia; e-mail: VarshavskyPR@mpei.ru

LIMIT BEHAVIOUR OF DYNAMIC RULE-BASED SYSTEMS

Gennady Osipov

Abstract: The paper suggests a classification of dynamic rule-based systems. For each class of systems, limit behaviour is studied. Systems with stabilizing limit states or stabilizing limit trajectories are identified, and such states and trajectories are found. The structure of the set of limit states and trajectories is investigated.

Keywords: Dynamic rule-based systems, set of attainable states, limit trajectories.

ACM Classification Keywords: I.2.8 Problem Solving, Control Methods, and Search

Introduction

Dynamic intelligent systems and dynamic knowledge bases are typically understood as a result of integrating expert systems with simulation systems and automatically updated knowledge bases [1].

Although research in this area has a rather long history, there are some issues that still remain unaddressed — those of global behaviour of dynamic intelligent systems, their attainable sets, stability, and other issues that usually come up when studying dynamic systems [2].

In this paper, we focus on the kind of systems that are dynamic systems whose states, behavioural laws and other dynamic system "attributes" are described in a special way.

This special way consists in using intelligent system techniques for describing both states and behavioural laws for such systems (by intelligent system techniques we mean methods for knowledge representation, modelling of reasoning and behaviour modelling that are common in artificial intelligence). The said does not mean that functions and variables defined in any other way cannot be used as components of such systems; what is more

even, it is supposed that the systems in question allow for integration with various models, such as differential equation systems, finite automata, and others.

In a general case, attainability of knowledge-based systems is determined by their knowledge bases and control strategies [3, 4]. In the case when a set of rules is used as knowledge representation in a system [5], attainability is entirely determined by the structure of the set of rules, by the general principles of rule organization, and the control strategy being used [6].

If by a rule base architecture we mean 'structure of the rule set + rule structure + rule application logic' then, consequently, attainability of rule-based systems is entirely determined by the rule base architecture.

Let us remind the basic definitions following [4].

1 Rule-Based Systems

A rule [6] is said to be a triple of sets:

$\Pi = \langle C, A, D \rangle$, where:

C is the applicability condition for the rule;

A is the set of facts to be added by the rule Π ;

D is the set of facts to be removed by the rule Π .

C, A and D are sets of formulas of a language L, e.g. a multi-sorted first-order predicate calculus language, whose alphabet contains variables of sort t that take values from a linearly-ordered discrete set T. $A \cap D = \emptyset$ for every rule.

The word "fact" is used here as a synonym for the expression "closed atomic formula of a first-order predicate calculus language".

Formulas from C, A and D are turned into facts by some substitutions that will be described below.

Every rule will be assigned to one of two classes τ or θ and denoted as a τ -rule or θ -rule, respectively.

With each τ -rule, we associate either an action which is performed by an actuator in the environment or a procedure that computes and assigns to a variable the values of certain database attributes based on values taken by other attributes in the current state.

No actions are associated with θ -rules, as the latter do not affect the real world and merely update our knowledge of it.

It should be stressed that in conditions C of θ -rules first-order language formulas are such that the value of t ($t \in T$) in the formulas from the condition is the same as the value of t in the formulas from the sets A and D. This means that the result of a θ -rule execution changes the state in which its condition is satisfied.

As far as τ -rules are concerned, if $(\forall t) (n \leq t \leq m) (\exists x) F_C(t, x)$ is a formula from condition C and $(\forall t) (p \leq t \leq q) (\exists y) P_A(t, y)$ is a formula from the list A of formulas to be added (where n, p and m, q are the discrete start and end time points, respectively, for the "validity" term of facts $F_C(t, x)$ and $P_A(t, y)$), then the integer $v = p - n$, which is the time lag between the start point of the fact $P_A(t, y)$ being added and the start point of the period when the condition $F_C(t, x)$ is true, is a characteristic of each rule and is associated with it.

Things are the same with the sets of formulas to be removed.

Let us now look at the basic computational process in rule-based systems.

For this purpose, we need the following concepts to be introduced [3]: database and strategy of control over the system's rules.

1.1 Database

Database is a collection of finite relations, or tables (e.g. like those in relational databases), the number of which equals the number of different predicate symbols in the rules. Table columns correspond to the sorts of individual variables in atomic formulas. Interpretation of language L in the database is taken to be defined in a standard way.

One can therefore talk of satisfiability or non-satisfiability of rules' conditions.

1.2 Control Strategy

Control strategy picks up a rule from the set of rules, checks if its condition is satisfied in the current state of the working memory and, if so, applies the rule, i.e. performs the actions as prescribed by the rule; otherwise, it picks up the next rule and carries out the same manipulations on it.

For the sake of definiteness, we assume that the set of rules is ordered, e.g. in a lexicographic way.

Then the control strategy looks as follows:

1. Pick up the next rule Π_i from the set of rules.
2. Check whether condition C_i is true in the current state of the working memory.
3. If C_i is true, then substitute all free variables in formulas from C_i , A_i and D_i by the corresponding values from the database. Otherwise go to 1.
4. Apply the rule, i.e. write down to the working memory the values that make true the formulas from A_i and remove from the working memory the values that make true the formulas from D_i .
5. Go to 1.

The condition for the completion of the process is either stabilization of the working memory or exhaustion of the set of applicable rules.

Typically, the choice of rule depends on the task or domain specifics; the general principle consists in that the rule's condition should hold. If there is more than one such rule in a current state then the so-called conflict set resolution strategies are applied. With the latter not being the subject of this paper, we take the control strategy to be such that the choice of rule will only affect the computational complexity and not the result of the process. To put it differently, in what follows we are not going to be concerned with rule applicability, and we will get back to this later.

2 Dynamic Rule-Based Systems

Let X be a set of facts, $\chi \in 2^X$, $\Pi \in (\tau \cup \theta)$. Let $K(\chi, \Pi)$ denote the control strategy we described above, and we assume

$$K(\chi, \Pi^\theta) = \Phi(\chi), \text{ where } \Pi^\theta \in \theta,$$

$$K(\chi, \Pi^\tau) = \Psi(\chi), \text{ where } \Pi^\tau \in \tau.$$

$\Phi(\chi)$ will be referred to as a closure function, $\Psi(\chi)$ as a transition function.

Then

$$H = \langle X, T, \Phi, \Psi \rangle \tag{1}$$

will be referred to as a dynamic rule-based system.

The fixed point of equation

$$\Phi(\chi) = \chi$$

will be referred to as a state of the system (1), and the fixed point of equation

$$\Psi(\Phi(\chi)) = \chi \text{ (if such exists) with } t \rightarrow \infty,$$

will be referred to as the limit state of the system (1).

3 Classification of Rule-Based Dynamic Systems

As the basis for classification we will take the form of system's rules and certain correlations on the sets of rule components.

First, we identify classes of systems which differ in the form of rules.

In system H1 the rules are of the form:

$$\Pi_1 = \langle C, \{P(t,y)\}, \emptyset \rangle$$

(here $P(t, y)$ is a fact to be added).

In system H2 the rules are of the form:

$$\Pi_1 = \langle C, \{P(t,y)\}, \{\Phi(t,z)\} \rangle$$

(here $P(t, y)$ is a fact to be added, $\Phi(t, z)$ is a fact to be removed).

In system H3 the rules are of the form:

$$\Pi_1 = \langle C, P(t,y), F(t,z) \rangle$$

(here $P(t, y)$ is a set of facts to be added, $F(t,z)$ is a set of facts to be removed).

Let us now identify classes of systems, based on some correlations on the sets of rule components. Let S_0 be the initial state.

Then system H21 is a system H2, such that:

$(\cup\{P\}) \cap (\cup\{\Phi\}) = \emptyset$ (where $(\cup\{P\})$ and $(\cup\{\Phi\})$ is the union of facts being added and removed, respectively, over all of the rules of system H2);

system H22 is a system H2, such that $S_0 \cap (\cup\{\Phi\}) = \emptyset$;

system H23 is a system H2, such that $(\cup\{P\}) \cap (\cup\{\Phi\}) \neq \emptyset$ and $S_0 \cap (\cup\{\Phi\}) \neq \emptyset$;

system H31 is a system H3, such that $(\cup P) \cap (\cup F) = \emptyset$;

system H32 is a system H3, such that $S_0 \cap (\cup F) = \emptyset$;

system H33 is a system H3, such that $(\cup P) \cap (\cup F) \neq \emptyset$ and $S_0 \cap (\cup F) \neq \emptyset$ (here $(\cup P)$ and $(\cup F)$ stand for the union of the sets of facts being added and removed, respectively, over the entire set of the rules of system H3).

4 Limit States of Dynamic Rule-Based Systems

Let us give a few rather simple statements without proof:

Statement 1. The limit state of system H1 equals $S_0 \cup (\cup\{P\})$.

Statement 2. The limit state of system H21 equals $(S_0 / (\cup\{\Phi\})) \cup (\cup\{P\})$.

Statement 3. The limit state of system H31 equals $(S_0 / (\cup F)) \cup (\cup P)$.

Statement 4. In systems H22, H23, H32, H33 stabilization of states never occurs, as a matter of fact, but every state of every one of these systems lies in the set $S_0 \cup (\cup\{P\})$ — for systems H22 and H23 or the set $S_0 \cup (\cup P)$ — for systems H32 and H33.

Statement 5. The trajectories of systems H22, H23, H32 or H33, with t large enough, look as shown by diagram 1, where for H22 and H32 for all i : $S_i \subseteq S_{i+1}$.

It is appropriate to call such trajectories *limit trajectories*.

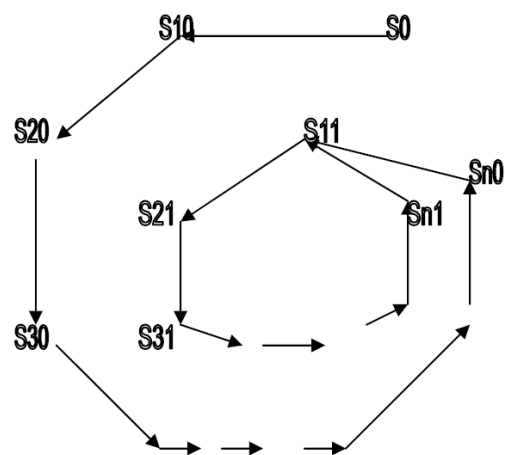


Diagram 1

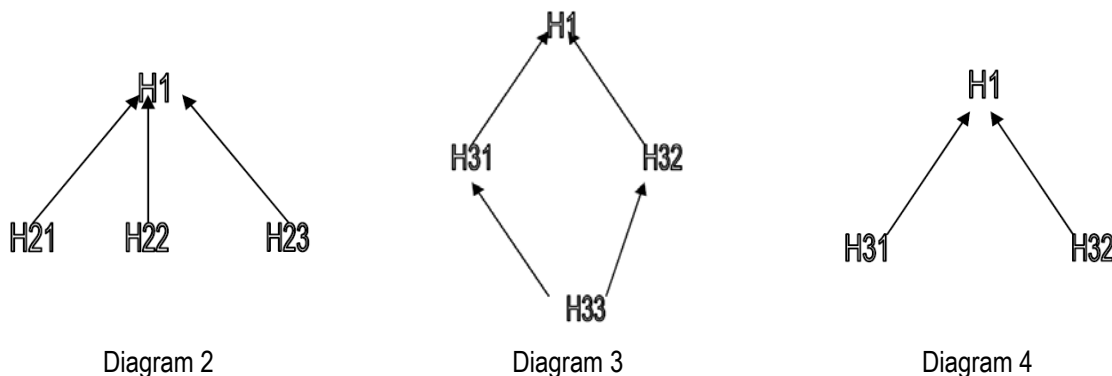
5. Structure of the Set of Limit States

Let us represent the structure of the set of limit states as set inclusion diagrams. The arrow pointing from a smaller set to a larger one plays the role of the inclusion relation. We also assume that all facts to be added and removed of all the systems under consideration belong to set X. Then it is evident enough that the following diagram 2 holds for systems H1, H21 and H22.

Let now systems H31, H32 and H33 be such that to each rule of system H31 a rule of system H32 is related in such a way that for each rule $\Pi(H31)$ of system H31 there is such a rule $\Pi(H32)$ in system H32 that $C(H31) = C(H32)$, $P(H31) = P(H32)$, and the inverse holds true (where C and P are the applicability conditions and the sets of facts to be added of systems H31 and H32, respectively). Then, if system H33 is such that for each rule $\Pi(H33)$ of system H33 there is a rule $\Pi(H31)$ in system H31 and there is a rule $\Pi(H32)$ in system H32 such that $C(H31) = C(H32) = C(H33)$, $P(H31) = P(H32) = P(H33)$ and $F(H31) \subseteq F(H33)$ and $F(H32) \subseteq F(H33)$, then the following diagram 3 holds.

In the last two diagrams, the inclusions of H22, H23, H32 and H33 in H1 have a slightly different meaning from others: they mean inclusion in H1 of every state of the limit trajectory.

Now let us come back to the postponed question of rule applicability. The situation is as follows: taking rule applicability into account may lead to some rules proving inapplicable on a certain step. It can be shown that if one sticks to the control strategy described in Section 1.2 then diagram 2 will remain the same. Diagram 3 will change its appearance to that of diagram 4.



Conclusion

In the paper, classes of dynamic systems have been identified from the point of view of their architecture. It has been shown that it is precisely the architectural specifics of such systems that determine their behaviour. The classes of systems with stabilizing limit states have been specified, and these states have been found. For systems with no stable limit states, the limit trajectories have been found (in case of a finite rule set).

The structures of limit states and trajectories have been established based on the criterion of attainable set inclusion.

Bibliography

- [1] V.L.Stefanuk. Dynamic Expert Systems. The International Journal of Systems & Cybernetics. V. 29, Issue 5, pp. 702-709, MCB University Press, 2000.
- [2] M.Mesarovich, Ya.Takakhara. General Theory of Systems: Mathematical Foundations. Moscow: Mir, 1978 (in Russian).
- [3] G.S.Osipov. Dynamics in Integrated Knowledge-Based Systems. Proc. of the 1998 IEEE International Symposium on Intelligent Control (September 14-17,1998). Gaithersburg, Maryland, USA, 1998, pp.199-2003.
- [4] T.G.Lebedeva, G.S.Osipov. Architecture and Controllability of Knowledge-Based Discrete Dynamical Systems. Journal of Computer and System Sciences International. NY, Vol. 39, No.5, 2000.

[5] N.Nilson. Principles of Artificial Intelligence. Moscow: Radio i svyaz', 1985.

[6] A.N.Vinogradov, L.Yu.Zhilyakova, and G.S.Osipov. Dynamic Intelligent Systems: II. Computer Simulation of Task-Oriented Behavior. Journal of Computer and System Sciences International. NY, Vol. 42, No. 1, 2003

Authors' Information

Gennady Osipov – Institute of Systems Analysis, RAS, 9 Prospekt 60-letiya Oktyabrya, Moscow 117312; e-mail: gos@isa.ru

МНОГОКРИТЕРИАЛЬНАЯ ОЦЕНКА И ОПТИМИЗАЦИЯ ИЕРАРХИЧЕСКИХ СИСТЕМ

Альберт Воронин

Аннотация. Показано, что любая многокритериальная задача может быть представлена иерархической системой, на нижнем уровне которой осуществляется оценка объекта по отдельным свойствам при помощи вектора критериев, а на верхнем уровне посредством механизма композиции получается оценка объекта в целом. Предложен метод решения сложных многокритериальных задач оценки и оптимизации, основанный на подходе вложенных скалярных сверток векторных критериев. Метод позволяет достаточно просто решать задачи структурного и параметрического синтеза многокритериальных иерархических систем.

Ключевые слова: выбор альтернатив, многокритериальность, иерархические системы, композиция критериев, вложенные скалярные свертки.

Введение

В теории принятия решений [1,2] различают два подхода к оценке объектов (альтернатив), подлежащих выбору. Один из них – оценка объекта в целом и выбор альтернативы по непосредственному сравнению объектов как гештальтов (гештальт: целостный образ объекта без детализации свойств). Второй – детализация и оценка тех или иных векторов характеристик (свойств) объектов и принятие решений по результатам сравнения этих свойств. Схема задачи принятия решений может быть представлена формулой [3]

$$\{\{\chi\}, \Phi\} \rightarrow \chi^*,$$

где $\{\chi\}$ – множество альтернатив (объектов); Φ – функция выбора, правило, по которому устанавливается предпочтительность в множестве альтернатив; χ^* – выбранные альтернативы (одна или более).

Целостный подход предусматривает выбор χ^* непосредственно по функции выбора Φ . Механизм векторного подхода требует осуществить декомпозицию (разложение) функции Φ на совокупность (вектор) функций выбора φ . Под декомпозицией функции выбора Φ понимается [1] ее эквивалентное представление (в смысле результатов выбора на всех альтернативах $\{\chi\}$) с помощью определенной совокупности других функций выбора φ , композицией которых является исходная функция выбора Φ .

Оба подхода имеют свои преимущества и недостатки. Выбор по непосредственному сравнению объектов может существенно отличаться от выбора, основанного на сравнении векторов-характеристик объектов

[2]. Это объясняется тем, что, как бы тщательно ни выбирались характеристики объектов, информация о векторах иногда дает недостаточно адекватное описание объектов выбора. Некоторая часть информации об объектах теряется при описании их набором характеристик. Другая часть информации, не имеющая непосредственного отношения к сравниваемым объектам, привносится при этом в модель. Кроме того, сам по себе выбор подходящей системы свойств (характеристик) объекта в известной мере субъективен. К тому же существует предположение [2], что человеческое мышление не приспособлено эволюцией к естественному с формальной точки зрения переходу от предпочтений на множестве объектов к предпочтениям на множестве наборов их характеристик.

Тем не менее, современная тенденция в теории принятия решений состоит в использовании именно векторного подхода. Это объясняется его объективностью и всесторонностью, а также принципиальной возможностью применения формализованных методов. Учитывается также конкретность и четкость подхода, так как по узкому вопросу меньше расхождений во мнениях, легче собрать бесспорные факты [4].

Предполагается, что в отношении отдельного свойства существенно проще сказать, какая из альтернатив предпочтительней для лица, принимающего решение (ЛПР). Так, в задаче выбора наилучшего проекта самолета гораздо уверенней можно говорить о том, что проект А лучше проекта В *по свойству* комфортности, или надежности, или грузоподъемности, нежели о том, что проект А лучше проекта В *в целом* [3]. Выделение свойств альтернатив является декомпозицией, приводящей к иерархической структуре свойств. Свойства первого иерархического уровня могут делиться на следующие наборы свойств и т.д. Глубина деления определяется стремлением дойти до тех свойств, которые удобно сравнивать друг с другом. Действительно, в примере с самолетом судить о комфортности, конечно, легче, чем о самолете в целом, но такое качественное свойство для сравнения также не совсем удобно и требует дальнейшей декомпозиции для удобства и объективности сопоставления свойств. Поэтому свойство комфортности, в свою очередь, подвергается декомпозиции на: а) уровень шумности в салоне, б) уровень вибрации пола, в) расстояние между креслами и др. Эти характеристики выражаются в числах и объективны.

Свойства, для которых существуют объективные численные характеристики, принято называть критериями. Более строго: **критериями** называются количественные показатели свойств объекта, числовые значения которых являются мерой качества объекта оценки по отношению к данному свойству. Получение набора критериев – конечный итог иерархической декомпозиции. Количество уровней зависит от требуемой глубины декомпозиции. Сложность заключается в том, что для каждого из начальных свойств глубина декомпозиции может быть различной, а на каждом уровне иерархии необходимо нормировать разнородные множества критериев.

Подход сравнения по отдельным свойствам, при всей своей привлекательности, порождает серьезную проблему обратного перехода к требуемому сравнению альтернатив в целом. Эта проблема предполагает решение задачи *композиции* критериев по уровням иерархии, что достаточно непросто, особенно при значительной глубине декомпозиции свойств. В простейшем и наиболее распространенном случае (двухуровневая иерархия) задача композиции решается традиционным получением однократной скалярной свертки критериев. Но уже при наличии трехуровневой иерархии требуются другие подходы.

Изложенное дает основание утверждать, что *любая многокритериальная задача может быть представлена иерархической системой*, на нижнем уровне которой осуществляется оценка объекта по отдельным свойствам при помощи вектора критериев, а на верхнем уровне посредством механизма композиции получается оценка объекта в целом. Центральной здесь является проблема композиции критериев по уровням иерархии.

Анализ состояния проблемы

В теории принятия решений в наибольшей мере разработан случай, когда многокритериальная задача представляется *двухурвневой* иерархической системой. Здесь задача композиции обычно решается либо методом главного критерия (критериальных ограничений), либо с помощью механизма однократной скалярной свертки векторного критерия [3]. Чаще используется расчет скалярной свертки, численная величина которой является оценкой качества данного объекта (альтернативы) в целом.

Скалярную свертку в традиционном виде удобно использовать в случаях, когда количество частных (парциальных) критериев s не слишком велико (в обычной практике $s \leq 10$). Тогда каждый из критериев играет самостоятельную роль, и все они сопоставимы по степени важности. Однако имеются сложные многокритериальные задачи, которые характеризуются большим числом (например, несколько десятков) частных критериев. В этих случаях значения отдельных критериев нивелируются и каждый из них слабо влияет на конечный результат решения многокритериальной задачи. Целесообразно объединять их в группы (кластеры, рубрики), в которых скалярные свертки рассматриваются как новые, более весомые частные критерии. Эти новые агрегированные критерии в свою очередь подвергаются операции скалярной свертки и уже более весомо сопоставляются между собой в процессе решения многокритериальной задачи. Таким образом, при увеличении размерности критериального пространства изначально двухурвневая иерархическая система критериев трансформируется в многоурвневую и требует разработки механизма композиции критериев по уровням иерархии.

В качестве примера приведем задачу альтернативной оценки научных космических проектов в области биологических исследований в космосе [5]. Для оценки эффективности таких проектов используются 28 частных критериев. В процессе консультаций с экспертами определены четыре группы (рубрики), в которых эти критерии объединяются: 1) общие критерии, 2) критерии научного развития, 3) экономические критерии, 4) социальные критерии (таблица 1).

Табл. 1. Оценка качества научного космического проекта «Биосорбент»

№ п/п	Критерии оценки качества проекта	Число баллов по 10-балльной шкале
I. Общие критерии		
1	Соответствие проекта заданиям Космической программы Украины	10,00
2	Степень интеграции проекта в международные программы биологических исследований в космосе	9,00
3	Вероятность того, что предложенный подход приведет к желаемым результатам	7,50
4	Полнота физической реализуемости проекта в заданных условиях космического эксперимента	8,30
II. Критерии научного развития		
5	Мера соответствия структуры работы и методов исследования целям проекта	9,75
6	Мера способствования проекта развитию знаний о влиянии факторов космического полета на фундаментальные физиологические процессы	8,25
7	Мера способствования проекта развитию знаний о механизмах адаптации биологических объектов к условиям космического полета	8,25
8	Степень новизны выполняемых исследований	8,00
9	Мера оригинальности и степень новаторства поставленных целей	8,75

10	Влияние результатов исследований на научные концепции и методы в области космической биологии и медицины	9,75
11	Вероятность того, что результаты исследований позволят осуществлять новые прорывные проекты	8,50
12	Мера опровержения существующих парадигм	5,30
13	Доля тех исследований в проекте, которые выполняются на мировом уровне	8,00
14	Мера способствования повышению международного престижа Украины	8,50
15	Степень поддержки проекта общественными организациями	9,70
16	Мера освещения исследований в научной литературе	9,70
17	Использование результатов работы в учебном процессе	7,70
18	Мера популяризации и распространения знаний	7,75
III. Экономические критерии		
19	Возможность передачи разработанных технологий для внедрения в экономику Украины	9,33
20	Мера обеспеченности коллектива достаточным количеством специалистов для доведения работ до практического внедрения	10,00
21	Способствование привлечению инвестиций	9,33
22	Уменьшение производственных затрат	9,33
23	Увеличение объема продажи продукции	8,67
24	Степень достаточности объема финансирования для выполнения запланированных работ	9,33
IV. Социальные критерии		
25	Увеличение количества рабочих мест	10,00
26	Повышение уровня квалификации персонала	8,00
27	Способствование развитию малого и среднего бизнеса	7,67
28	Влияние на деятельность общественных и молодежных организаций	10,00

В правой колонке таблицы приведены результаты обработки данных экспертизы одного из проектов данного направления, а именно научного космического проекта «Биосорбент», который был включен в программу экспериментов на борту Международной космической станции. Эти данные представляют собой нижний уровень трехуровневой иерархической системы критериев для оценки проекта в целом.

В монографии [6] рассмотрен общий случай разработки технологии поддержки принятия решений, когда требуется выбрать альтернативу из множества неоднородных альтернатив, для которых нельзя сформулировать единое множество количественных критериев оценки. В этом случае задача решается методами, основанными на иерархическом целевом оценивании альтернатив, без привлечения критериального анализа. Проблема композиции при наличии качественных свойств может быть решена методами бинарных отношений, например, методом анализа иерархий [7]. Но задача существенно облегчается, когда для оценки альтернатив можно привлечь количественные (или сводимые к ним) критерии, допускающие операции в нормализованном критериальном пространстве. Для таких задач применима теория многокритериальной оценки и оптимизации. В настоящей статье рассматривается именно такой класс задач исследования иерархических структур.

Постановка задачи

Состояние иерархической системы для данной альтернативы определяется параметрами:

$I = \{1, 2, \dots, n\}$ – множество элементарных подсистем, оцениваемых по критериям нижнего уровня иерархии;

$\{y_i\}_{i \in I}, y = \{y_1, \dots, y_n\}$ – оценки по скалярным критериям элементарных подсистем и векторный критерий нижнего уровня иерархии. Эффективность каждого высшего уровня зависит от значений оценок по критериям низшего уровня иерархии.

Дополнительные условия, определяющие собственно иерархическую структуру:

$J = \{1, 2, \dots, m\}$ – множество иерархических уровней;

$\{I_j\}_{j \in J}$ – распределение подсистем по уровням, $I_j = \{1, 2, \dots, n_j\}$;

$\{\lambda_j\}_{j \in J}$ – векторы приоритета.

Требуется найти аналитическую оценку φ^* и качественную оценку эффективности иерархической структуры, а из имеющихся альтернатив выбрать лучшую.

Метод решения

Предлагается для аналитической оценки эффективности иерархических структур применить метод вложенных скалярных сверток [8]. Композиция осуществляется по «принципу матрешки»: *скалярные свертки взвешенных компонент векторных критериев низшего уровня служат компонентами векторных критериев высшего уровня*. Скалярная свертка критериев, полученная на самом верхнем уровне, автоматически становится выражением для оценки эффективности всей иерархической системы в целом.

Формализм решения задачи методом вложенных скалярных сверток представляется последовательностью операций взвешенной скалярной свертки векторных критериев каждого уровня иерархии с учетом векторов приоритета на основе выбранной схемы компромиссов

$$\{(\varphi^{(j-1)}, \lambda^{(j-1)}) \rightarrow \varphi^{(j)}\}_{j \in J}, \varphi^{(1)} \equiv y,$$

а поиск оценки эффективности всей иерархической системы в целом выражается задачей определения скалярной свертки верхнего уровня иерархии:

$$\varphi^* = \varphi^{(m)}.$$

В задаче выбора решений количество вариантов (альтернатив) составляет $n_a \geq 1$. Каждый вариант характеризуется своей иерархической структурой. При $n_a = 1$ поставленная задача трансформируется в задачу оценки данной иерархической структуры. Если $n_a > 1$, то каждая структура оценивается как данная и выбирается тот вариант, иерархическая структура которого получила наилучшую оценку. Поэтому при дискретной многокритериальной оптимизации в качестве базовой здесь рассматривается задача оценки *данной* иерархической структуры. Конечно, так поступать можно только в случае относительно небольшого числа альтернатив n_a , когда метод простого перебора не вызывает значительных вычислительных трудностей. При больших объемах множеств альтернатив следует применять другие методы оптимизации, например, изложенные в [9].

Схема компромиссов

В качестве базовой схемы компромиссов для метода вложенных скалярных сверток предлагается использовать нелинейную схему, описанную в [10]. Установлено, что без потери общности предпосылкой для ее применения является то, что все частные критерии подлежат минимизации и являются ограниченными:

$$y_i \leq A_i, A = \{A_i\}_{i=1}^n, i \in [1, n],$$

где A – вектор ограничений.

Простой содержательной моделью функции полезности ЛПР на нижнем уровне иерархии при минимизируемых критериях в соответствии с концепцией нелинейной схемы компромиссов является скалярная свертка

$$\varphi(\lambda, y) = \sum_{i=1}^n \lambda_i [A_i - y_i]^{-1}$$

или

$$\varphi(\lambda, y_0) = \sum_{i=1}^n \lambda_i [1 - y_{0i}]^{-1},$$

если количественные критерии нормированы по формуле $y_0 = y/A$.

Качественные (но сводимые к количественным) критерии обычно определяются экспертами оценками по шкале баллов. Экспертам предъявляется анкета, в которой приводятся частные критерии. Критериям сопоставляется непрерывная шкала, разделенная, например, на десять интервалов. Цифра 0 на шкале соответствует понятию "никакой ценности", цифра 10 – "максимальная ценность". Эксперта просят оценить относительное влияние каждого из частных критериев на общую оценку в заданных условиях и приписать ему соответствующую точку на шкале, характеризующую числом f . Разрешается выбирать точки между числами или приписывать несколько критериев одной точке на шкале.

Анализ процессов принятия решений показал, что при оценке объектов по 10-балльной шкале эксперты руководствуются градациями так называемой фундаментальной шкалы. Понятие фундаментальной шкалы описано в [7] и представлено таблицей 2.

Табл.2. Фундаментальная шкала

Категория качества	Фундаментальная шкала f	Обращенная нормированная фундаментальная шкала y_0, φ_0
Неприемлемое	0 – 3	1,0 – 0,7
Низкое	3 – 5	0,7 – 0,5
Удовлетворительное	5 – 6	0,5 – 0,4
Хорошее	6 – 8	0,4 – 0,2
Высокое	8 - 10	0,2 – 0,0

Здесь показана связь между качественными градациями качества объектов и соответствующими количественными оценками f . Можно сказать, что в терминах теории нечетких множеств [11] фундаментальная шкала выступает как универсальная функция принадлежности для перехода от числа к соответствующей качественной градации и обратно. Осуществляется переход от лингвистической переменной (удовлетворительное качество, высокое качество и пр.) к соответствующим количественным

оценкам f по шкале баллов – (5,5; 7,0), т.е. переход от нечетких качественных градаций к числам и обратно.

Оценки f получаются по прямой 10-балльной фундаментальной шкале для *максимизируемых* критериев. Применяемая же в статье методика многокритериальной оценки по нелинейной схеме компромиссов разработана для *нормированных минимизируемых* критериев y_0 , оценки которых получаются из f по формуле [12]:

$$y_0 = 1 - 0,1 \cdot f, y_0 \in [0; 1],$$

что отражается обращенной нормированной фундаментальной шкалой в табл. 2. По этой же шкале измеряются и нормированные скалярные свертки критериев φ_0 .

Учет приоритетов

Областью определения коэффициентов приоритета $\lambda \in \Gamma_\lambda$ является симплекс

$$\Gamma_\lambda = \left\{ \lambda \left| \lambda_i \geq 0, \sum_{i=1}^n \lambda_i = 1 \right. \right\}. \quad (1)$$

Здесь $\lambda_i = \text{const}$ – формальные параметры, имеющие двоякий физический смысл. С одной стороны – это коэффициенты приоритета, выражающие предпочтения ЛПР по отдельным критериям. С другой – это коэффициенты содержательной регрессионной модели, построенной на основе концепции нелинейной схемы компромиссов. Определение коэффициентов λ на каждом уровне иерархии может быть выполнено путем оптимизации на симплексе с использованием дуального подхода, описанного в [10] или по формуле

$$\lambda_{ik}^{(j-1)} = \frac{f_{ik}}{n_k^{(j)} \sum_{i=1} f_{ik}}, k \in I_j,$$

где $\lambda_{ik}^{(j-1)}$ – i -я компонента вектора приоритета на $(j-1)$ -м уровне иерархии при расчете оценки эффективности k -й подсистемы j -го уровня; f_{ik} – параметр значимости i -й подсистемы $(j-1)$ -го уровня для k -й подсистемы j -го уровня (определяется экспертами по 10-балльной шкале); $n_k^{(j)}$ – количество подсистем $(j-1)$ -го уровня, поддерживающих k -ю подсистему j -го уровня.

В наиболее простом и достаточно распространенном случае формулируется и решается многокритериальная задача без приоритетов, когда ЛПР полагает, что все параметры значимости для всех подсистем *одинаковы*. В этом случае используется простейшая скалярная свертка по нелинейной схеме компромиссов в унифицированной форме [10].

При композиции критериев по уровням иерархии все скалярные свертки снизу доверху целесообразно вычислять на основе концепции нелинейной схемы компромиссов. При этом оценка эффективности k -й подсистемы на j -м уровне иерархии в виде нормированной вложенной скалярной свертки с учетом коэффициентов приоритета вычисляется по формуле

$$\varphi_{0k}^{(j)} = N_k^{(j)} \sum_{i=1}^{n_k^{(j)}} \lambda_{ik}^{(j-1)} [1 - \varphi_{0ik}^{(j-1)}]^{-1}, k \in I_j, \varphi_0^{(1)} \equiv y_0, \quad (2)$$

где $\varphi_{0ik}^{(j-1)}$ – оценка i -й компоненты нормированного векторного критерия на $(j-1)$ -м уровне иерархии при расчете оценки k -й подсистемы j -го уровня; $N_k^{(j)}$ – нормирующий множитель.

Условия нормировки

Для излагаемой теории важное значение имеет нормирование вложенных скалярных сверток на каждом уровне иерархии. В работах [5,8] рассмотрена возможность расчета условий нормировки исходя из принципа «солидарной ответственности» критериев. В множестве нормированных критериев выделяется значение наибольшего (наихудшего) из них. Считается, что если этот критерий достиг наихудшего значения, то и остальным нормированным критериям приписывается возможность достижения таких же значений, которые и составляют компоненты нормирующего вектора. Такой подход обладает достоинством простоты, но он работает только при условии действительно «равноправных» критериев.

Логично утверждать, что если оценки по всем парциальным относительным критериям $\varphi_{0ik}^{(j-1)}, i \in [1, n_k^{(j)}]$ будут одинаковыми и равными $\varphi_{0ik}^{(j-1)} \equiv \varphi_{0k}$, то их нормированная скалярная свертка в формуле (2) должна выражать такую же аналитическую и качественную оценку по обращенной нормированной фундаментальной шкале:

$$\varphi_{0k} = \frac{N_k^{(j)}}{(1 - \varphi_{0k})} \sum_{i=1}^{n_k^{(j)}} \lambda_{ik}^{(j-1)}.$$

Учитывая, что по условиям нормировки (1) $\sum_{i=1}^{n_k^{(j)}} \lambda_{ik}^{(j-1)} = 1$, выражение для нормирующего множителя приобретает вид

$$N_k^{(j)} = \varphi_{0k} (1 - \varphi_{0k}). \quad (3)$$

Проведем по формуле (3) калибровочные вычисления значений нормирующего множителя $N_k^{(j)}(\varphi_{0ik}^{(j-1)})$ для оценок $\varphi_{0ik}^{(j-1)}, i \in [1, n_k^{(j)}]$. Составим меру суммарной квадратичной погрешности, возникающей из-за того, что вместо точных значений нормирующего множителя в калибровочных точках $N_k^{(j)}(\varphi_{0ik}^{(j-1)})$ используется искомый множитель $N_k^{(j)}$:

$$M = \sum_{i=1}^{n_k^{(j)}} [N_k^{(j)} - N_k^{(j)}(\varphi_{0ik}^{(j-1)})]^2.$$

Используя необходимое условие экстремума функции

$$\frac{\partial M}{\partial N_k^{(j)}} = 0,$$

получим значение нормирующего множителя

$$N_k^{(j)} = \frac{1}{n_k^{(j)}} \sum_{i=1}^{n_k^{(j)}} N_k^{(j)}(\varphi_{0ik}^{(j-1)})$$

и с учетом (3)

$$N_k^{(j)} = \frac{1}{n_k^{(j)}} \sum_{i=1}^{n_k^{(j)}} \varphi_{0ik}^{(j-1)} (1 - \varphi_{0ik}^{(j-1)}).$$

Заключение

Рекуррентная формула (2) позволяет рассчитать количественные и получить качественные оценки нормированных по обращенной фундаментальной шкале скалярных сверток критериев по всем уровням иерархии вплоть до верхнего: $\varphi_0^* = \varphi_0^{(m)}$.

Решение оценочной многокритериальной задачи для примера, представленного таблицей 1, привело к следующим результатам. Скалярная свертка частных критериев нижнего уровня иерархии, относящихся к рубрике «Общие критерии», дала агрегированный критерий второго уровня иерархии $\varphi_{01}^{(2)} = 0,121$. Аналогичным образом, для группы «Критериев научного развития» получено значение агрегированного критерия $\varphi_{02}^{(2)} = 0,143$, для «Экономических критериев» $\varphi_{03}^{(2)} = 0,022$ и для «Социальных критериев» $\varphi_{04}^{(2)} = 0,100$.

Скалярная свертка указанных агрегированных критериев второго уровня иерархии позволила получить нормированную оценку в целом научного космического проекта «Биосорбент» в виде агрегированного критерия третьего уровня иерархии $\varphi_0^* = \varphi_0^{(3)} = 0,094$. Сопоставление с обращенной нормированной фундаментальной шкалой (табл.2) позволяет сделать вывод, что все агрегированные критерии находятся в пределах градации «Высокое качество». Отметим, что в расчетах данного примера предполагалась одинаковая значимость всех критериев, т.е. решалась многокритериальная задача без приоритетов.

Оценка данной и выбор лучшей из имеющихся альтернатив относятся к классу задач структурного синтеза. Задачи параметрического синтеза, решаемые методом вложенных скалярных сверток, описаны в работе [12].

Литература

1. Макаров И.М., Виноградская Т.М., Рубчинский А.А., Соколов В.Б. Теория выбора и принятия решений. – М.: Наука, 1982. – 328 с.
2. Юдин Д.Б. Вычислительные методы теории принятия решений. – М.: Наука, 1989. – 320 с.
3. Губанов В.А., Захаров В.В., Коваленко А.Н. Введение в системный анализ. – Л.: Изд-во ЛГУ, 1988. – 232с.
4. Ларичев О.И. Проблемы принятия коллективных решений в малых группах // Нелинейная динамика и управление (к 70-летию акад. С.В. Емельянова). – М.: Эдиториал УРСС. – 1999. – с. 91-103.
5. Воронин А.Н. Системный анализ и многокритериальная оценка космических проектов экспертными методами // Проблемы управления и информатики. – 2004. – № 1. – С. 121-135.
6. Тоценко В.Г. Методы и системы поддержки принятия решений. – К.: Наукова думка, 2002. – 382 с.
7. Saaty T.L. Multicriteria Decision Making: The Analytical Hierarchy Process. – N.Y.: McGraw-Hill, 1990. – 380 p.
8. Воронин А.Н. Вложенные скалярные свертки векторного критерия // Проблемы управления и информатики. – 2003. – № 5. – С. 10-21.
9. Воронин А.Н. Адаптивные аппроксимационные модели в задачах оптимизации // Кибернетика и системный анализ. – 1994. – № 5. – С. 83-93.
10. Воронин А.Н., Зиятдинов Ю.К., Козлов А.И. Векторная оптимизация динамических систем. – К.: Техніка, 1999. – 284 с.

-
11. Заде Л. Понятие лингвистической переменной и его применение к принятию приближенных решений. – М.: Мир, 1976. – 165 с.
 12. Воронин А.Н. Векторная оптимизация иерархических структур // Проблемы управления и информатики. – 2004. – № 6. – С. 26-34.
-

Сведения об авторе

Воронин Альберт Николаевич – профессор, доктор технических наук, профессор кафедры компьютерных информационных технологий Национального авиационного университета, проспект Комарова, 1, Киев-58, 03058 Украина.

СТАТИСТИЧЕСКИЙ АНАЛИЗ РАСПРЕДЕЛЕНИЙ ПОЛЕЗНОСТЕЙ

Виктор Семенов

Резюме: Рассматривается распределение полезностей доходов. Переход от анализа распределения доходов к исследованию распределения индивидуальных полезностей дает возможность достичь значительного уровня общности при анализе распределительной политики и сформировать подход к анализу институциональных и общественно-политических сдвигов и их импликаций на характер распределения общественных благ в рамках теории социального выбора.

Ключевые слова: Бокс–Кокс преобразование, функция полезности, индекс неравенства, индекс бедности.

Введение

Социально-экономические реформы в Украине насчитывают более 15 лет. За это время осуществлены существенные преобразования в политической, институциональной и социальной сферах. Появились разные социальные группы населения, которым присущи разные адаптационные возможности современных стратегий выживания и которые отличаются между собой показателями жизненного уровня и мотивационными характеристиками социального поведения. Социальные и экономические трансформации на постсоветском пространстве восстановили интерес научной общественности к исследованию неравенства доходов. Тем не менее, за незначительным исключением [1], эти исследования были в основном посвящены феноменологическому анализу этого явления. В предложенной работе рассматриваются обобщенные модели распределения доходов. Обобщение достигается путем перехода от анализа распределения доходов к исследованию распределения индивидуальных полезностей доходов. Переход от анализа распределения доходов к исследованию распределения индивидуальных полезностей дает возможность достичь значительного уровня общности при анализе распределительной политики и сформировать подход к анализу институциональных и общественно - политических сдвигов и их импликаций на характер распределения общественных благ в рамках теории социального выбора. В данной работе автор выходит из идеи, согласно которой в рамках теории социального выбора для обеспечения общественно-политического равновесия позиции медианного избирателя и избирателя со средним доходом в распределениях полезностей, полученных от реализации индивидуальных возможностей (в том числе доходов), должны совпадать. Таким свойствам удовлетворяет нормальная функция.

Постановка задачи. Некоторые социально-политические импликации

Предположим, что полезность доходов распределена нормально. Правомерность такого предположения вытекает из теории социального выбора [1]. Таким образом, для преобразования эмпирического распределения доходов используем функцию плотности нормального распределения:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \quad (1)$$

где μ - среднее значение, σ^2 - дисперсия. Преобразование, которое широко используется в разных областях, хотя и несколько в другом контексте, - это Бокс-Кокс-нормальное преобразование с постоянной маргинальной эластичностью [2,3], которому отвечает функция полезности

$$u = \frac{x^\alpha - \tilde{n}^\alpha}{\alpha}, \alpha < 1, \alpha \neq 0. \quad (2)$$

В предельном случае ($\alpha = 0$) получаем логнормальное распределение. При $\alpha < 1$ имеем: $u' > 0, u'' < 0, d \ln u / d \ln x = \alpha x^\alpha / (x^\alpha - c^\alpha)$. Т.е., это вогнутая функция с маргинальной полезностью α . Функция плотности имеет вид:

$$f_{\hat{A}\hat{E}} = \frac{x^{\alpha-1}}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}\left(\frac{x^\alpha - c^\alpha}{\alpha} - \mu\right)^2\right\}, \quad (3)$$

где μ и σ^2 - среднее значение и дисперсия полезностей, c - произвольная константа в единицах измерения x . В таком случае c может быть, например, минимальной заработной платой, не облагаемым налогами минимальным доходом, модальным или медианным доходом, конкретной долей медианного или среднего дохода (нормативно определенной чертой бедности). Если для заданного эмпирического распределения доходов мы выбираем функцию полезности $u = (x^\alpha - m^\alpha)/\alpha$, то, после соответствующего нормирования, получаем индексы неравенства из класса функций обобщенной энтропии (ОЭ) [4]:

$$I_\alpha = \frac{1}{\alpha(\alpha-1)} \int_0^\infty \left[\left(\frac{x}{m} \right)^\alpha - 1 \right] f(x) dx, \alpha \neq 0, 1 \quad (4)$$

с соответствующими предельными случаями при $\alpha = 0, 1$. При $\tilde{n} = z$, где z - черта бедности, на отрезке $[0, z]$ получаем для $0 < \alpha < 1$ индекс бедности, при $\alpha = 0$ имеем индекс бедности Ватса. Индексы ОЭ допускают аддитивную декомпозицию неравенства по демографическим подгруппам:

$$I_\alpha = \sum_{i=1}^k p_i \left(\frac{m_i}{m} \right)^\alpha I_{\alpha i} + I_{\alpha B}, \quad (5)$$

где $I_{\alpha i}$ - внутригрупповое неравенство группы i ; k - количество групп, $p_i = (n_i/n)$ - доля группы i , n_i - количество лиц в группе i , n - общее количество получателей доходов, $I_{\alpha B}$ - неравенство между группами, при вычислении которого предполагается, что каждый член группы получает одинаковый групповой доход. При $c = 0$ и $\alpha = 1 - \varepsilon$ получаем функцию полезности Аткинсона [1] с постоянной эластичностью по доходам:

$$u(x) = \frac{x^{1-\varepsilon}}{1-\varepsilon}, \quad \varepsilon > 0, \varepsilon \neq 1, u' > 0, u'' < 0. \quad (6)$$

Для $\varepsilon = 0$ получаем логарифмическую полезность: $u = \ln x$. Исходя из функции полезности вида (6), Аткинсон получил свой знаменитый нормативный индекс неравенства.

Рассмотрим некоторые импликации предложенных обобщенных моделей распределения доходов. В политологии чрезвычайно важную роль играет позиция медианного избирателя. В большинстве формальных моделей политического процесса система преимуществ медианного избирателя является доминирующей при принятии решений общественной значимости. Тем не менее, при условии сильной правосторонней скошенности распределения доходов (в общем случае благ, в том числе общественных), имеют место значительные расстояния между медианным избирателем и избирателем со средним уровнем доходов. Значительный разрыв между медианным избирателем и избирателем со средним уровнем доходов может приводить к политической нестабильности. Сближение их позиций по полезностям дает возможность избегать таких ситуаций. При $\alpha = 1$ ($\varepsilon = 0$) имеет место общество, в котором учитывается только агрегатный рост благосостояния, отсутствуют любые механизмы регулирования распределения благ, а общественная мысль абсолютно безразлична к неравенству. При увеличении ε возрастает антипатия общества к неравенству. Когда $\varepsilon \rightarrow \infty$, общество склонно предоставлять трансферы беднейшей части, игнорируя при этом их источники, а распределение доходов среди небедных не имеет никакого общественно-политического значения. Возрастание общественно-политического веса беднейших групп населения означает меньшую институциональную толерантность к неравенству и увеличение ε . При отсутствии эндогенных влияний поляризация общества ведет к уменьшению ε , а увеличение этого параметра обусловлено возрастанием политического и экономического влияния среднего класса на социальные процессы: мощный средний класс всегда поддержит прогрессивное перераспределение благ от богатых к бедным. Таким образом, можно сделать вывод, что изменение в значениях ε (α) может быть объяснено латентными факторами, которые отображают сдвиги в институциональном развитии, социально-политическом сознании, политико-экономической определенности и уровне экономической и политической культуры популяции (медианный избиратель и его окрестность должны сознавать свою доминантную роль в принятии общественно-значимых решений). В исторической ретроспективе разные значения ε (α) могут быть идентифицированы с процессом институциональных сдвигов, общественно-значимыми событиями, распространением политико-экономических знаний и уровня образования, уровнем человеческого развития и инвестициями в человеческий капитал.

Выводы

Общественные стремления реализуются в институциональных решениях. После достижения ситуации, когда созданы условия и достигнуто состояние, при котором молодые поколения, которые выходят на рынок труда имеют равные начальные условия, общественный интерес к проблемам социальной справедливости падает и общество переходит от утилитарного к либеральному пути развития. Соответственно, основное внимание отводится экономической эффективности, такие категории, как социальная справедливость и равенство отходят на второй план. Наблюдается либерализация налоговой политики, смягчаются (или исчезают вовсе) перераспределительные компоненты в пределах поколений и между поколениями. Это приводит к постепенному росту неравенства и общественной антипатии к неравенству, что, в свою очередь, приводит к включению институциональных механизмов

противодействия таким сдвигам. Таким образом, можно сделать предположение, что изменение значений ε имеет, аналогично U- гипотезе Кузнецца относительно неравенства, циклический характер. Значения параметра ε (антипатии к неравенству) могут быть использованные для смещения модальной и медианной точек с целью повышения равновесия в моделях социального взаимодействия. В исторической ретроспективе идентификация фактических значений α и ε с общественно-политическими и институциональными сдвигами дает возможность отождествить латентные факторы, которые связывают эти сдвиги с изменениями значений α и ε . Таким образом, цикличность экономического развития отображается на цикличности изменений в отношении общества к неравенству. Знание характера изменений α и ε и их взаимосвязей с институциональными изменениями дает возможность избегать точек неустойчивости и поддерживать равновесие в межгрупповых взаимодействиях (примером является U-гипотеза Кузнецца, которая резко проявляется в экономиках, которые развиваются, и значительно слабее в экономиках с высоким уровнем институционального и общественно-политического развития).

Бібліографія

- [1] Новіков В.М., Сітнікова Н.П., Мусина Л.А., Семенов В.В. Соціальні трансформації: міжнародний та вітчизняний досвід. - Київ: ІЕ НАНУ, 2003.-253с.
- [2] Draper N. R. and Cox D. R. "On Distribution and their Transformation to Normality" // Journal of the Royal Statistical Society, Ser. B, vol. 31, №3, 1969, p. 472-476.
- [3] Schlesselman J. "Power Families: A Note on the Box and Cox Transformation"// Journal of the Royal Statistical Society, Ser. B, vol. 33,№2, 1971, p.307-311.
- [4] Shorroks A. F. "The class of additively decomposable inequality measures"// Econometrica, vol. 48, 1980, p.613-625.

Інформація об авторе

Виктор Васильевич Семенов – Інститут демографії і соціальних досліджень НАН України
01011 МСП Київ 011, ул. Панаса Мирного, 26, Україна; e-mail: demographe@idss.org.ua

VARIANTS OF ENCODING FOR SELECTION OF OPTIMAL SUBSET OF DIAGNOSTIC TESTS

Anna Yankovskaya, Yury Tsoy

Abstract: *The paper concerns problem of selection of optimal subset of irredundant unconditional diagnostic tests by means of evolutionary approach. Three different variants of genetic encoding to solve this problem are described. Also new view on the optimal tests subset selection problem considering multi-objective variant of the well-known traveling salesman problem is introduced. The suggestion is made that evolutionary programming approach would be more appropriate then genetic algorithm because of disadvantage of crossover use for multi-objective problems solution.*

Keywords: *optimal tests subset selection, evolutionary multi-objective optimization, diagnostic test, intelligent systems, coevolution, genetic encoding.*

ACM Classification Keywords: G.1.6 [Mathematics of Computing]: Optimization – Constraint optimization

Introduction

Selection of "good" irredundant unconditional diagnostic tests (IUDT) is of great importance for decision making in intelligent systems, since quality of obtained solutions depends significantly on properties of the used tests. However such a selection doesn't necessarily lead to an optimal solution because total number of features in selected tests set can be too large as well as time consumption and cost. Also one should take into consideration damage (risk), caused in result of features measuring for the object under investigation, for example, in geoeological or biomedicine problems.

This research continues our previous work on optimal subset of IUDTs selection [Yankovskaya, 2002, Yankovskaya, Mozheiko, 2004, Kolesnikova et al., 2005, Yankovskaya, Tsoy, 2005]. For the first time the optimization criteria and the problem of optimal tests subset selection has been formulated in the paper [Yankovskaya, 2002]. In the paper [Yankovskaya, Mozheiko, 2004] logical-combinatorial algorithm for optimal IUDTs subset selection was presented. In the paper [Kolesnikova et al., 2005] optimization criteria were further elaborated and three algorithms providing satisfaction of those criteria were proposed: logical-combinatorial with sequent satisfaction of the prescribed criteria, algorithm of optimal tests set selection on the base of hierarchies analysis method, and genetic algorithm (GA).

For solution of the optimal IUDT subset selection problem we will use evolutionary algorithm (EA) which presents heuristic search concept similar to "trials-and-errors" method. In this paper we propose two new variants of genetic encodings for candidate-solutions and also present another view on the formulated problem introducing multi-objective free traveling salesman problem – MOFTSP.

During last decade a number of models of GAs were developed, such as NSGA-II [Deb et al., 2002], PAES [Knowles, Corne, 2000], SPEA2 [Zitzler, Laumanns, Thiele, 2001], PPREA [Hallam, Graham, Blanchfield, 2006] to solve multi-objective optimization (MOO) problems. Also alternative approaches on a basis of particles swarm optimization [Alvarez-Benitez, Everson, Fielden, 2005] and differential evolution [Becerra, Coello Coello, 2006] were proposed. Some researches are aimed at reduction of the optimization criteria number (see for example [Brockhoff, Zitzler, 2006]) and this certainly appears to be promising for the optimization results, though search of competent universal method of reduction of criteria number is rather challenging (if ever possible) due to great variety of existing MOO problems.

One of the critical conditions for the success of EA in MOO problem is preserving as many undominated (incomparable) solutions within one population as possible. Such solutions correspond to different points on the Pareto front. To preserve population of undominated solutions an idea of grouping of individuals according to some similarity/difference measure emerges in various forms, for example, as niching, or as specific non-dominated selection [Deb et al., 2002]. Considering this condition the idea of GAs use to solve MOO problems looks rather contradictory, from the authors point of view, since the main searching operator in GA is crossover and use of this operator traditionally involves risk of recombination of incompliant values of the optimization parameters due to crossing of different parent individuals, though the last can be situated rather close to each other in parameters space.

We are planning to examine this by investigation of MOO problem solution using evolutionary programming (EP) algorithm, which doesn't adopt crossing of individuals. The results of EP optimization will be compared with those of GA.

Basic Notions and Definitions

Let's introduce a number of definitions [Yankovskaya, 2002, Yankovskaya, Mozheiko, 2004, Yankovskaya, 1996] and notations used in this paper.

Test is a set of features distinguishing any pair of objects belonging to different patterns.

The test is called *irredundant* if after the removal of any feature the test is not a test.

The feature is called *obligatory* if it is contained in all irredundant tests [Yankovskaya, 2000].

The feature is called *pseudoobligatory* if it is not obligatory and enters the set of irredundant tests used in decision making.

Let $\mathbf{T} = \{t_{ij} \mid i = 1, \dots, n, j = 1, \dots, m\}$ be the matrix of IUDTs and \mathbf{T}_i corresponds to the i^{th} IUDT (the i^{th} row of matrix \mathbf{T}). We denote set of characteristic features as $\mathbf{z} = \{z_j \mid j = 1, \dots, m\}$ and for each feature z_j we define its weight w_j [Yankovskaya, 1996], cost w'_j [Yankovskaya, Mozheiko, 2004] and damage w''_j [Yankovskaya, Tsoy, 2005].

The case of binary matrix \mathbf{T} is considered therefore the weight W_i of the i^{th} IUDT is

$$W_i = \sum_j w_j t_{ij}.$$

Then average test weight along all tests inside the IUDT matrix equals to:

$$\bar{W} = \frac{\sum_i W_i}{n}.$$

Number η_i of features in each test is given by $\eta_i = \sum_j t_{ij}$ and average number of features along all tests in \mathbf{T} is

$$\bar{\eta} = \frac{\sum_i \eta_i}{n}.$$

Setting of a Problem

For the given tests matrix \mathbf{T} with defined values of features weight, cost and damage it is necessary to find such submatrix \mathbf{T}_0 with n_0 rows, which corresponds to the set \mathbf{N}^0 of tests that would provide satisfaction of the following criteria (in order of significance descend):

1. \mathbf{N}^0 should contain as many pseudo-obligatory features as possible.
2. \mathbf{N}^0 should contain in total as small number features as possible.
3. \mathbf{N}^0 should have maximum possible total weight.
4. \mathbf{N}^0 should have minimum possible total cost.
5. \mathbf{N}^0 should have minimum possible total damage.

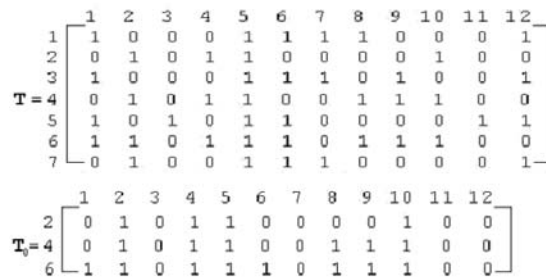
Statement of this problem accounting 5 optimization criteria was firstly introduced in the paper [Kolesnikova et al., 2005]. Since solving of the problem at hand is considered with use of evolutionary algorithm, which is known to be a heuristic search method, then as a consequence there is no guarantee that the optimal submatrix \mathbf{T}_0 (subset of IUDTs) will be found. In other words obtained solution is most likely to be suboptimal.

The problem formulated in this Section can also be considered as a modification of the well-known traveling salesman problem but here salesman is traveling for free and can visit only n_0 cities (not all the n ones) and in each city he has definite income (from sales) and expenses (cost of staying in the city). The task is to find such a path which provides the largest total income and the least total expenses. We will refer to this problem formulation as a multi-objective free traveling salesman problem – MOFTSP.

Genetic Encodings

We are going to use for comparison the following encoding schemes (for example shown on fig. 1a):

1. Candidate-solutions are encoded in binary chromosomes (strings) of length n , where each i^{th} symbol denotes inclusion ("1") (exclusion ("0")) of the i^{th} IUDT in (from) the resulting set of tests (fig. 1b). Note that number of units in the chromosome (number of IUDTs included in the resulting subset) can be unequal to n_0 therefore an additional constraint should be added for control of the number of units in chromosomes.



a) Initial matrix \mathbf{T} and solution submatrix \mathbf{T}_0

- b) Solution representation for the 1st encoding scheme: {0,1,0,1,0,1,0}
- c) Solution representation for the 2nd encoding scheme: {2,4,6}
- d) Solution representation for the 3rd encoding scheme: {0,1,0,1} ∪ {0,1,0}

Fig 1. Example of solution representation for different encoding schemes.

2. In case of the 2nd encoding scheme each chromosome includes n_0 integer-coded parameters where each parameter corresponds to the ordinal number of the IUDT in the initial matrix \mathbf{T} (fig. 1c). In the case of this encoding scheme each chromosome should contain only distinct (mutually unequal) values of parameters. From the MOFTSP viewpoint the salesman should not come twice (or more) to the one and the same city.

3. The 3rd encoding scheme uses cooperative coevolution idea [Potter, De Jong, 2000]. There are several subpopulations. Each one deals with its range of rows (submatrix of \mathbf{T}) such that submatrices for different subpopulations do not overlap. Chromosomes for each subpopulation are considered as binary strings analogous to the 1st encoding scheme. The candidate-solution is constructed by concatenation of the representative chromosomes from different subpopulations resulting in the binary chromosome similar to the chromosome for the 1st encoding scheme (example for the case of 2 subpopulations where the 1st one deals with rows 1-4 and the 2nd – with rows 5-7, is shown in fig 1d).

Let's make some comments on encodings under use.

First of all note that in case of use of the 1st and the 3rd encoding schemes there is additional optimization constraint with the greatest weight. Therefore we can expect that certain number of generations in the beginning of the evolutionary search will be spend to find the candidate-solutions that correspond to the IUDTs subset of power n_0 . The search of the solution satisfying to the prescribed optimization criteria can be performed only when this stage is over. In this connection search time for the case of the 1st and the 3rd encoding schemes is expected to be larger than that of for the 2nd encoding case. To overcome this deficiency of the 1st and the 3rd encodings an initialization of the binary chromosomes including exactly n_0 units can be proposed.

Use of the 2nd encoding scheme is connected with the problem mentioned above in this section. Since no IUDT can be included twice or more in the resulting subset, there should be a mechanism that eliminates incorrect candidate-solutions. Next, note that enumeration order of the numbers of tests included at the resulting subset doesn't matter. In other words, permutations of parameters inside the chromosome doesn't change the result (since the salesman is traveling for free). For example, solution shown in fig. 1c can also be presented as {2,6,4}

or {6,2,4} etc. Such an uncertainty involves the probability of presenting inside the population different permutations of the one and the same candidate-solution and thus slows the evolutionary search. To avoid this we will sort parameters inside the chromosome in the increasing order.

Objective Function

We will calculate fitness of the individual with chromosome h by evaluation of quality of corresponding submatrix $\mathbf{T}(h)$ as follows [Yankovskaya, Tsoy, 2005]:

$$f_h = \sum_{k=1}^5 v_k e_h^{(k)} + 100(U(h) - n_0)^2,$$

where v_k is a weight coefficient for the k^{th} optimization criterion corresponding to its significance; $U(\boldsymbol{\psi})$ gives number of units in binary string $\boldsymbol{\psi}$; $e_h^{(k)}$ is a penalty function for violation of the k^{th} criterion:

$$e_h^{(1)} = \frac{m - U_c(\mathbf{T}_0(h))}{m}, \quad e_h^{(2)} = \frac{U_d(\mathbf{T}_0(h))}{m},$$

$$e_h^{(3)} = \frac{S_w(\mathbf{T}) - S_w(\mathbf{T}_0(h))}{S_w(\mathbf{T})}, \quad e_h^{(4)} = \frac{S_{w'}(\mathbf{T}_0(h))}{S_{w'}(\mathbf{T})},$$

$$e_h^{(5)} = \frac{S_{w''}(\mathbf{T}_0(h))}{S_{w''}(\mathbf{T})},$$

where $S_w(\boldsymbol{\Psi})$, $S_{w'}(\boldsymbol{\Psi})$ and $S_{w''}(\boldsymbol{\Psi})$ – total weight, cost and damage correspondingly along all tests of the set of IUDTs corresponding to matrix $\boldsymbol{\Psi}$; $U_c(\boldsymbol{\Psi}) = U\left(\bigwedge_i \boldsymbol{\psi}_i\right)$ and $U_d(\boldsymbol{\Psi}) = U\left(\bigvee_i \boldsymbol{\psi}_i\right)$ – correspondingly number of units in conjunction and disjunction along all rows of binary matrix $\boldsymbol{\Psi}$. Evolutionary search is aimed at minimization of f .

In order to respect priorities of criteria mentioned above we will reduce weights of penalties with growth of penalty number k . Then the following penalties weights will be used: $v_1 = 40$, $v_2 = 30$, $v_3 = 15$, $v_4 = 10$, $v_5 = 5$. Note that penalties weights depend on the specific application.

Conclusion

Three variants of genetic encoding schemes to solve problem of optimal tests subset selection had been introduced in this paper. Also a new variant of the problem under consideration: the multi-objective free traveling salesman problem – MOFTSP had been introduced. It's worth noting that the optimal tests subset selection problem can also be reduced to a problem of search of optimal row coverings for Boolean matrix [Yankovskaya, Gedike, 1999].

In result of critical analysis of application of GA for solution of MOO problems and suggested deficiencies involved by crossover operator, use of EP algorithm instead of GA is proposed.

Future work is connected with experimental comparison of use of GA and EP with different encodings for the solution of the formulated problem of optimal IUDTs subset selection.

Implemented algorithms will be used in instrumental intelligent tool IMSLOG [Yankovskaya et al., 2003] for regularities revealing and decision making on the basis of test pattern recognition.

Bibliography

[Yankovskaya, 2002] A.E. Yankovskaya. Construction of logical tests of prescribed properties and logic-combinatorial pattern recognition on them // Abstracts of the Intellectualization of Information Processing (IIP2002). Simferopl, 2002. P. 100-102. (in Russian).

-
- [Yankovskaya, Mozheiko, 2004] A.E. Yankovskaya, V.I. Mozheiko. Optimization of a set of tests selection satisfying the criteria prescribed // 7th International Conference on Pattern Recognition and Image Analysis: New Information Technologies (PRIA-7-2004). Conference Proceedings. Vol. I. – St. Petersburg: SPbETU 2004. – Pp.145-148.
- [Kolesnikova et al., 2005] S.I. Kolesnikova, V.I. Mozheiko, Y.R. Tsoy, A.E. Yankovskaya. Algorithms of selection of optimal set of irredundant diagnostic tests in intelligent decision making systems // Proceedings of the First International conference "System analysis and information technologies" (SAIT-2005). Vol. 1. Moscow: KomKniga, 2005. P. 256-262.
- [Yankovskaya, Tsoy, 2005] A.E. Yankovskaya, Y.R. Tsoy. Optimization of a set of tests selection satisfying the criteria prescribed using compensatory genetic algorithm // Proceedings of IEEE East-West Design & Test Workshop (EWDTW'05) Odessa, Ukraine, September 15-19, 2005. Kharkov: SPD FL Stepanov V.V. P. 123-126.
- [Deb et al., 2002] K. Deb, S. Agrawal, A. Pratap, T. Meyarivan. A fast and elitist multi-objective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation, 2002, vol. 6, no. 2, pp. 182–197.
- [Yankovskaya, 2000] Yankovskaya A.E. Logical Tests and Cognitive Graphic Means in intelligent system // New Information Technologies in Investigations of Discrete Structures: Proceedings of the 3-d All-Russian Conf. with Foreign Participants. – Tomsk: SO RAS, 2000. – P. 163-168. (in Russian).
- [Knowles, Corne, 2000] J.D. Knowles, D.W. Corne. Approximating the non-dominated front using the Pareto archived evolution strategy. Evolutionary Computation Journal, 2000, vol. 8, no. 2, pp. 149–172.
- [Zitzler, Laumanns, Thiele, 2001] E. Zitzler, M. Laumanns, L. Thiele. SPEA2: Improving the strength pareto evolutionary algorithm for multi-objective optimization. In K. C. Giannakoglou, D. T. Tsahalis, J. P'eriaux, K. D. Papailiou, and T. Fogarty, editors, Evolutionary Methods for Design Optimization and Control with Applications to Industrial Problems, pages 95–100, Athens, Greece, 2001.
- [Hallam, Graham, Blanchfield, 2006] N. Hallam, K. Graham, P. Blanchfield. Solving multi-objective optimization problems using the potential pareto regions evolutionary algorithm // T.P. Runarsson et al. (eds.): PPSN IX, LNCS 4193, pp. 503-512. Springer-Verlag, 2006.
- [Alvarez-Benitez, Everson, Fieldsen, 2005] J.E. Alvarez-Benitez, R.M. Everson, J.E. Fieldsen. A MOPSO algorithm based exclusively on pareto dominance concepts // In C. Coello-Coello et al. (eds.): Evolutionary Multi-Criterion Optimization, vol. 3410, pp. 459-473. Springer, 2005.
- [Becerra, Coello Coello, 2006] R.L. Becerra, C.A. Coello Coello. Solving hard multi-objective optimization problems using -constraint with cultured differential evolution // T.P. Runarsson et al. (eds.): PPSN IX, LNCS 4193, pp. 543-552. Springer-Verlag, 2006.
- [Brockhoff, Zitzler, 2006] D. Brockhoff, E. Zitzler. Are all objectives necessary? On dimensionality reduction in evolutionary multi-objective optimization // T.P. Runarsson et al. (eds.): PPSN IX, LNCS 4193, pp. 533-542. Springer-Verlag, 2006.
- [Yankovskaya, 1996] A.E. Yankovskaya. Design of Optimal Mixed Diagnostic Test With Reference to the Problems of Evolutionary Computation // Proceedings of the First International Conference on Evolutionary Computation and Its Applications (EVCA'96). Moscow, 1996. P. 292-297.
- [Potter, De Jong, 2000] M. Potter, K. De Jong. Cooperative coevolution: An architecture for evolving coadapted subcomponents // Evolutionary Computation, 2000, vol. 8, no. 1, pp. 1-29.
- [Yankovskaya, Gedike, 1999] A. Yankovskaya, A. Gedike. Finding of All Shortest Column Coverings of Large Dimension Boolean Matrices// Proceedings of the First International Workshop on Multi-Architecture Low Power Design (MALOPD). – ISBN 5-93576-002-9, Moscow, 1999. – pp. 52-60. (<http://www.dice.ucl.ac.be/~anmarie/MALOPD>).
- [Yankovskaya et al., 2003] A.E. Yankovskaya, A.I. Gedike, R.V. Ametov, A.M. Bleikher. IMSLOG-2002 Software Tool for Supporting Information Technologies of Test Pattern Recognition// Pattern Recognition and Image Analysis, 2003, vol. 13, no. 4, pp. 650-657.
-

Authors' Information

Anna E. Yankovskaya – Tomsk State University of Architecture and Building, 2, Chitinskaya Str, apt. 28, 634003, Tomsk, Russia; e-mail: yank@tsuab.ru

Yury R. Tsoy – Tomsk Polytechnic University, 84, Sovetskaya Str, 634050, Tomsk, Russia; e-mail: gai@mail.ru

АСПЕКТЫ НЕСТАБИЛЬНОСТИ ТЕРМОМЕТРИЧЕСКИХ ХАРАКТЕРИСТИК ПРЕОБРАЗОВАТЕЛЕЙ ТЕМПЕРАТУРЫ

Святослав Яцишин, Богдан Стадник, Юрий Лега,
Ярослав Луцик, Вениамин Мельник

Abstract: *The investigations of different kinds of thermometric materials have shown the possibilities of developing of high quality temperature transducers by means of thermodynamics of irreversible processes.*

Введение

На сегодня, в арсенале термометрических материалов (далее по тексту - ТМ), кроме традиционных поли- и монокристаллических ТМ, с успехом используются новые классы ТМ – специальные микропористые и аморфные ТМ [1]. На очереди проектирование, изготовление и использование наноматериалов [2]. Однако их применение тормозится недостаточной изученностью последних, отсутствием планомерного и дифференцированного подхода к их изучению при наличии широкого спектра факторов возмущения.

Достижения метрологии в направлении создания интеллектуальных термопреобразователей (далее по тексту - ТП) позволяют сосредоточиться исключительно на таких характеристиках ТМ, как воспроизводимость и стабильность во времени под действием множества реальных факторов возмущения, не принимая во внимание величины и знаки параметров, обеспечивающих функцию преобразования. Логичным представляется применение статистически-термодинамических представлений относительно упомянутых характеристик, учитывая эффективность привлечения термодинамики неравновесных процессов [3] к подобным объектам.

При разработке и использовании преобразователей температуры повышенной точности необходимо учитывать особенности термодинамического состояния термометрических материалов. Поэтому проведены исследования аморфных, моно- и поликристаллических ТМ и разработаны алгоритмические основы корректирования функций преобразования соответственно влиянию конкретных термодинамических сил, потоков и их комбинаций.

Цель работы

Разработка статистически-термодинамического алгоритма проектирования ТП средств электротермометрии, базируясь на совокупности экспериментально-теоретических исследований разных классов ТМ для широкого диапазона температур.

Основные материалы исследований

Деградационные процессы, протекающие в частично открытых термодинамических системах, способных обмениваться энергией или/и потоками энтропии с внешней средой, закладывают термодинамические основания для вывода со строя различных конструктивных элементов ТП. Отказ происходит в ослабленной, энергетически напряженной зоне, какой считается зона горячего спая ТП.

Снижение уровня электромеханохимических шумов ТМ способствовало повышению метрологической и механической надежности ТП, так как отказы обусловлены деградацией основных характеристик ТП. Неразрывность конструктивно-технологично-эксплуатационного подхода относительно разработки, изготовления и эксплуатации высоконадежных ТП позволила предложить типовую конструкцию кабельного ТП (тип ТХА-1388) с повышенными параметрами надежности до $\lambda = 0,98$ за 25000 часов эксплуатации. Максимальной стабильностью функции преобразования характеризовались

термоэлектрические и термошумовые ТП с аморфными ТМ, у которых отклонений не выявлено (в тех же условиях отклонения для поликристаллических ТМ достигали 10%).

В результате проведенных работ разработаны принципы и реализованы способы построения высокоточных ТП средств электротермометрии. На их основе создан алгоритм оценки воздействия функций влияния на функции преобразования ТП с использованием предварительно установленных факторов и методологии статистической термодинамики. Для этого определены процессы переноса термодинамических параметров (сил и потоков), присущие ТМ ТП средств электротермометрии, в конкретных условиях эксплуатации. Соответственно благодаря этому (рис.1) происходило перемещение в поле Алгоритма, причем путь перемещения определялся посредством коэффициентов $K_1; K_2; K_3$.

Указанным алгоритмом рассмотрено 6 степеней свободы основного уравнения термодинамики. Главными для ТМ ТП средств электротермометрии считаются электрическая и тепловая степени свободы, которые формировали электрические шумы. Именно они определяли функции преобразования терморезистивных, термоэлектрических и термошумовых ТП - соответственно $\sigma(T); U(T); P(T)$. Все другие степени свободы формировали функции влияния, действие которых оценено последовательно, начиная с наименее существенных для типовых условий использования ТП (такими считаются магнитная и поверхностная степени свободы). Их механизм действия реализован через процессы переноса определенных термодинамических сил и потоков, отличающихся для каждой из следующих степеней свободы: химической, механической, поверхностной и магнитной.

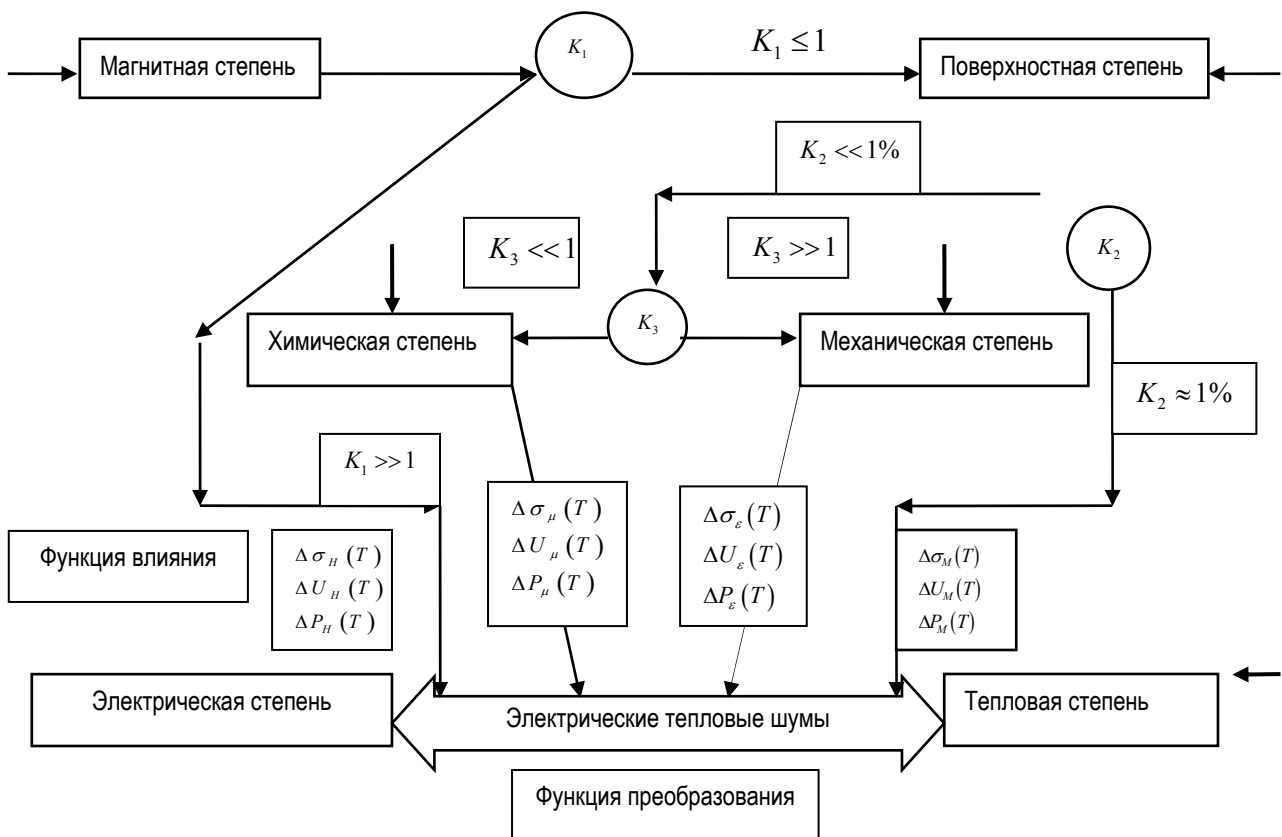


Рис.1. Формирование функций преобразования и влияния ТП средств электротермометрии зависит от веса определенных степеней свободы основного уравнения термодинамики для ТМ:

$$\text{Постоянные определены из уравнений } K_1 = \frac{\mu H}{kT}; K_2 = \frac{\rho - \rho_0}{\rho} 100\%; K_3 = \frac{X_{\sigma\sigma}}{X_{Tt}}$$

Для оценки влияния магнитной степени и возможности пренебрежения ею введено критерий $K_1 = \frac{\mu H}{kT}$.

При $K_1 \gg 1$ (большие магнитные поля и «гелиевые» температуры) влияние магнитной степени свободы считается определяющим. Когда $K_1 < 1$, переходят к оценке влияния поверхностной степени на функцию преобразования. В условиях существенной ($K_2 > 1\%$) пористости ТМ металлокерамической технологии производства данная степень оказывала непосредственное влияние на функцию преобразования через соответствующую функцию влияния. Следуя данному Алгоритму, функция влияния поверхностной степени определялась посредством шумов, связанных с изменением плотности. При $K_2 < 1\%$ (монокристаллический ТМ) функцию влияния можно считать несущественной. Поэтому можно осуществить переход к оценке влияния химической и механической степеней свободы. Здесь предварительными установками заданы температурно-механически-временные режимы, которые определяли превалирование химической степени (процессы термодиффузии) или механической степени (процессы деформационного уплотнения). Для этого рассчитывали коэффициент K_3 согласно диффузионных $X_{t,T}$ и «механических» $X_{\sigma,\varepsilon}$ смещений атомов матрицы ТМ. Значения $K_3 \ll 1$ соответствовали превалированию химической степени, а $K_3 \gg 1$ - механической степени. Величины функций влияния, обусловленные воздействием конкретных процессов переноса, рассчитывали согласно представленных выражений или производили их оценку согласно результатам дополнительных исследований, для чего предусмотрено возможность расширения Алгоритма.

Основная погрешность ТП средств электротермометрии, обусловленная изменениями функции преобразования под действием функций влияния, сведена к суммарному коэффициенту неустойчивости K_{Σ} . Это - граничное значение относительной погрешности конкретного типа термопреобразователя, выраженное в виде комбинации трех безразмерных коэффициентов неустойчивости:

$$K_{\Sigma} = (K_x + K_m) K_T,$$

где K_x ; K_m ; K_T - соответственно химический, механико-поверхностный и температурный коэффициенты неустойчивости функции преобразования.

Каждому из приведенных коэффициентов, зависимо от особенностей изготовления и эксплуатации, присущи собственные происхождение и механизм действия, а также результирующее значение. Данное выражение получено в результате применения статистической термодинамики к изучению проблемы стабильности термометрических характеристик преобразователей температуры.

Выводы

В результате проведенного анализа предложено пути совершенствования ТП средств электротермометрии, которые заключаются в учете термодинамического состояния ТМ. Оптимизировано методика и разработаны алгоритмические основания корректировки функций преобразования согласно влиянию конкретных термодинамических сил и потоков на аморфные, моно- и поликристаллические ТМ, что позволяет решить проблему сквозного проектирования, изготовления и применения ТП.

Литература

1. Б.И.Стадник и др. Температурные измерения. Справочник. Киев: Наукова думка, 1986, 282 с.
2. P.I.Skoropad, B.I.Stadnyk, S.P.Yatsyshyn. Technological and Thermodynamic Changes in Parameters of Thermoelectric Materials // Journal of Thermoelectricity. # 2, 2004, p. 30-36.
3. Н.К.Булатов, А.Б.Лундин. Термодинамика необратимых физико-химических процессов. Москва: Химия, 1984, 334 с.

Информация об авторах

Яцишин Святослав Петрович, к.т.н., доцент; **Луцик Ярослав Теодорович**, д.т.н., проф. - Национальный университет "Львівська політехніка", Львов, Украина; yaroslav.lutsyk@gmail.com

Стадник Богдан Иванович, д.т.н., проф. – Жешовская политехника, Жешув, Польша;

Лега Юрий Григорьевич, д.т.н., проф.; **Мельник Вениамин Васильевич**, к.т.н., доцент – Черкасский государственный технологический университет, Черкассы, Украина

КОМБИНИРОВАННЫЙ МЕТОД РЕШЕНИЯ ЗАДАЧИ О РЮКЗАКЕ

Татьяна Пермякова, Владимир Морозенко

Аннотация: Статья посвящена решению проблем применимости генетического алгоритма для решения задачи о рюкзаке и описанию комбинированного метода для решения этой задачи, включающей в себя улучшение решения, полученного в результате работы генетического алгоритма, при использовании идей, составляющих основу метода ветвей и границ. Результаты проведенного тестирования показывают, что разработанный метод характеризуется достаточно высокой степенью точности.

Keywords: задача о рюкзаке, генетический алгоритм, метод ветвей и границ.

ACM Classification Keywords: F.2 Analysis of Algorithms and Problem Complexity: F.2.1 Numerical Algorithms and Problems, G.2 Discrete Mathematics G.2.1 Combinatorics, G.1.6 Optimization, I.2 Artificial Intelligence: I.2.8 Problem Solving, Control Methods, and Search - Heuristic methods.

Введение

Задача о рюкзаке – это одна из классических задач дискретной оптимизации. Пусть имеется множество предметов, каждый из которых имеет определенную стоимость и вес. Требуется составить такой набор этих предметов, который имел бы суммарную стоимость, максимально возможную среди всех наборов, чей суммарный вес не превосходит заданной величины – вместимости рюкзака. Более точно, пусть $c_i > 0$ и $a_i > 0$ – соответственно стоимость и вес i -го предмета, где $i = 1, 2, 3, \dots, n$, а n – число предметов. Требуется найти такой булев вектор (x_1, x_2, \dots, x_n) , чтобы была максимальной сумма

$$\sum_{i=1}^n c_i x_i ,$$

и выполнялось неравенство

$$\sum_{i=1}^n a_i x_i \leq P,$$

где $P > 0$ – вместимость рюкзака.

К этой задаче часто сводятся практические проблемы, касающиеся поиска оптимального распределения некоторого ресурса при наличии ряда ограничивающих факторов. Задача о рюкзаке, как и многие оптимизационные комбинаторные задачи, принадлежит к классу NP -полных задач [Сигал, 2003]. Её можно решить полным перебором всех допустимых вариантов заполнения рюкзака имеющимися предметами, однако при больших массивах входных данных такой переборный алгоритм практически неприемлем, поскольку имеет экспоненциальную сложность относительно длины входа. Если же применить идеи «метода ветвей и границ», то в большинстве случаев объем перебираемых вариантов можно сократить. Подобные алгоритмы с ограниченным перебором могут иметь неплохую сложность в среднем, но в худшем случае все равно остаются экспоненциальными. Поэтому на практике часто используют различные модификации «жадного» алгоритма, который имеет полиномиальную сложность, что и является его основным достоинством. Главный же недостаток «жадного» алгоритма в том, что он не всегда находит правильное решение. В этом смысле его можно отнести к классу приближенных алгоритмов.

В задаче о рюкзаке требуется найти максимум функционала – суммарной стоимости положенных в рюкзак предметов. Пространство поиска при этом ограничивается дискретным множеством всех допустимых вариантов заполнения рюкзака, число которых может достигать величины 2^n , где n – количество предметов. Все эти особенности рассматриваемой задачи позволяют воспользоваться для её решения генетическим алгоритмом. Такой алгоритм при разумном кодировании допустимых решений, как известно, может оказаться вполне эффективным [Вороновский, 1992]. Однако его тоже следует отнести к классу приближенных алгоритмов, поскольку он не всегда дает оптимальный вариант заполнения рюкзака.

Одна из основных причин получения неправильного ответа в результате работы генетического алгоритма заключается в возможной преждевременной сходимости алгоритма. Это случается, когда все особи, образующие очередную популяцию, оказываются сгруппированными в «окрестности» локального экстремума целевой функции, и при этом никакие усилия операторов скрещивания, мутации и отбора не приводят к «смещению» наиболее приспособленных особей в «направлении» искомого оптимального решения, доставляющего глобальный экстремум целевой функции. Увеличение числа популяций также не дает ожидаемого эффекта, а существенное увеличение их численности мало будет отличаться от метода полного перебора вариантов. В данной работе предпринята попытка избежать преждевременной сходимости генетического алгоритма за счет использования идей метода «ветвей и границ».

Генетический алгоритм

При разработке генетического алгоритма для решения задачи о рюкзаке необходимо ответить на следующие вопросы [Рутковская, 2004]:

1. Каким образом в хромосоме будет закодировано допустимое решение?
2. Как вычислять функцию приспособленности особи?
3. Как реализовывать генетические операторы скрещивания и мутации, чтобы полученные в результате их применения особи являлись допустимыми решениями?
4. На каком основании будет происходить формирование новой популяции?
5. Каково условие завершения работы алгоритма?

Применим следующий естественный способ кодирования решения. Изначально все предметы пронумерованы некоторым образом. Каждая хромосома будет представлять собой булев вектор (x_1, x_2, \dots, x_n) , в котором i -й ген (т.е. элемент x_i) равен единице, если i -й предмет положили в рюкзак, и равен нулю в противном случае. Очевидно, что не любой булев вектор (x_1, x_2, \dots, x_n) указанного вида будет кодировать допустимое решение, поскольку существует ограничение на набор предметов, положенных в рюкзак, обусловленное его вместимостью.

Функция приспособленности (фитнесс-функция) особи характеризует степень «близости» этой особи к точному решению задачи. Чем больше значение функции приспособленности особи, тем ближе соответствующее ей решение задачи к искомому максимуму. Приспособленность особи (x_1, x_2, \dots, x_n) будем вычислять по формуле

$$f(x_1, x_2, \dots, x_n) = \sum_{i=1}^n c_i x_i,$$

т.е. приспособленность особи – это суммарная стоимость предметов, положенных в рюкзак.

При описанном выше способе кодирования допустимых решений возникает проблема, связанная с использованием стандартных генетических операторов скрещивания (кроссовера) и мутации. При применении односточечного оператора скрещивания из допустимых хромосом родителей может получиться потомок, кодирующий недопустимое решение, т.е. такой булев вектор, для которого суммарный вес выбранных предметов превышает заданную вместимость рюкзака. При использовании обычного оператора мутации из допустимой хромосомы также может получиться хромосома, кодирующая недопустимое решение.

Чтобы избежать этих трудностей предлагается использовать стандартные операторы односточечного или равномерного кроссовера с последующей коррекцией получаемых хромосом, если это потребуется. Коррекция состоит в том, что в полученной недопустимой хромосоме случайным образом выбираются единичные гены и заменяются нулевыми генами до тех пор, пока не получим допустимую хромосому. Аналогично поступаем и для недопустимых хромосом, получаемых в результате мутации.

Работу генетического алгоритма можно завершить в двух случаях. Во-первых, когда значения фитнес-функции, соответствующие наиболее приспособленной особи в нескольких подряд идущих популяциях, оказались достаточно близки. Например, можно использовать следующее эвристическое правило: завершать работу алгоритма, когда на протяжении последних 50 поколений итераций не изменилась максимальная приспособленность особи популяции. Во-вторых, работу алгоритма можно остановить при достижении заранее выбранного количества поколений.

Комбинированный метод решения задачи

Основным недостатком описанного выше генетического алгоритма является неприятная возможность преждевременного завершения его работы, когда найденное им решение доставляет локальный максимум фитнес-функции, но не является точкой её глобального максимума. В то же время алгоритмы, основанные на методе ветвей и границ, лишены этого недостатка, поскольку являются точными алгоритмами. Однако они обладают экспоненциальной сложностью, и это ограничивает их возможности при решении задач большой размерности. Что касается генетического алгоритма, то значительное увеличение размерности задачи не так сильно отражается на его вычислительных возможностях.

С учетом указанных достоинств и недостатков этих двух типов алгоритмов предлагается комбинированный подход, сочетающий в себе генетический алгоритм с идеями метода ветвей и границ.

Данный подход состоит в том, что сначала применяется генетический алгоритм, а затем найденное им решение улучшается за счет использования метода ветвей и границ.

Пусть с помощью генетического алгоритма был найден некоторый вариант заполнения рюкзака. Он не обязательно является наилучшим. Однако, как и любое решение, получаемое генетическим алгоритмом, оно кодируется в хромосоме строкой (y_1, y_2, \dots, y_n) из нулей и единиц длины n , где n – количество предметов. Позиции единиц в этой строке определяют номера предметов, положенных в рюкзак. Далее будем считать, что предметы пронумерованы в порядке убывания их удельной ценности (т.е. стоимости единицы веса). Известно, что на практике неплохо работает «жадный» алгоритм полиномиальной сложности, где в рюкзак в первую очередь укладывают наиболее ценные предметы [Гэри, 1982]. Отбросим последние h элементов строки (y_1, y_2, \dots, y_n) . Получим строку $(y_1, y_2, \dots, y_{n-h})$, которая соответствует такому варианту заполнения рюкзака, когда все предметы, положенные в рюкзак согласно решению генетического алгоритма, и имеющие номера, большие, чем $n-h$, удаляются из рюкзака. Вполне вероятно, что полученную укороченную строку $(y_1, y_2, \dots, y_{n-h})$ можно теперь продлить строкой длины h из нулей и единиц так, чтобы получился новый вариант заполнения рюкзака, имеющий большую суммарную стоимость по сравнению с решением генетического алгоритма (y_1, y_2, \dots, y_n) . Более того, используя метод ветвей и границ, можно найти наилучшее продолжение строки $(y_1, y_2, \dots, y_{n-h})$. Действительно, укороченная строка $(y_1, y_2, \dots, y_{n-h})$ указывает на предметы, которые уже положены в рюкзак. Это означает, что вместимость той части рюкзака, которая пока ещё остается свободной, уменьшилась ровно на сумму весов предметов, уже положенных в рюкзак. Таким образом можно считать, что у нас имеется новый рюкзак вместимости P' и стоимости C' , где

$$P' = P - \sum_{i=1}^{n-h} a_i y_i, \quad C' = \sum_{i=1}^{n-h} c_i y_i.$$

Наша цель – уложить в него часть оставшихся предметов так, чтобы получился вариант заполнения рюкзака максимально возможной стоимости. Иными словами, после работы генетического алгоритма мы свели исходную задачу о рюкзаке к такой же задаче, но с меньшим числом предметов и с рюкзаком меньшей вместимости.

Применим к новой задаче метод ветвей и границ. Будем строить бинарное дерево высоты h , каждой вершине которого соответствует строка длины n из нулей и единиц. Единицы в такой строке указывают на номера предметов, положенных в рюкзак. Корень дерева имеет высоту h , и ему соответствует строка $(y_1, y_2, \dots, y_{n-h}, 0, \dots, 0)$. Вершине высоты k соответствует строка $(y_1, y_2, \dots, y_{n-h}, z_1, z_2, \dots, z_{h-k}, 0, \dots, 0)$, а двум её прямым потомкам – строки $(y_1, y_2, \dots, y_{n-h}, z_1, z_2, \dots, z_{h-k}, 1, 0, \dots, 0)$ и $(y_1, y_2, \dots, y_{n-h}, z_1, z_2, \dots, z_{h-k}, 0, 0, \dots, 0)$, где $k = 1, 2, 3, \dots, h-1$. Они кодируют два «близких» варианта заполнения рюкзака, которые отличаются лишь тем, что в первом случае рюкзак содержит предмет с номером $n-k$, а во втором – нет. Заметим, что не всякая строка является допустимой, т.е. описывает допустимый вариант заполнения рюкзака с учетом его вместимости. Поэтому вершина, которой соответствует недопустимая строка, не имеет потомков.

Предположим, что вершине высоты k соответствует допустимая строка $(y_1, y_2, \dots, y_{n-h}, z_1, z_2, \dots, z_{h-k}, 0, \dots, 0)$, где $k = 2, 3, \dots, h$. Вычислим для этой вершины величину перспективности $persp$ по формуле

$$persp = \sum_{i=1}^{n-h} c_i \cdot y_i + \sum_{j=1}^{h-k} c_{n-h+j} \cdot z_j + \left(P - \sum_{i=1}^{n-h} a_i \cdot y_i - \sum_{j=1}^{h-k} a_{n-h+j} \cdot z_j \right) \cdot c_{n-k+1} / a_{n-k+1}.$$

Перспективность вершины является ограничением сверху на суммарную стоимость укладки рюкзака, которая может быть найдена в результате достраивания дерева вниз от этой вершины. Более точно,

перспективность вершины высоты k равна максимальной стоимости укладки рюкзака, на которую можно рассчитывать, если известно, какие из предметов с номерами $1, 2, 3, \dots, n - k$ уже положены в рюкзак. Перспективность вершины первого уровня, если ей соответствует допустимая строка, вычисляется по формуле

$$Persp = \sum_{i=1}^{n-h} c_i \cdot y_i + \sum_{j=1}^{h-1} c_{n-h+j} \cdot z_j + c_n.$$

Согласно идее метода ветвей и границ описанное бинарное дерево строится постепенно сверху вниз, т.е. от корня к листьям. На каждом шаге построение дерева продолжается из вершины с максимальной величиной перспективности.

Таблица 1. Исходные данные

Номер предмета	1	2	3	4	5	6
Стоимость	5	7	8	6	4	1
Вес	2	3	4	3	2	1
Удельная ценность	5/2	7/3	2	2	2	1

Приведем пример, в котором таким образом удалось улучшить решение, найденное генетическим алгоритмом. Пусть имеется 6 предметов, веса и стоимости которых указаны в таблице 1. Будем считать, что вместимость рюкзака $P = 7$.

Предположим, что результатом работы генетического алгоритма стала хромосома (101001). Для такой укладки рюкзака суммарный вес уложенных предметов равен 7, а их суммарная стоимость равна 14. Положим параметр h равным 4. Тогда, используя описанный выше комбинированный метод, получим дерево, изображенное на рис. 1.

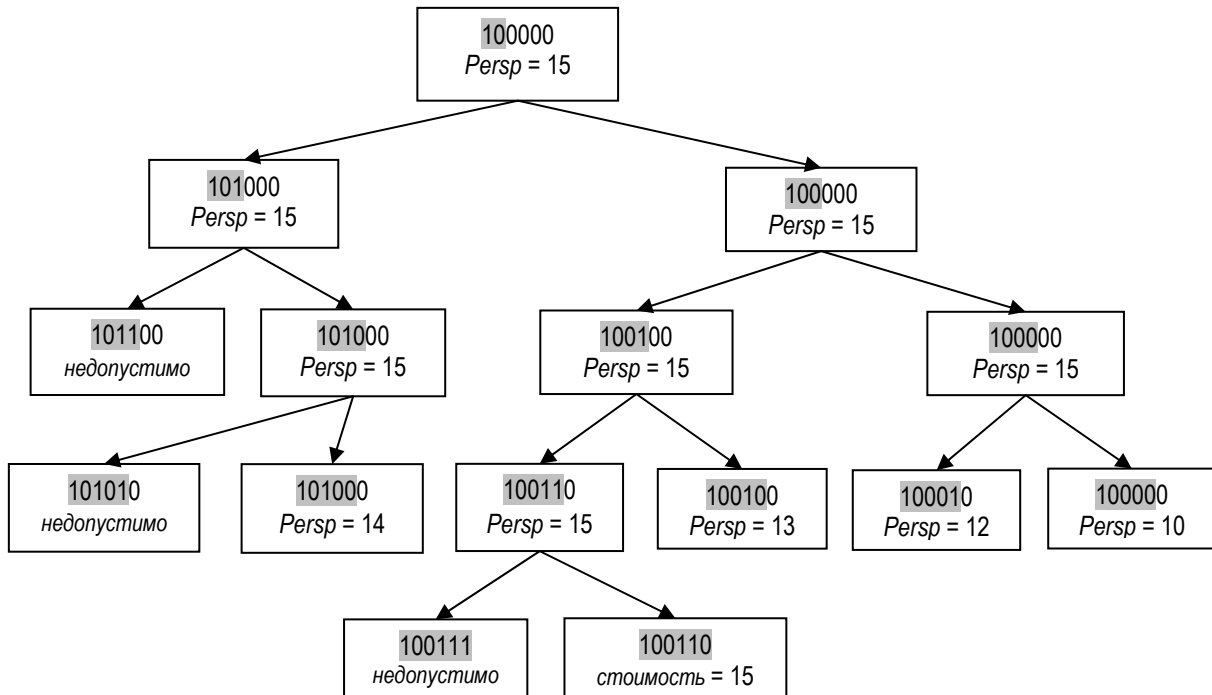


Рис. 1. Дерево, построенное комбинированным методом

Построение дерева начинается с вершины, которой соответствует строка (100000). Она находится на высоте 4, а её перспективность равна 15. Затем строим её потомков на высоте 3, которым соответствуют строки (101000) и (100000). Величины их перспективности также равны 15. Продолжаем строить дерево из вершины (101000). Образует два её потомка (101100) и (101000) на высоте 2. Первый потомок является недопустимым вариантом заполнения рюкзака, а второй – допустимым с перспективностью 15. Рассматривая его потомков (101100) и (101000), получаем, что один из них кодирует недопустимый вариант заполнения рюкзака, а другой – допустимый с перспективностью 14. Поэтому далее переходим к рассмотрению потомков вершины (100000), расположенной на высоте 3 и т.д. В итоге получаем концевые вершины дерева (101100), (101010), (100111), кодирующие недопустимые варианты заполнения рюкзака, концевую вершину (100110), кодирующую вариант стоимости 15, и вершины (101000), (100100), (100010) и (100000), чья перспективность ниже, чем 15. Следовательно, строка (100110) является «условно лучшим» вариантом заполнения рюкзака (при условии, что предмет с номером 1 будет положен в рюкзак, а предмет с номером 2 – нет). На самом деле «абсолютно лучший» вариант заполнения рюкзака для данной задачи имеет стоимость 16 и кодируется строкой (110010).

Таким образом, используя метод ветвей и границ, в конкретном рассмотренном примере удалось улучшить решение, полученное генетическим алгоритмом. Однако необходимо заметить, что, во-первых, в общем случае такое улучшение возможно лишь при достаточно больших значениях параметра h . Например, нетрудно убедиться, что в рассмотренной задаче решение генетического алгоритма было бы невозможно улучшить, если бы мы выбрали $h < 4$. Во-вторых, найденное улучшенное решение все равно может оказаться не абсолютно лучшим, как это и оказалось в рассмотренном примере. Можно лишь утверждать, что «качество» получаемого решения повышается одновременно с ростом параметра h . В крайнем случае, когда параметр h принимает максимально возможное значение, равное n – числу предметов, предложенный комбинированный метод вообще не использует генетический алгоритм, а с самого начала работает как метод ветвей и границ.

В изложенном комбинированном методе время работы генетического алгоритма пропорционально размеру популяции и числу поколений, и полиномиально зависит от параметра n , а время работы метода ветвей и границ экспоненциально относительно параметра h . Выбирая разные значения этих параметров, можно регулировать время работы алгоритма.

Для оценки качества полученного комбинированного алгоритма случайным образом было сформировано около 100 тестовых примеров с количеством предметов от 100 до 500. В этих тестовых примерах для поиска решения использовались следующие алгоритмы:

- 1) «жадный» алгоритм (ЖА);
- 2) генетический алгоритм (ГА);
- 3) генетический алгоритм с добавлением в начальную популяцию хромосомы, соответствующей решению «жадного» алгоритма (ЖА+ГА);
- 4) комбинированный метод для улучшения решения генетического алгоритма (КМ).

Решения, полученные указанными выше приближенными методами, сравнивались с точными решениями, полученными в результате работы метода ветвей и границ.

Средние отклонения решений, полученных этими алгоритмами, от точного решения приведены на рис. 2. Отклонение определялось по следующей формуле:

$$E = \frac{A - \tilde{A}}{A} \cdot 100\% ,$$

где A – стоимость точного решения, полученного методом ветвей и границ, \tilde{A} – приближенное решение, являющееся результатом работы одного из перечисленных алгоритмов.

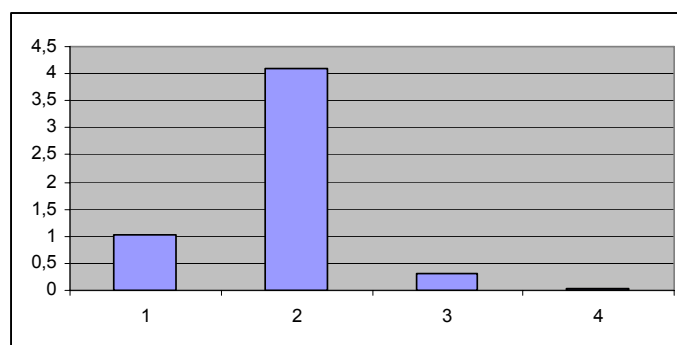


Рис. 2. Отклонение от точного решения (в процентах).

1- ЖА, 2 –ГА, 3 – ЖА+ГА, 4 – КМ

Первый столбец показывает, насколько эффективен «жадный» алгоритм. Второй несет в себе информацию о работе генетического алгоритма, начальная популяция которого сформирована случайным образом. Третий – результат работы генетического алгоритма, в начальную популяцию которого включена хромосома, соответствующая решению, полученному «жадным» алгоритмом. Четвертый показывает насколько результативно применение комбинированного метода для улучшения решения генетического алгоритма. Для тестирования генетического алгоритма использовалась популяция большой численности, а условием остановки алгоритма являлась стабилизация решения, т.е. неизменность лучшей особи на протяжении 50 итераций.

Заключение

Ранее был указан один из существенных недостатков в работе генетического алгоритма – возможная сходимость популяции к точке локального экстремума целевой функции, в результате чего генетический алгоритм выдает неправильный ответ. Снизить вероятность такой преждевременной стабилизации популяции можно было бы путем расширения исследуемой окрестности за счет увеличения числа особей в популяции. Однако в данной работе предложен другой вариант расширения области поиска – применение метода ветвей и границ к вспомогательной задаче о рюкзаке меньшей размерности, возникающей после завершения работы генетического алгоритма и учитывающей результаты его работы.

Проведенные вычислительные эксперименты показали, что комбинированный метод позволяет быстро найти близкое к точному решение задачи о рюкзаке. Как и ожидалось, улучшение решения, полученного генетическим алгоритмом, происходило при достаточно больших значениях параметра h , причем не всегда найденное улучшенное решение оказывалось правильным. Однако качество окончательных решений можно регулировать за счет выбора параметра h .

Библиографический список

- [Вороновский, 1992] Вороновский Г. К. Генетические алгоритмы, искусственные нейронные сети и проблемы виртуальной реальности. – Х.: ОСНОВА, 1992.
- [Гэри, 1982] Гэри М., Джонсон Д. Вычислительные машины и труднорешаемые задачи. – М.: МИР, 1982.
- [Гук, 2001] Гук А. К. Математическая логика и теория алгоритмов. – М.: ФИЗМАТЛИТ, 2001.
- [Кормен, 2001] Кормен Т., Лейзерсон Ч., Ривест Р. [Алгоритмы: построение и анализ](#). — М.: МЦНМО, 2001.

[Рутковская, 2004] Рутковская Д., Пилинский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы: Пер. с польск. И. Д. Рудинского. – М.: Горячая линия – Телеком, 2004.

[Сигал, 2003] Сигал И.Х., Иванова А.П. Введение в прикладное дискретное программирование: модели и вычислительные алгоритмы. – М.: ФИЗМАТЛИТ, 2003.

Сведения об авторах

Татьяна Пермякова – Пермский государственный университет, студентка магистратуры кафедры математического обеспечения вычислительных систем; Россия, г. Пермь, 614990, ул. Букирева, д. 15; e-mail: tpermjakova@mail.ru

Владимир Морозенко – Пермский государственный университет, доцент кафедры математического обеспечения вычислительных систем; Россия, г. Пермь, 614990, ул. Букирева, д. 15; e-mail: v.morozenko@mail.ru

ИНФОРМАЦИЯ, ЭКОНОМИКА, ЭКОЛОГИЯ – КОМПОНЕНТЫ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОГО ПОТЕНЦИАЛА ТЕРРИТОРИИ (ИННОВАЦИОННЫЕ ХАРАКТЕРИСТИКИ РЕГИОНА И ТРЕХМЕРНОЕ ПРЕДСТАВЛЕНИЕ КРИВОЙ С. КУЗНЕЦА)

Анатолий Крисиллов, Виктор Крисиллов

Аннотация: Рассмотрены особенности «знаниевой», инновационной экономики, ее роль в формировании социально-экономического потенциала. Описаны возможности количественной оценки инновационной деятельности; сделано трехмерное обобщение кривой С. Кузнеца (*knowledge based impact*).

Ключевые слова: инновационная деятельность, количественное описание и оценка, информационная и социально-экономическая пирамида.

Вводные замечания

Представляется, что основной темой, метазадачей большинства работ, вот уже ряд лет эффективно представляемых на конференциях KDS, является рассмотрение вопросов информатики в широком смысле. Среди этих вопросов не последнее место занимает проблема связи и взаимного влияния характеристик хозяйственно-экономического комплекса, с одной стороны, и информационной деятельности, – с другой.

В чем конкретно проявляется насыщение экономики знаниями, в чем информационная/инновационная компонента будет сказываться на уровне и качестве жизни, каковы особенности инновационной составляющей социально-экономического потенциала территории и как эту составляющую можно оценить количественно – вот некоторые вопросы, рассмотренные в настоящей работе.

1. Актуальные характеристики инновационной деятельности

В последние годы все настойчивее проводится разными авторами идея необходимости формирования и реализации на Украине стратегии инновационного развития. Имеют место: значительное количество

публикаций в научной и общественной печати, создание и упразднение инновационных фондов, принятие различных документов на разных уровнях.

Если отвлечься от фактора моды, от декларативных и нереальных заявлений, то остается большой комплекс серьезных представлений об объеме и значимости инновационной деятельности и инновационных процессов в экономике вообще, в социально-экономической действительности, темпах и путях развития страны, – в особенности. Вкратце эти представления сводятся к следующему.

1. Происходит значительный рост вклада инновативных процессов в мировой валовой продукт. Особенно это характерно для развитых стран. Еще Й. Шумпетер связывал интенсивное развитие с инновациями: «причина развития экономики – стремление к сверхприбыли за временную монополию, получаемую от внедрения инновации» [1]. Р. Солоу стал нобелевским лауреатом в области экономики именно за анализ и демонстрацию роли этого фактора.

2. Ощутимое развитие будет обеспечено, если инновационная деятельность не ограничивается одним-двумя изделиями, одним-двумя направлениями. Эффект, особенно – территориальный, будет иметь место при условии встраивания инноваций, больших и маленьких, в разные виды производства, в сервисное обслуживание, в инфраструктуру, в деловые и административные практики.

3. Основным носителем и выражением инноваций, основной средой и механизмом продуцирования и использования инноваций, является новая экономика, экономика знаний, экономика, основанная на знаниях, knowledge-based economy. Результатом деятельности такой экономики (и одновременно – ее товаром) являются знания, особенно – новые знания. Таким образом, основные усилия затрачиваются на производство, распределение и использование знаний, в отличие от того, что для предшествующих стадий экономического развития в центре этих усилий был вещественный продукт. Содержанием этого нового «знаниевого» продукта является информация, во всех возможных видах и формах; в связи с этим закрепляется и новое название – информационная экономика.

4. Для Украины крайне актуальным является переход на стратегии инновационного развития, на строительство своей экономики знаний. Несмотря на трудности затянувшегося кризиса, на социальные и экономические потери периода «реформ», – социально-экономический потенциал страны и, в частности, ряда ее регионов достаточен для организации такого развития.

Приведенные соображения, будучи развернуты и детализированы, могут представлять собой исходные послышки для формирования соответствующей инновационной концепции, в том числе – региональной, – для создания в регионе структурных и «средовых» компонент экономики знаний.

Для развития экономики, для развития социума необходимо создание дополнительной стоимости. На протяжении прежних стадий развития экономики (заметим – при разных формациях!) дополнительная стоимость создавалась в массовом виде за счет товара, затем, для ряда стран – за счет качества товара. В последние десятилетия в передовых капиталистических странах вперед выдвигаются такие факторы формирования дополнительной стоимости, как качество услуг и качество менеджмента. Наконец, в информационной экономике ведущим фактором создания дополнительной стоимости становятся знания, информация. На наших глазах происходит формирование такой стоимости, которая как политэкономическая категория делается все более нематериальной.

Явление это имеет глубокие корни, оно связано не с той или иной общественно-политической формацией, а с общим социальным и научно-техническим прогрессом. Да, становление экономики знаний мы наблюдаем в развитых капиталистических странах, однако рыночная система является формой, хорошей формой реализации, – **содержанием являются коренные изменения в обществе, в производстве, в распределении и потреблении, в содержании работы как таковой, в моделях ведения бизнеса.** Общество научилось использовать то самое качество человека, которым он отличается от всех остальных существ на Земле, – **понимание и умение оперировать символами** [2]. Эта мощная компонента социально-экономического потенциала общества стала в массовом порядке

актуализироваться. Результатом и явилось то, что дополнительная стоимость начала возникать в процессе генерирования и «промышленного» использования **знаний**.

Инновационная деятельность имела место во все времена, больше или меньше, – и при господстве сельскохозяйственного сектора, и в индустриальной экономике. Однако в настоящее время, в период третьей и четвертой «тоффлеровской волны» инновационная деятельность приобрела новые черты. К ним относятся:

- массовый характер инноваций;
- явная информационная выраженность деятельности;
- не обязательное наличие собственного высокого технологического уровня (примеры – Ирландия, Малайзия, Финляндия и др.);
- возникновение глубоких положительных обратных связей в отраслевых и территориальных инновационных процессах (развитие инновационной деятельности в одних направлениях стимулирует таковую в других);
- опора на **средовую** ситуацию, **задающую восприимчивость** инноваций.

Учет этих факторов для Украины представляется очень важным.

Несколько лет назад Организация Экономического Сотрудничества и Развития (ОЭСР) опубликовала отчет «Map 2001 – Towards a Knowledgebased Economy», в котором перечислены основные черты экономики стран из первой десятки по суммарной величине капиталовложений в индустрию знаний и формирование интеллектуального капитала данной территории. Оказалось, что отличительными характеристиками этих стран являются:

- в структуре активов в экономике данной страны увеличивается доля нематериальных составляющих – патентов, ноу-хау, информационных/телекоммуникационных систем и сетей;
- происходит увеличение доли высокотехнологичной продукции в объеме международной торговли в расчете на 1 человека (для стран с низким среднедушевым доходом – 5%, с высоким – 20% и выше);
- имеет место рост «знание-емкости» ВВП (увеличение в нем «мягкого» продукта);
- рост расходов на НИОКР;
- рост инвестиций, в первую очередь, бюджетных, – на науку, образование, здравоохранение, окружающую среду.

По сути дела, это – ориентиры для правительств, которые выбирают путь строительства у себя экономики знаний. Практически все эти критерии могут быть использованы региональными властями.

Еще два замечания об описании и оценке инновационного (интеллектуального) потенциала региона. Первое относится к необходимости учета его глубокой, генетической связи с трудовым потенциалом региона, поскольку в живой действительности, в жизнедеятельности общества инновационный потенциал выражается через общественный труд, через приобретение знаний, через научную работу. Это значит, что чем выше трудовой потенциал (высокий образовательный уровень населения, значительная доля работников промышленности, строительства, транспорта с высокими разрядами, наличие квалифицированного менеджмента и пр.), – тем выше возможности инновационного развития. И наоборот, – если трудовой потенциал низкий, – при организации инновационной деятельности будут возникать помехи и затруднения.

Второе замечание относится к необходимости использования качественных характеристик наряду с количественными и, кроме того, к учету соответствия тех или иных проектов специфике региона и перечню собственно региональных приоритетов. Лишь объединение этих подходов может дать объективное представление об инновационном потенциале любого региона. Обсчитывать, количественно измерять можно лишь ту его часть, которая, во-первых, вообще поддается такой оценке, а во-вторых,

отображается в региональной статистике и включает в себя уровень общей образованности, высшее образование (не число институтов, а число жителей с высшим образованием), квалификационные характеристики научных работников, работников просвещения, культуры, творческой интеллигенции.

2. Инновационная деятельность и территориальное развитие

Под инновационной деятельностью, как это принято, понимается совокупность процессов зарождения новой идеи, генерирования и реализации новшеств и их встраивания в реальную жизнь, в действительность. Таким образом, эта деятельность требует *специфических знаний о той среде, в которой данное новшество должно быть реализовано.*

Это означает, что в данном регионе:

- а) должен регулярно проводиться *всесторонний маркетинг*, внутренний и внешний, с мощным информационным оснащением и оповещением;
- б) должна существовать, быть признанной и соблюдаться система промышленных, экологических и социальных стандартов;
- в) должен иметь место развитый *комплекс гуманитарных технологий* (обучение, консалтинг, аудит, информ-сфера, реклама, public relations и т.д.).

Тогда будут обеспечены комфортные, *естественные условия для реализации инноваций, а не «внедрение».* Иными словами, речь идет о создании системы, которая обеспечивает не просто генерирование инноваций, а *среду для их восприятия, мало того, – нуждающуюся в них.*

Для Причерноморского региона Украины, например, это сложная, но посильная задача. Ниже очень кратко перечислены положительные и отрицательные региональные факторы, в большой мере определяющие возможность формирования информационной экономики в указанном регионе.

Табл. 1

РЕГИОНАЛЬНЫЕ ФАКТОРЫ ВОЗДЕЙСТВИЯ НА ФОРМИРОВАНИЕ ЭКОНОМИКИ ЗНАНИЙ	
<i>Негативные и тормозящие факторы</i>	<i>Позитивные, стимулирующие факторы</i>
Разрозненность – нет реальных кристаллизующих институций	Возможность формирования стратегических целей развития региона, поддержка региональных властей
Отсутствие работающей инфраструктуры, средовой компоненты	Наличие не до конца утерянного научно-технического потенциала и производственных традиций
Отсутствие начального инновационного капитала	Определенная законодательная основа
Отсутствует масштабный инновационный маркетинг (в интересах региона), и, соответственно, – информация	Наличие в регионе информационно-коммуникационной основы и собственных технологий
Слабая восприимчивость хозяйственного комплекса	Общий достаточно высокий уровень образования и культуры

Для развития и поощрения региональной активности в формировании экономики знаний целесообразно выделить определенные индикаторы, критерии и показатели. Европейский опыт подсказывает в этом отношении целую гамму средств. Шведское организационно-информационное агентство ISA ("Invest in Sweden" Agency – Агентство «Инвестируйте в Швецию») опубликовало свой перечень критериев, в частности, для демонстрации преимуществ своей страны. Ниже с некоторыми модификациями приведен этот перечень.

Табл. 2

Области и направления деятельности	Рекомендуемые индикаторы
1. Генерирование и коммерциализация инноваций («инновационность» региональной действительности)	<ul style="list-style-type: none"> – объем НИОКР, их доля в валовом региональном продукте; – расходы на науку, их доля в ВРП; – число и структура научных и научно-технических публикаций; – внутренний рынок обмена инновациями
2. Создание знаний с коммерческим потенциалом	<ul style="list-style-type: none"> – расходы на прикладные НИОКР; – их доля в ВРП; – число патентов, зарегистрированных в «тройке» (Европа, США, Япония), на 1 млн. жителей региона;
3. Инновации (изобретения, патенты, рацпредложения) в промышленности и в хозяйственном комплексе	<ul style="list-style-type: none"> – доля (и структура!) промышленных предприятий, <ul style="list-style-type: none"> а) генерирующих и б) использующих инновации: по отраслям, объемам, типам
4. Качество интеллектуального потенциала и человеческого капитала	<ul style="list-style-type: none"> – структурные и внешние (выходные) показатели системы образования; – вся наука (доля научных работников в общей численности занятых, количество и качество работ, структура работников...); – миграция (+, –) высококвалифицированных работников;
5. Информационно-коммуникационные и компьютерные технологии	<ul style="list-style-type: none"> – показатели сети, локальных сетей, информационных систем; – он-лайн-овые государственные веб-узлы; – стоимость сообщений и объем обмена;
6. Предпринимательская деятельность	<ul style="list-style-type: none"> – количество создаваемых частных компаний; – удельный вес малого и среднего бизнеса в структуре хозяйств. комплекса (в деталях); – число предприятий-промоторов; – число венчурных фирм;
7. Поддержка инноваций и изобретений со стороны властей	<ul style="list-style-type: none"> – доля государственных расходов на НИОКР в объеме ВРП; – доля предоставленного венчурного финансирования в ВРП;
8. Емкость рынка знаний и динамика международного обмена знаниями	<ul style="list-style-type: none"> – количество и объем проданных за рубеж патентов и НИОКР; – количество и объем закупленных за рубежом патентов; – характеристики международного «безвалютного» обмена; – обмен студентами, аспирантами, докторантами;

Данная таблица (в приведенном или модифицированном виде) может служить материалом для формирования, в том числе – на региональном уровне, *регуляторных мер по развитию инновационного потенциала территории* и, как видим, – для укрепления конкурентоспособности региона. Разрабатывая это направление, мы в состоянии получить, по крайней мере, две важных характеристики инновационного потенциала: структурную (по предметным сферам) и атрибутивную, качественную. Это и означает конкретное продвижение к ответам на вопрос: как идеи и знания (и какие именно) превратить в общественное благосостояние и деньги? Применительно к социально-экономическому потенциалу и такой

его компоненте, как инновационный потенциал, – возникают задачи измерения (см., например, вышеприведенную таблицу) и оценки, т. е. мониторинга. Для этих работ с успехом может быть использована модель, описанная в [3].

В качестве конкретного направления, актуального для Причерноморского региона и для страны в целом, рассмотрим вкратце вопросы развития и внедрения экологически чистых технологий. Эта область чрезвычайно интересна и в методологическом отношении, так как на ее примере можно показать взаимосвязанность различных составляющих социально-экономического потенциала и, соответственно, значимость системных, комплексных решений.

3. О взаимосвязи чистых технологий, общих объемов выбросов и материально-финансовых характеристик населения

Не вдаваясь в технические, экономико-экологические, организационно-правовые вопросы этой проблемы, отметим лишь некоторые важные для нас аспекты.

К мощным социально-экономическим результатам применения чистых технологий следует отнести тот факт, что они являются не только *мультипликатором* в системе регионального и национального социально-экономического развития, но и *акселератором* в целом ряде отраслей и сфер жизни.

Сильный мультипликативный эффект высокой технической и технологической культуры проявляется в одновременном уменьшении потребления невозобновляемых ресурсов и нагрузки на среду. Лишь одна из сторон применения чистых технологий состоит в том, что уходит меньше топлива, меньше материалов, дольше и надежнее эксплуатация произведенной продукции. Действует так называемый Factor 4 (умножение на 4): «затрат – половина, выгода – двойная», – один из последних докладов Римскому клубу. (При этом следует помнить, что просто «применение в производстве чистых технологий» еще само по себе не является гарантией уменьшения нагрузок на окружающую среду, на хозяйственный комплекс в целом). Известен пример с персональными компьютерами: микроминиатюризация схем и автоматизация изготовления и сборки способствовали удельному уменьшению потребляемой энергии и стоимости; в итоге выросли спрос и объем продаж; суммарные объемы потребления энергии увеличились. Имеем дело с характерным примером сложности социально-экономических процессов и неоднозначности принимаемых решений и действий.

Кроме собственно материально-производственной сферы, чистые технологии находят все большее применение в сфере услуг, в здравоохранении, в образовании, в менеджменте, в предпринимательской деятельности, в социальных практиках. Каждое из этих направлений – очень перспективная и результативная сфера применения чистых технологий, каждое направление базируется на важных компонентах социально-экономического потенциала и ощутимо совершенствует его, каждое заслуживает специальной разработки. Отметим, что механизм внедрения чистых технологий, как и другие сложные механизмы, работает одновременно и на *формирование* социально-экономического потенциала территории, и на его *реализацию*.

Особое место в этом занимает выросшая из логистики культура деловых и, шире, гражданских отношений (английское выражение *culture to deliver* на русском языке обозначает по смыслу взаимную высокую ответственность и обязательность). Если говорить не только о деловой, но и о социальной ответственности и обязательности, при этом – реализуемой всеми фигурантами региональной жизнедеятельности, местным и национальным управлением, населением и предпринимателями, – то переоценить эти новые социальные технологии – невозможно.

Акселератором чистые технологии являются по отношению к ряду направлений. Достаточно привести пример с социально-трудовыми отношениями, – чистые технологии требуют создания высококвалифицированных рабочих мест, с соответствующими оплатой труда и уровнем знаний у работников (все это – основные компоненты и характеристики продуктивной занятости). А это, в свою

очередь, предъявляет высокие требования к образованию, формирует высокие зарплаты, повышение платежеспособного спроса, повышение уровня и качества жизни.

Отдельного рассмотрения заслуживает постоянная и многофункциональная положительная обратная связь между разработкой и внедрением чистых технологий, с одной стороны, и наукой, разными ее направлениями, – с другой; особенно – в контексте нашего рассмотрения проблемы эффективного формирования инновационного потенциала региона.

4.Трехмерное представление экологической кривой Кузнеця (ЭКК).

Наиболее интересным для нас аспектом является выявление связи между объемами выбросов, материально финансовыми характеристиками населения и расширением экологически чистых производств, далее – как эти связи и взаимовлияния выглядят в динамике, а также – как эта динамика связана (зависит, взаимодействует) с социально-экономическим потенциалом территории.

Достаточно давно известна так называемая Экологическая Кривая С. Кузнеця (ЭКК). Построенная, вообще говоря, для других целей, она показывает, как зависит в стране или регионе общий объем загрязнений (выбросов, отходов и т. д.) от *среднедушевого дохода граждан* или ВРП (кривая **A** на рис. 1 в плоскости $P - M$). Сперва с увеличением этого дохода объем загрязнений растет и достигает определенного максимума (точка P_1). Затем денег в обществе (больше $m_{кр}$) оказывается достаточно для более качественной очистки, расширения природоохранной деятельности, новых средств контроля и т. д. Естественно, объем загрязнений при этом начинает уменьшаться. Высказанные ранее в качестве гипотезы, эти соображения позднее получили эмпирическое подтверждение [4].

Упомянутая модель, будучи представленной на плоскости «деньги – выбросы», имела двумерное изображение.

Введем в рассмотрение еще одну координату (см. рис. 1). По этой третьей оси K (*knowledge*) будем откладывать обобщенную агрегированную величину, характеризующую объем знаний в обществе: развитие чистых технологий, рост инновационной деятельности, новые знания производителей, более высокую грамотность потребителей и управленцев, – данные, приведенные в Табл. 2. Таким образом производится увеличение размерности модели за счет введения измерения, описывающего уровень информационной, «знаниевой» насыщенности данной социально-экономической системы.

На рис. 1 изображены три модельных варианта кривых A_i , построенные для разных k_i .

Из рисунка видно, что при условии $k_3 > k_2 > k_1$, мы получим:

$$p_3 < p_2 < p_1.$$

Это означает, что можно ожидать последовательного падения ординат p_i , то есть уменьшения суммарного максимального объема выбросов.

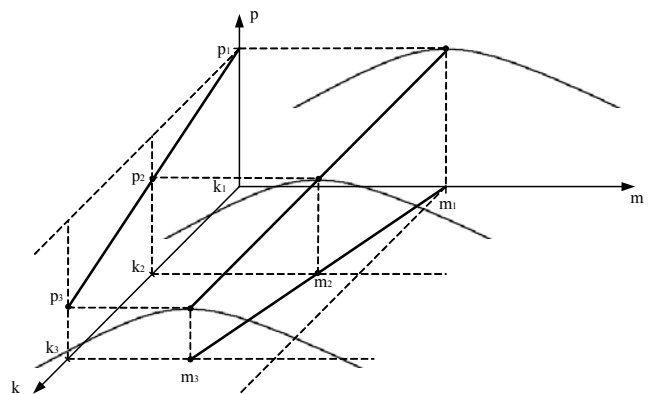


Рис. 1.

Построенное на рис. 1 условное изображение связи трех параметров (информационных, социальных, экологических) назовем ИСЭ-пирамидой.

Приведенное здесь развитие представлений, получаемых из анализа ЭКК, изображение картины в объеме, учитывающем рост экономики знаний, является еще одной системной демонстрацией значимости использования чистых технологий, – при расширении их внедрения будет уменьшаться давление на внешнюю среду.

Но этого мало; из графика видно, что будет иметь место еще одна динамика: при росте знаний уменьшается не только ордината p_{max} . Максимальные критические значения объемов загрязнений, как можно ожидать, будут возникать при все меньших значениях душевого дохода (последовательно уменьшается величина абсциссы $t_{кр}$), что является крайне актуальным для сегодняшней ситуации на Украине.

Проведенные рассуждения можно подытожить следующим образом. На базе полученной трехмерной модели связи между общим объемом загрязнений в регионе, величиной среднедушевого дохода населения и насыщенностью экономики знаниями сформулирована гипотеза:

- а) об уменьшении общего объема антропогенных загрязнений с ростом знание-емкости хозяйственного комплекса;
- б) об уменьшении значений среднедушевого дохода населения, при которых наступает критическая (максимальная) величина загрязнений.

Можно думать, что оба эти вывода имеют очень важное значение при анализе социально-экономического потенциала региона в сложившихся условиях и при выборе управляющих воздействий.

Еще один естественный вывод – о результативности объемного представления: ведь на третьей оси можно откладывать различные важные характеристики, – продуктивную занятость, культуру управления, степень коррумпированности общества (в этом случае, по-видимому, получим не пирамиду, а расширяющуюся воронку, – при большей коррумпированности системы нужно больше денег, увеличивается нагрузка на природную среду), и т. д. Подобная многомерная модель может стать весьма эффективным инструментом анализа и управления. Добавим, что такая модель иллюстрирует также роль и значимость различных работ, представляемых на KDS, а также роль и значимость движения KDS в целом.

Библиография

1. Шумпетер Й. Теория экономического развития, ч. 1. – М.: «Прогресс», 1987.
2. Кастельс М. Информационная эпоха: экономика, общество и культура. – М.: ГУ ВШЭ, 2000.
3. Крисилов А., Крисилов В. Формирование целеориентированной векторной модели для построения агрегированных оценок сложных объектов. // Социально-экономич. проблемы информационного общества. Под ред. проф. Л. Мельника. – Сумы: Унив. Книга, 2005
4. Брюханов М. Проблемы экологизации экономического развития при переходе к информационному обществу. // Социально-экономические проблемы информационного общества. Под ред. проф. Л. Мельника. – Сумы: Унив. Книга, 2005

Информация об авторах

Крисилов Анатолий Данилович, к. т. н., доцент Международного Гуманитарного Университета, каф. экономики и моделирования экономических процессов; научн. руководитель Бюро инновационных процессов; ул. Педагогич., 22/40, Одесса-63, 65063, Украина; e-mail: <adkrissilov@list.ru>

Крисилов Виктор Анатольевич, д. т. н., проф., зав. кафедрой системного программного обеспечения Одесского национального политехнического университета; просп. Шевченко, 1, Одесса-44, 65044, Украина; e-mail: <victork@405.com.ua>

ОСНОВНЫЕ АСПЕКТЫ МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ УПРАВЛЯЕМОГО ДИАГНОСТИЧЕСКОГО ПРОЦЕССА В РАМКАХ ЕДИНОГО ИНФОРМАЦИОННОГО ПРОСТРАНСТВА СИСТЕМЫ ЗДРАВООХРАНЕНИЯ

Игорь Долгополов, Наталия Семенова

Резюме: Рассматривается задача математического моделирования управления процессами клинической диагностики, которая описывается векторной задачей дискретной оптимизации в условиях неопределенности данных. Исследуются возможности объединения индуктивного обобщения и дедуктивного вывода в рамках интеллектуальной системы управления процессами клинической диагностики в объеме целенаправленных диагностических систем с использованием методов распознавания образов и многокритериальной дискретной оптимизации.

Ключевые слова: распознавание образов, классификация, клиническая диагностика, средства диагностических технологий, аналитические технологии, многокритериальные задачи, дискретная оптимизация.

Введение

Традиционно в технических системах распознавания считалось, что диагноз, как решение о принадлежности распознаваемого объекта ω соответствующему классу Ω_i , $i = 1, \dots, m$, принимается только после определения всей совокупности признаков этого объекта x_1, \dots, x_n . Однако в практике клинической медицины возможен другой подход. Задача распознавания должна решаться на основании информации об измеренных к текущему моменту клинической диагностики (КД) каждого очередного признака контрольного объекта: x_1, x_2 и т. д. При этом процесс набора дополнительной информации путем измерения последующего признака объекта или прекращение этого процесса происходит в зависимости от результатов полученного решения на каждом этапе исследований. Такая последовательная процедура решения задачи распознавания обязана своим возникновением одному из разделов статистики — последовательному анализу [2]. Технология последовательного и многократного решения задачи распознавания с использованием на каждом этапе исследований возрастающего числа измеренных признаков актуальна для клинической медицины, особенно когда для определения признаков используются средства диагностических технологий (СДТ), а сам процесс КД связан с определенным риском, и, что принципиально важно — это когда ряд заболеваний надежно распознается по неполному или ограниченному набору признаков.

Обозначим множество исследований, проводимых в рамках диагностических мероприятий, $A = \{a\}$. Назначить исследование a — это значит указать, какой признак и с помощью какого метода СДТ его необходимо определить. Каждое исследование имеет определенный исход: наличие или отсутствие признака, определение его числового значения, интервала и т. д. В процессе КД необходимо дифференцировать множество промежуточных $x = \{x_a\}$ (здесь x_a — общее обозначение исхода исследования) и финишных исходов $F_{ds} = x_1 + x_2 + \dots + x_n$.

На проведение процесса клинических исследований накладываются определенные ограничения, обусловленные рядом обстоятельств. Это, прежде всего, их стоимость, включающая временные и трудовые затраты, амортизация средств диагностических и телекоммуникационных технологий и др. Сумма затрат определяется объемом разделенных во времени и пространстве диагностических

мероприятий, неоднородностью их инструментальной базы, что часто ведет к дублированию данных, потере валидности и в связи с этим — необходимости поиска информации в распределенной базе данных или хранилище данных, в зависимости от типа архитектуры медицинской сети региона и др. При этом стоимость определения выявляемого класса состояния, в объеме достоверного диагноза, не должна превышать установленной величины. Таким образом, на множество $A = \{a\}$ накладывается система последовательных ограничений H , заданная следующим образом. Если для каждой системы H — допустима последовательность исходов $x_{a_1}, x_{a_2}, \dots, x_{a_k}$, в которой $a_1 \in A_1^G$, $a_2 \in A_2^H(x_{a_1}), \dots$, $a_k \in A_k^H(x_{a_1}, \dots, x_{a_{k-1}})$, то это означает, что определено множество исследований $(k+1)$ -й стадии $A_{k+1}(x_{a_1}, \dots, x_{a_k})$, с получением цепочки исходов x_{a_1}, \dots, x_{a_k} в результате исследований a_1, \dots, a_k . Совокупность диагностических исследований A с заданной системой ограничений H обозначим A^H .

В результате нескольких итераций инструментальных исследований полученная информация используется для принятия решения о принадлежности диагностируемого объекта одному из возможных классов состояний. Обозначим $z = \{z_1\}$ как множество промежуточных решений. Оно распадается на подмножества $z = \{z_1^k\}$, элементы которого z_1^k означают, что после проведения k стадий исследований принято промежуточное решение о принадлежности объекта ω к Ω_i^A -му классу, т.е. установлен предварительный диагноз, с некоторым упорядоченным множеством гипотез правдоподобных диагнозов $G_1 \in G$. Каждая гипотеза из множества G отличается от конечного решения — клинического диагноза DS — объемом недостающей информации ΔI . Порядок формирования гипотез G_1^k о диагнозе $DS_i(I_{V_i})$, $i = \overline{1, p}$, и предельная степень их конкретизации зависят от апостериорной, т.е. полученной после исследований информации I_V . В данном процессе выполняется соотношение $DS_1(I_V) \subset DS_2(I_{V-1}) \subset \dots \subset DS_p(I_{V_p})$, т.е. каждое подмножество правдоподобного диагноза $DS_p(I_{V_p})$ содержится во множестве $DS_i(I_{V_i})$, $i = \overline{1, p-1}$, и следует из него. Дальнейшую конкретизацию DS за счет сужения $DS_p(I_{V_p})$ можно осуществить с помощью дополнительной информации в процессе диагностических исследований. Для этого необходимо определить такой признак x_k , $k = 1, \dots, N$, измерение которого внесет в систему распознавания DS наибольшее количество информации, с учетом того, что на предыдущем этапе был определен признак x_l , т.е. $I(x_k / x_l = \max_j Ix_j / x_l)$.

Принятие окончательных решений в КД всегда сопряжено с определенной их стоимостью. Последовательное проведение дальнейших исследований, как и принятие окончательного решения о принадлежности объекта ω определенному классу A , на основании информации, полученной в результате этих исследований, сопряжено с наиболее чувствительными потерями для КД — времени, расходами трудового и ресурсного обеспечения и др. Если проведенные исследования a_1, \dots, a_k , завершились исходами x_{a_1}, \dots, x_{a_k} и принято окончательное решение z_1^k , то будем полагать, что значение стоимости диагностической траектории такого решения равно $C_\omega[z_1^k(x_{a_1}, \dots, x_{a_k})]$.

Значение расходов U_ω , дальнейшей диагностической траектории формирования признакового пространства приоритетной гипотезы (ΔI), усредненное по возможным ее вариантам определяется

правилом R , в соответствии с которым планируется последовательность и объем диагностических мероприятий, т.е. $\bar{U}_\omega = \bar{U}_\omega(R)$. Однако каждое правило R может строиться лишь с учетом ограничений H , накладываемых на мероприятия проводимых исследований. Поскольку заранее не известно, какой в информационном поле объект подвергается распознаванию, величина U_ω должна быть усреднена с помощью априорной вероятности появления объекта $P(\Omega_i)$ [3].

Качество каждого алгоритма, определяющего последовательное правило R , в соответствии с которым реализуется процесс управления, можно охарактеризовать функционалом, представляющим собой математическое ожидание от значения средних расходов:

$$U_p(R) = M[\bar{U}_\omega(R)] = \sum_{i=1}^m \bar{U}_\omega(R) P(\Omega_{ds}).$$

Требуется определить оптимальное правило $R \in R^H$, обеспечивающее минимум функционала, т.е. минимизацию математического ожидания расходов, связанных с реализацией процесса управления процессами КД. Данную задачу можно рассматривать в виде двухкритериальной задачи. В соответствии с подходом, используемым в финансовом анализе, будем искать решения Парето-оптимальные по двум критериям: минимуму среднего значения целевой функции и минимуму ее дисперсии. В работах [5-7] исследованы некоторые свойства различных видов эффективных решений и предложены методы поиска решений указанных задач.

Рассмотренные понятия позволяют построить метод управления процессами КД в виде правила последовательного поиска решений, обеспечивающего разработку оптимального плана проведения диагностических мероприятий. Суть оптимальности такого метода заключается в том, что каждый шаг диагностических мероприятий включает в себя следующий объем процедур: установление признаков; выбор метода и средства его регистрации; планирование исследования состояния больного; контроль его эффективности, которые определяются информацией, полученной в результате предыдущих исследований, как в рамках текущего диагностического процесса, так и всей предысторией настоящего состояния обследуемого. Это позволяет определить, какие признаки необходимо актуализировать и какие средства диагностических технологий (СДТ) использовать на каждом шаге мероприятий КД [4].

Выводы

Рассматривается задача математического моделирования управления процессами клинической диагностики, которая описывается векторной задачей дискретной оптимизации в условиях неопределенности данных. Исследуются возможности объединения индуктивного обобщения и дедуктивного вывода в рамках интеллектуальной системы управления процессами клинической диагностики в объеме целенаправленных диагностических систем с использованием методов распознавания образов и многокритериальной дискретной оптимизации. На основе предложенного математического аппарата показана принципиальная возможность создания медицинских диагностических информационных систем нового класса целенаправленного действия

Библиография

- [1] Рекомендации Европейского общества кардиологов по диагностике и лечению: www.Therapia.com.ua
- [2] Фу К. Последовательные методы в распознавании образов и обучении машин. - М.: Наука, 1971.
- [3] Горелик А.Л., Гуревич И.Б. Некоторые аспекты проблемы распознавания образов. - М.: Радио и связь, 1985. - 159 с.

-
- [4] Долгополов И.Н. Технология фрейма в проектировании интеллектуальных систем автоматизированной диагностики // Кибернетика и системный анализ. - 1998. - № 2. - С. 149–161.
- [5] Семенова Н.В. Методы поиска гарантирующих и оптимистических решений задач целочисленной оптимизации в условиях неопределенности данных // Кибернетика и сист. анализ. – 2007.– №1.– С. 103-114.
- [6] Лебедева Т.Т., Семенова Н.В., Сергиенко Т.И. Устойчивость векторных задач целочисленной оптимизации: взаимосвязь с устойчивостью множеств оптимальных и неоптимальных решений // Кибернетика и сист. анализ. – 2005.– №4.– С. 90-100.
- [7] Лебедева Т.Т., Семенова Н.В., Сергієнко Т.І. Умови оптимальності та розв'язуваності в задачах лінійної векторної оптимізації з опуклою допустимою множиною // Доповіді НАНУ. – 2003. – №10. – С. 80-85.
-

Информация об авторах

Игорь Николаевич Долгополов – Международный научно – учебный центр информационных технологий и систем ЮНЕСКО/МПИ, 03680, проспект академика Глушкова, 40, Украина; e-mail: dolgigo@ukr.net

Наталья Владимировна Семенова – Институт кибернетики им. В.М.Глушкова НАН Украины, старший научный сотрудник, 03680 МСП Киев 187, проспект академика Глушкова, 40, Украина; e-mail: nvsemenova@meta.ua

МЕТОДЫ МОДИФИКАЦИИ ОПТИМИЗАЦИОННЫХ МОДЕЛЕЙ С ЦЕЛЮ УВЕЛИЧЕНИЯ ДИАПАЗОНА ВЫБОРА ВАРИАНТОВ ПРИНЯТИЯ УПРАВЛЕНЧЕСКИХ РЕШЕНИЙ

Николай Кузубов

Аннотация: В докладе для оценки экономико-математических моделей в процессе выбора возможных вариантов функционирования экономической системы предложено использовать показатель вариантности моделей и на его основе выработки наиболее рационального направления развития системы. Использование этого показателя при совершенствовании моделей позволяет вести целенаправленный поиск методов их модификации. Положительный эффект дает лишь такая модификация, которая сопровождается повышением коэффициента вариантности моделей. Рассмотрены методы модификации моделей, которые основаны на принципе введения в модель простейших (маргинальных) переменных, отражающих возможность увеличения объема производства продукции каждым отдельно взятым видом экономической деятельности на макроуровне или отдельной отрасли производства – на предприятии.

Ключевые слова: Показатель вариантности, диапазон развития, экономическая система, метод модификации.

Введение

На современном этапе развития вычислительных средств и научных прикладных разработок вряд ли кто-то сможет отрицать, что наиболее надежным инструментом выработки научно обоснованного управленческого решения в сфере экономики являются методы математического моделирования. На макроуровне успешно используется модель межотраслевого баланса экономики страны или крупного

региона, на микро уровне – наиболее эффективны, в частности в аграрном секторе экономики, модели оптимизации отраслевой структуры производства предприятий.

Поиск наиболее эффективных вариантов развития экономической системы по модели межотраслевого баланса осуществляется, как известно, в диалоговом режиме. При этом нельзя гарантировать, что выбранный вариант развития данной экономической системы окажется наиболее рациональным, поскольку в модели отсутствует критерий оптимизации выбора.

Исследования автором известных моделей оптимизации отраслевой структуры производства сельскохозяйственных предприятий показали, что в названных моделях число технологических способов (переменных модели), различающихся между собой уровнем затрат ресурсов и выхода продукции конкретной отрасли ограничивалось в лучшем случае двумя переменными. Одна из них обозначала культуру, продукция которой используется на товарные цели, другая – на корма.

Таким образом, можно констатировать, что эти модели не отвечают одному из основных требований к таким моделям. Это требование сформулировано основоположником прикладного использования математических методов в экономических исследованиях Л.В.Канторовичем [Л.В.Канторович, 1960]. Он отмечает, что в «модели число технологических способов (*т.е. переменных различающихся между собой уровнем производственных затрат и выхода продукции* (курсив авт.)) или вариантов технологий, должно быть весьма велико».

Однако, во-первых, если это условие выполнимо для моделей производственного типа, то практически невозможно в модели межотраслевого баланса, поскольку не существует информации, которая характеризует различные уровни развития одной определенной отрасли.

С другой стороны, нельзя априорно определить количество технологических способов достаточное для обеспечения вариантности модели.

И наконец, следует подчеркнуть, что в понятие вариантности часто вкладывается различное толкование.

Именно этим вопросам автор уделил особое внимание в данной публикации.

Диапазон и вариантность развития экономической системы или процесса

Чаще всего «когда речь идет о вариантности экономико-математической модели, то имеется в виду наличие в модели переменных, обозначающих варианты развития экономической процессов, изменение которых носит дискретный характер» [А.В.Крушевский, 1982]. Например, варианты производственной мощности того или иного предприятия или наличие нескольких поставщиков товаров для определенной группы потребителей. Такое содержание понятия вариантности моделей, вполне очевидно, может быть обнаружено путем простого логического анализа и не требует каких либо дополнительных исследований.

С позиции математики вариантность экономико-математической модели (ЭММ) вычисляется по количеству в ней ограничений и переменных [Г.Вагнер, 1972]. И совершенно очевидно, что в этом случае речь идет лишь о возможном количестве вариантов решения данной модели.

В реальной практике большая часть процессов, включая экономические, имеет непрерывный характер развития. Например, на предприятии может быть произведено определенное количество узлов, агрегатов или единиц техники, на сельскохозяйственном предприятии – какое-либо количество определенного вида продукции. При этом предельный уровень выпуска продукции ограничивается на предприятии его мощностью и наличием производственных ресурсов, включая рабочую силу. В сельскохозяйственном предприятии, кроме того, наличием и качеством земельных ресурсов. Минимальный уровень выпуска в экономико-математических моделях обеспечивается ограничениями, предусматривающими этот уровень.

Таким образом, логично вытекает вывод, что ЭММ, отражающая условия функционирования экономической системы, имеет определенные пределы или диапазон развития, ограниченные ее минимальным и максимальным уровнем развития. В этом диапазоне находятся все возможные

варианты развития данной экономической системы. Совершенно очевидно также, что это есть характеристика области допустимых решений ЭММ.

Диапазон развития любой экономической системы (процесса) имеет числовую характеристику, как в абсолютных единицах измерения (разность между максимальным и минимальным значениями показателей уровня развития данной системы), так и в относительных (отношение этой разницы к максимальной величине данного показателя). Относительные показатели, характеризующие диапазон выбора вариантов развития экономической системы или процесса, приняты нами в качестве показателей их вариантности [Н.Кузубов, 1986].

В ЭММ показатель вариантности модели (K) можно получить при ее решении на максимум ($F \rightarrow \max$) и минимум ($F \rightarrow \min$) выбранной целевой функции. Формула его вычисления имеет вид:

$$K = \frac{F \max - F \min}{F \max}.$$

По нашему мнению, есть необходимость выполнить логический анализ параметров показателя вариантности ЭММ и определить наиболее целесообразное его значение. В соответствии в приведенной выше формулой, показатель вариантности ЭММ может приобрести значение близкое или равное единице. Т.е. модель будет обладать максимальной вариантностью, если при решении модели на минимум целевой функции ее значение будет приближаться к нулю. В моделях производственного типа такая ситуация может возникнуть, когда задания по гарантированному объему производства некоторой совокупности продукции будут полностью отсутствовать, поскольку именно эти задания определяют нижний (минимальный) предел развития данной экономической системы.

В реальной действительности такая ситуация не может иметь места и чаще бывает наоборот – задания по гарантированному объему производства продукции в экономической системе превышает ее ресурсные возможности, т.е. условия модели являются несовместимыми, полностью отсутствует область допустимых решений.

При условии, что ресурсов достаточно для выполнения гарантированных объемов производства продукции, показатель вариантности модели будет минимальным, если при ее решении на максимум и минимум значения целевой функции будут близки или совпадают, т.е. модель обладает незначительной или вообще нулевой вариантностью. По модели, не обладающей вариантностью можно получить заведомо известное решение, и таким образом, обосновать любое управленческое решение.

Из этого вытекает вывод, что прежде чем выполнить поиск оптимального варианта развития экономической системы (процесса) и обосновать, в результате решения модели направления ее перспективного развития, необходимо оценить вариантность ЭММ. При условии, что она окажется очень малой, возникает необходимость поиска методов ее модификации.

Модификация ЭММ макроуровня

Наибольший интерес, несомненно, представляет собой методы модификации моделей макроуровня. Исходной моделью экономики на макроуровне может служить модель межотраслевого баланса. В качестве объекта исследования выбрана модель межотраслевого баланса Украины*. Методом ее модификации избран предложенный Н.М. Губиным и др. [Н.Губин и др., 1993] способ включения в модель «приростных величин, характеризующих прирост функции, приходящийся на одну дополнительную «единицу» того или иного фактора». В нашем случае в модель межотраслевого баланса для каждого отдельно взятого вида экономической деятельности вместе с основными переменными, обозначающими эти виды деятельности, введены приростные (маргинальные) переменные. Эти переменные моделируют

* Данная часть исследования проведена совместно с аспиранткой Киевского национального университета имени Тараса Шевченко Единак Е.Н.

возможность увеличения объемов выпуска продукции каждым отдельно взятым видом экономической деятельности.

В качестве минимального уровня развития экономической системы (в данном случае экономики Украины) в модели принят уровень развития, достигнутый отраслей. Параметры возможного увеличения объемов выпуска продукции определялись при этом с учетом сложившихся темпов роста производства за последние годы.

Одним из ограничительных условий развития экономики Украины в модели является вектор-строка по суммарному уровню выбросов вредных веществ в атмосферу.

Приведем показатели вариантности этой модели. При условии, что темпы увеличения количества вредных выбросов будут соответствовать их фактически сложившимся темпам прироста, значения функции, полученные при решении на максимум и минимум ВВП (в ценах 2004 года) будут равны соответственно – 565,8 и 345,1 млрд. грн. Показатель вариантности этой модели при этом равен 0,39. По нашему мнению, это достаточно высокий показатель вариантности модели.

Кроме того, просчитан вариант, учитывающий возможность некоторого увеличения темпов прироста вредных выбросов. При этом удалось лишь незначительно до 0,396 увеличить показатель вариантности модели. В этом случае был достигнут предельный уровень развития всех видов экономической деятельности, предусмотренный ограничениями на возможный прирост объемов выпуска каждого вида продукции.

Модификация ЭММ микроуровня

Широкие возможности модификации ЭММ имеются при моделировании экономики предприятий, особенно сельскохозяйственных. В этих моделях вместо переменных обозначающих отрасли, продукция которых используется на товарные и кормовые цели, вводится ряд переменных. При этом одна из них является основной, технико-экономические коэффициенты которой фиксируют сложившийся уровень затрат ресурсов и выхода продукции в расчете на принятую единицу измерения переменной. Другие (их может быть несколько) фиксируют прирост продуктивности растениеводческих отраслей в зависимости от количества вносимых удобрений, расхода поливной воды и т.п. в отраслях растениеводства. В отраслях животноводства прирост продуктивности животных возрастает в зависимости от потребляемого корма в расчете на голову животных.

Отдельными переменными могут быть введены направления целевого использования продукции. В качестве таких переменных могут быть приняты различные каналы реализации продукции (заготовительным и перерабатывающим предприятиями, на рынке и т.п.). Такая модификация дает возможность наиболее адекватно отражать в модели условия функционирования предприятия в рыночной экономике.

Наши исследования аналогичных моделей [Н.Кузубов, 1986] показали, что эти методы модификации позволили повысить показатель вариантности модели оптимизации отраслевой структуры с 0,17 до 0,248. Особо следует отметить, что по сравнению с традиционной моделью оптимизации структуры производства предприятия, появляется возможность, кроме того, выбрать наиболее эффективный вариант технологии производства и существенно расширить диапазон выбора возможных вариантов функционирования предприятия.

Выводы

Исследования известных в литературе оптимизационных моделей функционирования экономических систем на макро- и микроуровне показало их несовершенство. Для их оценки предложено использовать показатель вариантности моделей, который позволяет оценить величину диапазона выбора вариантов

функционирования данной экономической системы в исходной модели и его (диапазона) изменения в случае модификации модели.

В качестве одного из возможных вариантов модификации оптимизационных ЭММ следует рассматривать метод их модификации путем включения в модель приростных (маргинальных) переменных, отражающих возможность увеличения объемов производства продукции каждым отдельно взятым видом экономической деятельности на макроуровне или отраслю предприятия – на микроуровне.

Литература

- [Л.В.Канторович, 1960] Канторович Л.В. Экономический расчет наилучшего использования ресурсов – Изд-во АН ССР, Москва, 1960.
- [А.В.Крушевский, 1982] Крушевский А.В. Справочник по экономико-математическим моделям и методам. – «Техника», Киев, 1982.
- [Г.Вагнер, 1972] Вагнер Г. Основы исследования операций Т.1. – «Мир», Москва, 1972.
- [Н.Кузубов, 1986] Кузубов Н. Показатель вариантности экономико-математических моделей и его практическое использование. – Экономика Советской Украины, № 4, 1986. – с. 72-76.
- [Н.М. Губин и др., 1993] Губин Н.М. и др. Экономико-математические методы и модели в планировании и управлении в отрасли связи: Учебник для вузов / Н.М.Губин, А.С.Добронравов, Б.С.Дорохов. – 3-е изд. доп. и перераб. – «Радио и связь», Москва, 1993.

Сведения об авторе

Николай Кузубов – доктор экономических наук, профессор кафедры Межрегионального института экономики и менеджмента Киевского славистического университета. 02154 Киев, ул. Энтузиастов, 7/2, 80, Украина; e-mail: knik2005@ukrnet.ua

ПРОБЛЕМЫ ФОРМИРОВАНИЯ ВНУТРИЛИЧНОСТНОГО КОНФЛИКТА КАК ФАКТОРА, ВЛИЯЮЩЕГО НА ПРИНЯТИЕ РЕШЕНИЙ

Елена Шинкаренко

Аннотация. *Рассматриваются проблемы, связанные с источниками возникновения бессознательных внутриличностных конфликтов и то, каким образом наличие данного фактора отражается на процессе принятия решений личностью.*

Ключевые слова: *сознательное, бессознательное, сверхсознательное, внутриличностный конфликт, комплексы, комплекс самоуничижения, принятие решения.*

Введение

В современном мире постоянно происходит формирование новых направлений, объединяющих различные области научного знания. Именно к таким направлениям принадлежат исследования, относящиеся к проблеме принятия решений. Принятие решения является центральным моментом на всех уровнях переработки информации человеком, группами людей, системами "человек – машина". Естественно, что эта сложная комплексная проблема включает различные аспекты: соматический, психологический, кибернетический и другие. Психологические аспекты проблемы связаны не только с анализом роли и места процессов принятия решения в системе целенаправленной сознательной деятельности человека, но и касаются его бессознательной сферы.

В данной работе мы рассмотрим некоторые проблемы, связанные с источниками возникновения бессознательных внутриличностных конфликтов и каким образом наличие данного фактора отражается на процессе принятия решений личностью.

Принятие решения как процесс перестройки стереотипов личности

Самые трудные как осознанные, так и бессознательные внутриличностные решения – те, которые принимаются, с одной стороны, в условиях дефицита времени, а с другой – в условиях информационной неопределенности и противоречивости, связанной как с внешними (социальными), так и внутренними (личностными) факторами. Эти решения служат главным критерием зрелости личности, её готовности нести ответственность за принятие решений. Смягчить трудность их принятия можно только в том случае, если они будут готовиться заранее, что в процессе жизнедеятельности не всегда возможно. Принять решение – означает «снять» противоречие назревшей проблемы, с которой личность отказывается в дальнейшем примиряться. Необходимость в его принятии возникает тогда, когда невозможна стереотипная, привычная реакция на полученную информацию и нужно действовать по-новому.

Но такого рода перестройка – нелегкая нервная работа, а иногда и серьезное испытание для человека. Понять причину этих трудностей помогают знания раскрытых И.П. Павловым закономерностей высшей нервной деятельности. Как известно, в коре головного мозга происходит непрерывное систематизированное деление раздражительных и тормозных состояний, которые в условиях повторяющихся обстоятельств фиксируются все легче и автоматизированнее. Так возникает динамический стереотип, реализация которого нуждается во всё меньшей затрате нервного труда. Установленный стереотип бывает тяжело изменить и привести в соответствие с новыми обстоятельствами, новыми требованиями жизни. Фиксируя определенный опыт и сохраняясь долгое время, динамические стереотипы обнаруживают склонность к крепкому удерживанию. Формируясь в процессе повседневного уравнивания организма со средой, жизненные стереотипы составляют в известной мере основу человеческой личности, обычно достаточно стабильную по форме и динамическую по содержанию.

Если динамический стереотип хорошо зафиксирован, он представляет собой достаточно консервативную систему и его отмена и выработка новой – тяжелая для нервной системы задача, которая может привести к временному нарушению высшей нервной деятельности. Выдержать такое испытание может только сильный нервный тип. Только такой личности под силу в какой-то степени преодолеть мощь собственных комплексов, т.е. группу взаимосвязанных сознательных и неосознаваемых идей и ощущений, которые оказывают динамическое влияние на поведение личности.

Комплексы как составляющее бессознательной сферы психики

Чтобы более четко понять сущность комплексов, их влияние на деятельность человека, необходимо коротко остановиться на понимании бессознательного, сознательного и сверхсознательного, а также их роли в психике личности.

Предлагаемая модель личности, принятая в психоанализе, представляет собой взаимодействия трех уровней, находящихся между собой в определенных соотношениях. Это – «Оно» (Id), т.е., глубинный уровень бессознательных влечений, своеобразный резервуар бессознательных иррациональных психических реакций и импульсов, биологических по своей природе. Это основа деятельности личности, та психическая инстанция, которая руководствуется своими собственными законами. «Оно» является единственным источником психической энергии и руководствуется только принципом удовольствия. К сфере бессознательного принадлежат инстинкты (пищевой, половой, самозащиты и др.), которые порождают у человека неосознанные желания, эмоции, влечения и позже могут попадать в область сознания.

Сознание же (сознательное – «Его») в отличие от бессознательного позволяет человеку осуществлять высший контроль над своими психическими процессами и поведением, направлять ход своей психической и наглядной деятельности в нужное русло, а также анализировать свое собственное сознание. Оно включает отношение человека и к самому себе, и к другим людям, то есть выступает как самосознание.

Наконец, «Сверх-Я (Super-Ego)» – это внутриличностная совесть, т.е., инстанция, олицетворяющая собой ценности и установки общества.

Лишь сравнительно небольшая часть психической деятельности осознается человеком, остальное остается неосознанным. Но грань между осознанным и неосознанным подвижна. Да и свои эмоции, интуитивные предположения человек со временем может логично осмыслить, разобраться в мотивах своих иногда необдуманных поступков. Напротив, с формированием автоматизмов происходит переход их в сферу бессознательного. Причем, в случае неосознанных актов поведения человек хоть и получает из внешнего мира необходимую информацию, но она не осознается, обрабатываясь и используя на самых разных уровнях центральной нервной системы.

Благодаря включению бессознательного уменьшается нагрузка на сознание и увеличиваются возможности человека в научном или художественном творчестве. Оно проявляется через рефлексy, инстинкты, привычные автоматизированные действия и нормы (по мере овладения ими осознание их снижается); стремления, чувства и поступки, причины которых не осознаются; ошибочные действия, выразительные движения, свободные ассоциации, сновидения, невротичные реакции и аффективные комплексы.

Эти проявления способны вызывать в психике человека, соответственно З.Фрейду, постоянный конфликт между уровнями личности («Super-Ego», «Его», «Id»). Поэтому для смягчения ситуации напряжения и вины, вызванных таким конфликтом, психика создала ряд защитных механизмов, призванных «подавлять, вытеснять» неприятные или запрещенные мысли, желания, ощущения, поступки, устранять их из области сознания в область бессознательной сферы психики, неосознанно ставить препятствия на пути той информации, которая является противоречием относительно требований совести, каких-то определенных моральных качеств личности и может ее травмировать.

Защитные психологические механизмы по теории психоанализа	
Рационализация	псевдоразумное объяснение человеком своих желаний, поступков, в действительности вызванных причинами, признание которых грозило бы потерей самоуважения
Вытеснение	избегание внутреннего конфликта путем выключения из сознания неприемлемого мотива или информации
Забывание	частный случай вытеснения, который связан с забыванием неприятной информации и намерений
Перенос	замещение действия с недоступным объектом на действия с доступным
Проекция	бессознательное приписывание другому лицу собственных чувств, желаний, влечений, в которых человек не хочет себе сознаться, понимая их социальную неприемлемость
Регрессия	возврат на более раннюю стадию развития или к более примитивным формам поведения и мышления
Замещение	защитный механизм, связанный с переносом действия с недоступного объекта на доступный
Отрицание	процесс устранения, игнорирования травмирующих восприятий внешней реальности
Изоляция/отчуждение	обособление и локализация внутри сознания травмирующих человека факторов

Информация о плохих или позорных поступках человека, «запрещенных» желаниях, его травматические душевные переживания могут смещаться в бессознательную область психики, в результате чего там образуются неосознанные комплексы, способные влиять на сознание человека, его эмоциональное состояние и поведение, вызывать негативные изменения в физическом и психическом здоровье человека.

Поэтому личности, на которых собственные комплексы имеют значительное влияние («закомплексованные» люди), строго логично обосновать свое поведение не могут. При таких условиях нельзя говорить о принятии личностью адекватных решений.

Появившись в бессознательном человека сначала в зародышевом виде и выражаясь внешне в виде капризов, повышенной раздражительности, безрассудности поведения и т.п., комплексы, в дальнейшем, развиваясь и включая все большую часть психики, оказываются уже в значительно более серьезных состояниях, таких как патология личности, неврозы, психосоматические и психические болезни.

Сами по себе данные психические образования невидимы и переживаются в сопровождающих их эмоциях. Наглядно комплексы оказываются в стереотипных паттернах поведения и формируются не только сильными травмирующими переживаниями, но и незначительными постоянно повторяющимися психическими травмами. Если материал бессознательного комплекса затрагивается на уровне эмоций или значений, тогда активируются все бессознательные взаимосвязи и провоцируемые ими несоответственно стереотипные паттерны поведения. Такой процесс проходит автономно, поскольку комплекс бессознателен. При его активации человек реагирует излишне эмоционально, решения принимаются под влиянием бессознательного человека, потому что мы реагируем не на актуальную ситуацию, а на все подобные ситуации, какие переживались на протяжении нашей жизни, когда комплекс оставался бессознательным. Обычно такое его проявление заканчивается стыдом из-за того, что мы не смогли контролировать ситуацию. Чем большее возбуждение, которое провоцируется комплексом, чем большая эмоциональная интенсивность комплекса, тем меньшей свободой наделена наша воля.

Комплекс самоунижения как источник внутриличностного конфликта

Очень часто вследствие такой психической деятельности возникают внутриличностные конфликты, в свою очередь тоже имеющие сильнейшее влияние на принятие решений человеком.

Под внутриличностным конфликтом мы понимаем внутриличностное противоречие, которое воспринимается и эмоционально переживается личностью как значимая для нее психологическая проблема, требующая решения и вызывающая внутреннюю работу сознания, при этом не всегда направляя ее на преодоление данного противоречия.

Так, если мы возьмём личность, подверженную, например, комплексу самоунижения, то вполне возможно, что при возникновении внутренних противоречий, внутреннего конфликта человек пребывает в состоянии стресса. При этом поступающий в кровь адреналин может вызывать приятные (как осознанные, так и бессознательные) ощущения, которые человек неосознанно захочет воспроизвести. Чтобы лучше понять, как происходит такой процесс, предлагаем подробнее рассмотреть это на примере личности, у которой имеется комплекс самоунижения.

У такого человека внешнее проявление самооценки – заниженное. Поэтому, оценивая себя, они навешивают сами себе наиболее уничижительные ярлыки. При этом его собственные потребности не перестают существовать, они только подавляются и глубоко скрываются от собственного сознания, создавая неосознанный внутриличностный конфликт. Отсюда хроническая неудовлетворенность жизнью, что есть естественным следствием подобного вытеснения. Но, не смотря на вышеизложенное, человек с комплексом самоунижения глубоко переживает любой собственный, даже незаметный постороннему, изъян.

Это есть одним из проявлений внутреннего конфликта, поскольку на самом деле, при всей своей внешней скромности и демонстрируемой неприязнительности такие люди имеют высокий уровень претензий к тому, каким должен быть «достойный» человек. Именно разница между образом Идеального Я и восприятием себя реального является причиной данного комплекса.

При всём том, что человек с комплексом самоуничужения выглядит скромным и готовым угодать всем, кто его окружает, в действительности он полностью сконцентрирован на самом себе, постоянно переживая, как он выглядит в глазах других людей, что он сделал неправильно, в чём виноват. Он заботится прежде всего о себе, когда стремится услужить: нужно выглядеть «хорошим», «милым», скромным, нельзя раздражать других, выделяться на фоне окружающих, надо быть удобным для всех. Поскольку постоянно происходящие сравнения себя с окружающими всегда не в его пользу, то, в соответствии с идеальным образом в своих внутрличностных ожиданиях, он должен быть лучше этих людей. Такое сравнение заставляет его постоянно извиняться, потому что проще предупредить людей о собственной несостоятельности, чем переживать впоследствии их разочарование. Кроме того, предупредив, он показывает себя человеком самокритичным, совестливым и скромным, что, с другой стороны, не только подтверждает ему его значимость, но и в какой-то степени всё же возвышает его над остальными (т.е. срабатывает компенсаторный защитный механизм).

Самоуничужающийся человек чрезвычайно зависим от мнения окружающих людей, поэтому любая критика с их стороны для него невыносима. И в этом выражается ещё одно противоречие, подтверждающее наличие внутрличностного конфликта. Критикуя открыто самого себя, он чувствует не только боль и гнев, если с его словами соглашаются, но и затаивает враждебные чувства по отношению к таким людям. Он ругает себя не для того, чтобы с ним соглашались, а для того, чтобы возражали ему, что, в конечном счёте, снова-таки даёт возможность получить подтверждение собственной значимости.

Часто подобный тип личности терпит ограничения, наблюдая, как другие смело идут вперед. Доказывая свою ничтожность, он иногда даже осознанно стремится, чтобы ему доказывали его значимость (т.е. убедиться от обратного), но бывает настолько убедительным в доказательствах несостоятельности, что в результате многие начинают относиться к нему в соответствии с его поведением. Вследствие этого возникают тайные чувства мстительности и враждебности. Но эти чувства тут же сливаются с чувством стыда и вины.

Внутренний конфликт такого типа личности состоит в том, что в глубине души самоуничужающийся человек почти уверен, что заслуживает презрения, и одновременно он хотел бы выглядеть лучше других. Будучи не в состоянии принять самого себя таким, каков есть, он не способен поверить, что другие, зная все его недостатки, могут отнестись к нему с дружбой и принятием. Такая личность перестаёт воспринимать любые положительные чувства окружающих. При этом сам человек не осознаёт происходящие в нем искажения: он уверен, что полностью прав, поскольку срабатывают механизмы психологической защиты, являющиеся специальной регулятивной системой стабилизации личности, направленной на устранение или сведение к минимуму ощущения тревоги, связанной с осознанием негативной информации.

Иногда такие люди могут погрязать в нищете и разрушать себя алкоголем, наркотиками. В ряде случаев они могут вести вполне достойный образ жизни, много трудиться и даже достигать определенных успехов, и при этом боятся больших достижений, потому что однажды достигнутые результаты влекут за собой соответствующие ожидания окружающих. Но именно это очень пугает человека с комплексом самоуничужения, поскольку он на самом деле боится провала, а значит и разочарования других людей по поводу его личности. В дальнейшем он ещё более тщательно следит, чтобы не было успеха, поскольку ему тяжело переживать именно сами страхи удач и неудач. Поэтому и передает ответственность за свою жизнь значимому Другому. Сильный, активный, властный партнер, который готов распоряжаться его жизнью и диктовать свои условия, максимально соответствует самоуничужающемуся человеку, готовому растворяться и угодать, поскольку привык слушаться и быть незаметным. Такая личность чувствует себя абсолютно зависимой, но эта зависимость – единственно возможный для нее способ существования.

Соответственно, партнер должен обладать всеми качествами, позволяющими ему занимать доминирующую позицию. Можно предположить, что при длительном и близком общении со зрелой личностью, доброжелательном отношении человека, заботящегося о развитии своего избранника, самоуничижение постепенно могло бы уступить место вере в собственные силы и самоуважению. Но, как правило, самоуничижающийся человек чувствует влечение именно к доминирующим людям, а те видят в нем благодатную почву для удовлетворения своих потребностей во власти и контроле.

Причиной для развития такого комплекса служат, с одной стороны, завышенные требования со стороны родителей, оправдать которые ребенок явно не способен, и постоянные сравнения с другими детьми, которые он слышит от родителей и практически всегда не в его пользу. Родители, желая таким образом стимулировать у малыша стремление совершенствоваться и достигать успехов, получают обратное: ребенок понял, что он просто не может соответствовать столь высоким требованиям. В дальнейшем он сам к себе начинает предъявлять подобные требования: сравнивая себя с другими, постоянно сам себя критикует, усваивает, что если он ничего не делает, то получает гораздо меньше критических замечаний, чем когда начинает что-то делать и при этом ошибается.

С другой стороны, слишком требовательные родители никогда не бывают полностью удовлетворены действиями ребенка. Постепенно у него развивается страх провала. Это еще больше сковывает его и приводит к новым неудачам. Но, поскольку ему жизненно важно получить любовь, свою неловкость и неумелость малыш старается компенсировать ласковостью, послушанием и/или незаметностью. А если ему все-таки приходится что-то делать, он заранее извиняется за свою неполноценность, как бы предупреждая разочарование (но в любом случае он выигрывает: если не получилось – он же предупреждал и этим своим предупреждением становится как бы над теми, кто ждет от него положительных результатов; получилось – он значимый, что подтверждают те, кто подбадривал его, хотя в бессознательном он и сам знает, что он очень хороший).

Однако низкое самомнение ребенка вовсе не означает, что он действительно никуда не годится. Часто самоуничижение разъедает душу весьма талантливых людей. Поэтому успехи все-таки периодически достигаются, несмотря на все блокирующие обстоятельства. Но любое достижение малыша вызывает только одну реакцию его родителей: «ну вот, можешь, когда хочешь». Теперь от него ожидают еще большего. По существу, от него хотят безупречности и совершенства. Планка требований поднимается. Угроза провала увеличивается. Так у малыша развивается страх успеха. Он предпочитает подавить свои способности, чтобы даже нечаянно не вызвать пугающие его ожидания.

Или же кто-то из родителей передает ребенку собственный комплекс самоуничижения, помогая создавать установку не высовываться, поскольку есть много людей, которые умнее и лучше во всех отношениях. Такие люди не любят неудачников, поэтому безопаснее находиться незамеченным среди остальных.

Итак, формирование личности с комплексом самоуничижения происходит под воздействием таких факторов:

Родительские посылы:	Выводы ребенка	Формирующиеся черты характера личности
<p><i>Ребенок должен:</i></p> <p>1) соответствовать родительским ожиданиям;</p> <p>2) быть лучше других;</p> <p>3) выглядеть лучше других;</p> <p>4) раньше других</p>	<p><i>Если ребенок не может выполнять то, что от него требуют родители:</i></p> <p>1) значит, он неполноценный;</p> <p>2) любят только умных, умелых, успешных людей;</p> <p>3) он недостойн такой нужной ему родительской любви.</p>	<p>1) эгоцентризм («зацикленность» на себе);</p> <p>2) неуверенность в себе;</p> <p>3) застенчивость;</p> <p>4) высокий уровень требований к себе;</p> <p>5) стремление привлечь симпатию окружающих посредством жалости к себе;</p>

<p>научиться многому такому, что остальные не умеют;</p> <p>5) делать это в совершенстве;</p> <p>6) не высовываться, чтобы над ним не смеялись</p> <p><i>Если у ребенка не получается:</i></p> <p>1) он ничтожество;</p> <p>2) он расстраивает родителей;</p> <p>3) он недостоин любви родителей;</p> <p>4) он не достоин уважительного отношения.</p>	<p><i>Чтобы заслужить родительскую любовь, нужно:</i></p> <p>1) постараться, по крайней мере, поменьше раздражать родителей;</p> <p>2) быть милым и предупредительным;</p> <p>3) постараться предупреждать их заранее о возможности его неудач, чтобы они меньше сердились;</p> <p>4) чем реже он что-либо делает, тем меньше на него сердятся;</p> <p>5) нужно поменьше проявлять инициативу и свести все дела к минимуму;</p> <p>6) следить, чтобы его успехи не были слишком заметными, поскольку всё равно, это – дело случая, а потом он обязательно ещё больше разочарует родителей и его ещё меньше будут любить.</p>	<p>6) постоянное чувство вины и стыда;</p> <p>7) страх неудачи;</p> <p>8) страх успеха;</p> <p>9) обидчивость;</p> <p>10) тайная мстительность;</p> <p>11) болезненное восприятие критики;</p> <p>12) избегание конфликтных ситуаций, даже в ущерб себе;</p> <p>13) невозможность выразить своё мнение, постоять за себя;</p> <p>14) хроническая неудовлетворенность жизнью.</p> <p>15) стремление передать ответственность за важные решения и свою судьбу другим людям;</p> <p>16) склонность выбирать доминирующих партнеров, руководящих его жизнью;</p> <p>17) склонность выбирать критикующих и унижающих его партнеров (выполняется модель родительской семьи);</p> <p>18) склонность к депрессии.</p>
--	--	---

Постоянное подавление собственной энергии и стремление быть незаметным постепенно сказывается на всём облике человека, а затем и на всей его судьбе.

Определенных тестов на выявление людей с комплексом самоуничужения нет. Но если у человека даже без специального тестирования наблюдается значительно заниженная самооценка, если у него постоянно проявляется набор таких черт характера, как застенчивость, высокий уровень требований к себе, постоянное чувство вины и стыда, склонность к избеганию риска, обидчивость, болезненное восприятие критики, избегание конфликтных ситуаций даже в ущерб себе, невозможность выразить своё мнение и постоять за себя, стремление передать ответственность за важные решения и свою судьбу другим людям, неуверенность в себе, выбор доминирующих или же критикующих и унижающих его партнеров, стремление привлечь внимание окружающих посредством жалости к себе, склонность к депрессии и т.д., с большой долей вероятности даже на бытовом уровне мы можем говорить о том, что перед нами личность с чертами комплекса самоуничужения.

Проблемы принятия решения

Рассмотренный тип личности в большинстве случаев будет принимать решения под влиянием бессознательных эмоций, препятствующих конструктивной деятельности в данной ситуации. Возникающий вследствие этого внутрличностный конфликт втягивает человека как бы в замкнутый круг: Неадекватность принятого решения часто вызывает (как осознанно, так и бессознательно) чувство стыда, а это, в свою очередь, снова мешает принятию адекватного решения. Поэтому, существует высокая вероятность того, что при принятии решений такой тип личности будет руководствоваться следующим:

- стараться держаться в тени, избегая иметь собственное мнение;

- уклоняться от самостоятельного принятия решения вообще;
- предпринимать что-либо важное без подсказки со стороны какого-либо авторитетного лица;
- стараться передавать ответственность при решении любых вопросов кому-нибудь, по его мнению, более компетентному;
- в случае же неизбежности принятия решений он будет заранее извиняться за свою некомпетентность, непонимание происходящего, неумелость и т.д.

Таким образом, данный тип личности, согласно сетке Томаса-Килмана, при возникновении конфликтной ситуации приемлемыми способами выхода из нее для себя будет считать уклонение и/или приспособление. Кроме того, такие люди заранее уверены не только в провалах собственных планов, но и часто остерегаются давать оптимистические прогнозы относительно тех или иных проектов, если сами в них включены.

Заключение

Несомненно, любая типология приближительна и условна. Довольно редко можно встретить людей, точно отражающих любой описанный тип. Но психологические портреты позволяют лучше сориентироваться во внутреннем мире человека, быстрее увидеть скрытые «звенья», построить предположения и проверить их. Сталкиваясь в повседневной жизни, например, с данным типом людей мы можем не только понять, что в психике воздействует на принятие решения таким человеком, но и заранее спрогнозировать его. Поэтому, назначая экспертов, важно учитывать психологические факторы, влияющие на субъективность оценивания личностью того или иного явления.

Но если человек захочет изменить себя, свой тип личности, то в результате решения внутриличностных противоречий, возникающих под воздействием разнообразных комплексов, он может перейти (в первую очередь – через формирование адекватной самооценки) на качественно новый способ жизнедеятельности: при условии преодоления предыдущих препятствий внутриличностного развития, у человека есть возможность перевести их во внутренний опыт, который поможет ему в последующем построении конструктивного сценария своей жизни.

Библиография

- [Айсмонтас, 2004] Айсмонтас Б.Б. Общая психология. Схемы. – М., 2004.
- [Волошин, Мащенко, 2006] Волошин О.Ф., Мащенко С.О. Теорія прийняття рішень. – К, 2006.
- [Грибак, 1989] Грибак Л. П. Резервы человеческой психики. – М., 1989.
- [Гринсон, 1994] Гринсон Р. Практика и техника психоанализа. Новочеркасск, 1994.
- [Емельянова, 2004] Емельянова Е.В. Кризис в созависимых отношениях. – СПб, 2004.
- [Мерлин, 1970] Мерлин В.С. Проблемы экспериментальной психологии личности. – Пермь, 1970.
- [Томе, Кэхеле, 1996] Томе Х., Кэхеле Х. Современный психоанализ. Практика. В 2-х томах. Перевод с английского. – М., 1996.

Сведения об авторе

Шинкаренко Елена Владимировна – Кировоградский педагогический университет им. Владимира Винниченко кандидат филологических наук, доцент, Кировоград, Украина e-mail: oshink@kspu.kr.ua

КУЛЬТУРНАЯ ИДЕНТИФИКАЦИЯ КАК ФАКТОР СТАБИЛЬНОСТИ В ДВУЯЗЫЧНОМ ОБЩЕСТВЕ

Ирина Горицына, Александр Глущенко

Аннотация: Рассматривается проблема, которая возникает в двуязычном обществе. Культурная идентичность населения может помочь сохранению языков меньшинств в районах их компактного проживания. Эта проблема формулируется в терминах теории принятия коллективных решений. Приводятся условия «устойчивости» языковых групп игр.

Ключевые слова: культурная идентичность, принятие коллективных решений, равновесие Нэша.

Введение

С января 2005 года Украина является участником важного документа Совета Европы, известного под названием Европейская хартия региональных языков или языков меньшинств. Подписание Хартии от имени Украины состоялось уже 2 мая 1996 года. Но процесс ее ратификации продолжался несколько лет, сопровождаясь жестким политическим противостоянием и грубыми нарушениями Конституции Украины и украинского законодательства. Верховная Рада Украины 15 мая 2003 года одобрила Закон Украины № 804-IV «О ратификации Европейской хартии региональных языков или языков меньшинств».

Вопреки духу и содержанию Хартии [Хартия, 1992], этот закон возлагает на государство чрезмерные обязательства относительно особой защиты тех языков, которые в этом не нуждаются, и оставляет вне сферы своего действия ряд языков, которым действительно нужны особая защита и поддержка

Закон предусматривает одинаковые меры особой защиты к таким языкам меньшинств, как белорусский, болгарский, венгерский, гагаузский, греческий, еврейский, крымскотатарский, молдавский, немецкий, польский, русский, румынский и словацкий.

Часть этих языков не нуждаются в особой защите. Они являются официальными в соответствующих государствах и имеют необходимую государственную защиту. В границах Украины также отсутствуют препятствия для применения этих языков представителями соответствующих национальных меньшинств в местах их компактного проживания. Состояние языков этих меньшинств в Украине вполне удовлетворительное, поскольку украинское законодательство и практика традиционно уделяют большое внимание защите прав национальных меньшинств, в том числе и языковых.

Среди перечисленных языков самую сильную позицию занимает русский, который длительное время был «языком межнационального общения» в бывшем СССР и продолжает широко использоваться в границах всей Украины, особенно в ее восточной части. В Украине русский язык не является региональным или миноритарным, поскольку, в соответствии с Хартией, подпадает под определение языка, не связанного с определенной территорией. Это является дополнительным основанием применять по отношению к нему не особый, а базовый режим защиты.

В то же время особый режим защиты должен бы распространяться на такие языки, как гагаузский, караимский, крымскотатарский, крымчацкий, ромский. Эти языки не являются широко употребляемыми ни в Украине, ни за ее пределами. Они не защищены официальным статусом, поскольку народы, которые на них говорят, немногочисленны и не имеют собственных государственных образований [Василенко, 2006].

Наконец, в соответствии со статьей 3 Хартии, меры особой защиты должны применяться к украинскому языку, который (несмотря на его статус государственного языка) остается в состоянии, угрожающем его

существованию. В годы независимости наблюдалось даже ухудшение фактического состояния украинского языка и сужения его применения в определенных сферах общественной жизни. И это при том, 77,8% населения идентифицируют себя с этническими украинцами согласно данным Всеукраинской переписи населения 2001 г. В общем, распределение населения по культурной идентификации и родному языку по данным переписи 2001 года выглядит следующим образом:

Культурная идентификация	Считают родным языком		
	язык своей национальности	украинский язык	русский язык
Украинцы (77.8%)	85,2%	X	14,8%
Русские (17.3%)	95,9%	3,9%	X

Таким образом, в Украине существует три мегаязыковые группы: украиноязычные украинцы, русскоязычные украинцы и русскоязычные русские.

Таким образом, вопрос о том, каким образом украинское государство будет строить целостную политическую нацию, на 16-ом году независимости остается чрезвычайно актуальным в контексте развития языка, который в этой стране является государственным по закону, но не деле.

Вопрос языка, а скорее спекуляции на этом вопросе, использование языковых реалий в политических целях является фактором, который способен окончательно поляризовать общество, способствовать окончательному разделу между областями Украины.

В контексте процессов, которые происходят сейчас в Украине, проблема культурной идентичности приобретает особенно актуальное значение. Становление украинской государственности предусматривает акцент внимания на данном феномене. Речь идет не только о идентичности личности, но и о идентичности этноса в целом, его материальной и духовной культуры. Кто мы? Где наши корни? Каковы источники нашего бытия? Эти и другие вопросы возникают в процессе формирования государственных институтов независимой Украины.

Люди в основном сообщают о своей социальной идентичности посредством использования некоторых общих для группы символов или маркеров. Это может быть определенный особенный стиль одежды, прическа, определенные украшения, способ питания и т.д. Для многих этнических групп в качестве наиболее значимого символа коллективного членства выступает язык. Поэтому не случайно этнические сообщества всегда в первую очередь озабочены сохранением родного языка, полноценным обучением языку подрастающего поколения, его дальнейшим развитием. В то же время не принято абсолютизировать язык как единственный способ сохранения этнической идентичности. Существует немало примеров, когда главным идентификационным критерием выступает не язык, а религия, история и т.п.

Если вернуться к вопросу языка, то необходимо отметить двухстороннюю связь между языком и идентификацией – с одной стороны, использование языка как культурного символа и маркера является следствием идентификации, с другой стороны, существование языка является необходимым условием самоидентификации и самокатегоризации.

Предпосылки для построения модели

Данная статья является естественным продолжением работы [Горицына, 2006], которая обсуждалась на KDS-2006. Коротко напомним ее основные положения. Рассмотрим общество, разговаривающее на двух языках $\{a, b\}$. Для простоты предположим, что семьи состоят из одного родителя и одного ребенка.

Функция полезности родителя состоит из двух компонент. Во-первых, владение языком приносит определенную пользу, во-вторых, немаловажную роль имеет и любовь (забота) родителя к своему ребенку.

Польза (полезность) от владения языком выбрана так, чтобы учесть преимущества, получаемые при возможности на бытовом уровне облегчить общение с людьми, говорящих на двух языках, что немаловажно, осознавая себя частью другой культуры или хотя бы понимая ее. Это учтено в модели отдельно от материального потребления, поскольку нет достаточных доказательств того, что знание определенного языка существенно влияет на потребление товаров. Различия, которые мы видим в моделях потребления, главным образом определяется культурными факторами (религия, раса, этническая принадлежность, и т.д.).

Родительская функция полезности включает затраты на образование ребенка. Родитель считает своим долгом передать ребенку свою этническую (или культурную) идентичность. Культурная идентичность однозначно связана со знанием одного из двух языков.

Языковая функция полезности (v) - одинакова для родителя и для ребенка. Однако, для того, чтобы промоделировать желание передать культурную идентичность, предположим, что родитель и ребенок получает одинаковую пользу от владения одним и тем же языком при условии, что и все общество разговаривает на этом же языке. Другими словами, он получает такую же пользу, как будто ребенок живет в одноязыковой, монокультурной стране. В противном случае, он не получает никакой пользы от передачи культурной идентичности. Функция максимизации также включает отрицательную компоненту, соответствующую затратам на обучение ребенка одному или двум языкам.

Наличие языковых навыков представлено булевой переменной s , которая показывает уровень владения языком.

Функция социализация или функция изучения языка S зависит от знания родителем языка, вектора денежных "доходов" и удельного веса населения, которое разговаривает на этом языке. Функция зависит от булевых (двоичных) переменных и имеет ограниченные значения по каждому фактору. Предполагается, что обучение языку меньшинства обходится дороже в обществе, которое имеет другой язык большинства. В то время, как одинаковые потребности требуют одинаковых затрат, то чем ниже удельный вес населения, разговаривающего на языке меньшинства, тем больше значение функции изучения языка S .

В доходной части бюджетного уравнения заработная плата родителя зависит от его языковых способностей. Предполагается, что языковые способности родителя известны и все люди вначале имеют некоторые знания другого языка. Заработная плата имеет постоянную компоненту W , не связанную со знанием языков, и компоненту, обусловленную знанием языка и зависящую от доли населения, с которой человек может общаться, основываясь на своих знаниях языков.

Постановка задачи и решение

Задача Родителя состоит в максимизации функции полезности от потребления некоторых материальных благ в совокупности с некоторыми благами и преимуществами, которые дают культурная идентичность и знание языков, на которых разговаривают члены общества:

$$U = u(x) + s^i_p \cdot v(q^i_p + q^{ij}_p) + s^j_p \cdot v(1 - q^j_p) + u(x_c(s^i, s^j)) + s^i \cdot v(1) - C(\tau^i + \tau^j) \rightarrow \max,$$

Необходимо определить x , x_c , τ^i , τ^j в области допустимых решений (бюджетных ограничений для себя и своего ребенка):

$$px = W + w(s^i_p \cdot q^i_p + q^{ij}_p + s^j_p \cdot q^j_p)$$

$$px_c = W + w(s^i \cdot q^i + q^{ij} + s^j \cdot q^j),$$

$$\text{где } s^i = S(s^i_p, \tau^i, q^i_p + q^{ij}_p), \quad s^i \in \{0, 1\}, \quad s^i_p \in \{0, 1\}, \quad q^i + q^j + q^{ij} = 1,$$

$u(x)$	монотонно возрастающая, строго вогнутая функция полезности,
x	множество выборов родителя (наборы благ, на которые могут быть потрачены деньги);
x_c	множество выборов ребенка (наборы благ, на которые могут быть потрачены деньги);
$i \in \{a, b\}$	языки, на которых разговаривают, a - язык большинства, b - язык меньшинства;
s^i	языковая переменная ребенка, $s^i = 1$, если ребенок имеет способности к языку i , $s^i = 0$ в противном случае;
s^i_p	языковая переменная родителя, $s^i_p = 1$, если родитель владеет языком i , $s^i_p = 0$, в противном случае;
$q^i \in [0; 1]$	удельный вес детей, говорящих только на языке i ; $q^i = N(s^i = 1 \ \& \ s^j = 0)/N$, $q^j = N(s^i = 0 \ \& \ s^j = 1)/N$;
$q^{ij} \in [0; 1]$	удельный вес детей, говорящих на двух языках; $q^{ij} = N(s^i = 1 \ \& \ s^j = 1)/N$, $q^i + q^j + q^{ij} = 1$;
$q^i_p \in [0, 1]$	удельный вес родителей, говорящих только на языке i ;
$q^{ij}_p \in [0, 1]$	удельный вес родителей, говорящих на двух языках;
$v(q^i)$	дополнительная польза от владения языком i (монотонно возрастающая, строго вогнутая функция);
$v(1)$	функция полезности i -язычного родителя при наличии у ребенка способностей к родному языку;
$w(q^i)$	монотонно возрастающая, строго вогнутая функция заработной платы, связанная со знанием i -го языка;
$C(\tau^i)$	строго возрастающая функция стоимости, $C(0) = 0$, $dC(0)/d\tau^i = 0$;
τ^i	N -мерный вектор затрат;
S	механизм социализации, подготовки к жизни в обществе $S(0, 0, 0) = 0$, $dS/ds^i_p \geq 0$, $dS/d\tau^i \geq 0$, $dS/dq^i \geq 0$;
N	численность населения.

Функция социализации устанавливает однозначное соответствие между s^i и τ^i при условии, что знания языков родителей и распределения населения относительно знаний языков известны заранее. Поэтому, введем новую переменную для обозначения затрат, необходимых для изучения языка. Обозначим минимальные затраты, необходимые для обучения ребенка языку при условии, что и родитель говорит на этом же языке:

$$\tau^i \rightarrow \min$$

$$\text{при условии } S(1, \tau^i, q^i_p + q^{ij}_p) = 1.$$

Пусть τ^{i_1} – значение целевой функции этой задачи. Естественно, что родитель заинтересован в минимизации затрат на обучение языку. Аналогично обозначим через τ^{i_0} минимальные затраты (значение целевой функции), необходимые для обучения ребенка языку, на котором родитель не разговаривает:

$$\tau^i \rightarrow \min$$

$$\text{при условии } S(0, \tau^i, q^i_p + q^{ij}_p) = 1.$$

Так как затраты определены в задаче максимизации функции полезности, то существует только две альтернативы: обучать ребенка языку, или нет. Следовательно, максимальное значение $\tau^i \in \{0, \tau^{i_1}\}$, если родитель говорит на языке i или $\tau^i \in \{0, \tau^{i_0}\}$, если он не говорит на этом языке.

Стабильность распределения языковых групп

В дальнейшем мы абстрагируемся от факта, что минимальные затраты, необходимые для обучения определенному языку $\tau^{a_{11}}$, $\tau^{a_{01}}$ и т.д. зависят от времени, так как изменяется во времени и доля населения, которое говорит на языке $a - (q^a_p + q^{ab}_p)$. Т.е., соответствующее значение должно быть равным $\tau^{i_{11}}(q^i_p + q^{ij}_p)$. Однако для задачи оптимизации в рамках одного временного периода примем эти величины заданными.

Для нахождения решения с учетом возможного распределения населения в зависимости от знания языка (языков), будем считать величины q^a , q^b известными.

Определение расходов на образование ребенка можно рассматривать как игру. Игроками является все население страны (региона). В зависимости от знания языков и от своей культурной идентичности все игроки образуют четыре коалиции, внутри которой все игроки соблюдают одну стратегию, т.е. действуют как один игрок. Стратегиями игроков является желание обучить своего ребенка одному или двум языкам.

Для исследования этих игр будем использовать равновесие Нэша в чистых стратегиях [Волошин, 2006]. Следовательно, игра может быть представлена как игра четырех игроков с тремя стратегиями у каждого игрока. Игроками являются:

- ✓ игрок, владеющий только языком a (**A**),
- ✓ игрок, владеющий двумя языками, отождествляющий себя с культурой a (**2A**),
- ✓ игрок, владеющий двумя языками, отождествляющий себя с культурой b (**2B**),
- ✓ игрок, владеющий только языком b (**B**).

В качестве трех стратегий выбраны:

- обучение своего ребенка языку a ,
- обучение ребенка двум языкам,
- обучение ребенка языку b .

Представим схематически стратегии и выигрыши, принимая во внимание, что зависят q^a , q^b , $q^2=q^{ab}$ от выборов других игроков (табл. 1).

Таблица 1. Таблица выигрышей игроков

Стратегии Игроки	Обучение языку a	Обучение двум языкам	обучение языку b
Игрок 2A	$u\{W + w(q^a + q^2)\} + v(1) - C(\tau^{a_{11}})$	$u\{W + w(1)\} + v(1) - C(\tau^{a_{11}} + \tau^{b_{11}})$	$u\{W + w(1 - q^a)\} - C(\tau^{b_{11}})$
Игрок 2B	$u\{W + w(q^a + q^2)\} - C(\tau^{a_{11}})$	$u\{W + w(1)\} + v(1) - C(\tau^{a_{11}} + \tau^{b_{11}})$	$u\{W + w(1 - q^a)\} + v(1) - C(\tau^{b_{11}})$
Игрок A	$u\{W + w(q^a + q^2)\} + v(1) - C(\tau^{a_{11}})$	$u\{W + w(1)\} + v(1) - C(\tau^{a_{11}} + \tau^{b_{01}})$	$u\{W + w(1 - q^a)\} - C(\tau^{b_{01}})$
Игрок B	$u\{W + w(q^a + q^2)\} - C(\tau^{a_{01}})$	$u\{W + w(1)\} + v(1) - C(\tau^{a_{01}} + \tau^{b_{11}})$	$u\{W + w(1 - q^a)\} + v(1) - C(\tau^{b_{11}})$

Определим условия устойчивости для всех игроков.

Для группы игроков **2A**

$$u\{W + w(1)\} + v(1) - C(\tau^{a_{11}} + \tau^{b_{11}}) \geq u\{W + w(q^a + q^2)\} + v(1) - C(\tau^{a_{11}}), \quad (1)$$

$$u\{W + w(1)\} + v(1) - C(\tau^{a_{11}} + \tau^{b_{11}}) \geq u\{W + w(1 - q^a)\} - C(\tau^{b_{11}}). \quad (2)$$

Для группы игроков **2B**

$$u\{W + w(1)\} + v(1) - C(\tau_{11}^a + \tau_{11}^b) \geq u\{W + w(qa + q2)\} - C(\tau_{11}^a) \quad (3)$$

$$u\{W + w(1)\} + v(1) - C(\tau_{11}^a + \tau_{11}^b) \geq u\{W + w(1 - qa)\} + v(1) - C(\tau_{11}^b). \quad (4)$$

Для группы игроков **A** ($s^a = 1, s^b = 0$)

$$u\{W + w(q^a + q^2)\} + v(1) - C(\tau_{11}^a) \geq u\{W + w(1 - q^a)\} - C(\tau_{01}^b) \quad (5)$$

$$u\{W + w(q^a + q^2)\} + v(1) - C(\tau_{11}^a) \geq u\{W + w(1)\} + v(1) - C(\tau_{11}^a + \tau_{01}^b) \quad (6)$$

Для группы игроков **B** ($s^a = 0, s^b = 1$)

$$u\{W + w(1 - q^a)\} + v(1) - C(\tau_{11}^b) \geq u\{W + w(q^a + q^2)\} - C(\tau_{01}^a) \quad (7)$$

$$u\{W + w(1 - q^a)\} + v(1) - C(\tau_{11}^b) \geq u\{W + w(1)\} + v(1) - C(\tau_{01}^a + \tau_{11}^b) \quad (8)$$

Модель с сильным фактором осознания культурной идентичности

Чтобы понять, как сильный фактор осознания культурной идентичности меняет задачу, посмотрим на равновесие Нэша в этой же игре. Это означает, что в обществе никто не теряет свою культурную идентичность.

Если фактор передачи культурной идентичности

$$v(1) > u\{W + w(q^a + q^2)\} - u\{W + w(1 - q^a)\} + C(\tau_{11}^a + \tau_{11}^b) - C(\tau_{11}^a),$$

то представители различных языковых групп отдадут предпочтение своему языку.

Придание большого веса фактору культурной передачи означает, что никто теряет своей культурной идентификации.

Введем обозначения

$$\Delta u^{2a} = u\{W + w(1)\} - u\{W + w(q^a + q^2)\};$$

$$\Delta u^{2b} = u\{W + w(1)\} - u\{W + w(1 - q^a)\};$$

$$\Delta u^{ab} = u\{W + w(q^a + q^2)\} - u\{W + w(1 - q^a)\} = \Delta u^{2b} - \Delta u^{2a}.$$

Тогда условия (1')-(8') можно переписать:

$$\Delta u^{2a} - C(\tau_{11}^a + \tau_{11}^b) + C(\tau_{11}^a) \geq 0; \quad (1)$$

$$\Delta u^{2b} - C(\tau_{11}^a + \tau_{11}^b) + v(1) + C(\tau_{11}^a) \geq 0; \quad (2)$$

$$\Delta u^{2a} - C(\tau_{11}^a + \tau_{11}^b) + v(1) + C(\tau_{11}^a) \geq 0; \quad (3)$$

$$\Delta u^{2b} - C(\tau_{11}^a + \tau_{11}^b) + C(\tau_{11}^b) \geq 0; \quad (4)$$

$$\Delta u^{2b} - \Delta u^{2a} + v(1) - C(\tau_{11}^a) + C(\tau_{01}^b) \geq 0; \quad (5)$$

$$-\Delta u^{2a} - C(\tau_{11}^a) + C(\tau_{11}^a + \tau_{01}^b) \geq 0; \quad (6)$$

$$\Delta u^{2a} - \Delta u^{2b} + v(1) - C(\tau_{11}^b) + C(\tau_{01}^a) \geq 0; \quad (7)$$

$$-\Delta u^{2b} - C(\tau_{11}^b) + C(\tau_{01}^a + \tau_{11}^b) \geq 0;$$

Для двуязычного населения, отождествляющего себя с культурой **a**, стратегия превращения в **a**-одноязычных доминирует над стратегией превращения в **b**-одноязычных. Это видно из предположения

$$q^a + q^2 \geq q^b + q^2 = 1 - q^a,$$

а также из предположения, которое мы сделали относительно функции социализации. То есть требуется больше усилий для изучения языка меньшинства в среде (окружении) с меньшим удельным весом населения, говорящего на этом языке.

Формально это можно выразить так

$$\tau_{11}^b \geq \tau_{11}^a \text{ и } C(\tau_{11}^b) \geq C(\tau_{11}^a);$$

$$\min \tau^a \leq \min \tau^b$$

в области ограничений

$$S(1, \tau^a, q^a + q^2) = 1;$$

$$S(1, \tau^b, q^b + q^2) = 1.$$

Справедливость (4) и (1) влечет за собой справедливость (2) и (3), поскольку они содержат дополнительное положительное слагаемое v (1) в левой части неравенства. Далее, учитывая, что a – группа большинства, а b – группа меньшинства. Это означает, что $w(q^a + q^2) \geq w(q^b + q^2) = w(1 - q^a)$.

Если в обществе удельный вес меньшинства достаточно мал, тогда все двуязычные родители, отождествляющие себя с культурой большинства, не будут обучать своих детей языку меньшинства.

Это следует непосредственно из уравнения (1). Поскольку

$$-C(\tau_{11}^a + \tau_{11}^b) + C(\tau_{11}^a) < 0,$$

то

$$\Delta u^{2a} > 0 \Rightarrow u(W + w(1)) - u(W + w(q^a + q^2)) \gg 0.$$

Другими словами, для выполнения этого неравенства необходимо, чтобы величина q^j была достаточно близкой к единице.

Таким образом, условие (5) всегда верно, если группа a является группой большинства. Следовательно, мы можем сократить систему неравенств стабильности, исключив доминирующие стратегии, вычеркивая уравнения (2) и (5).

Устраняя доминирующие стратегии, т.е. условия (2) и (5) мы ограничиваем группу, отождествляющую себя с культурой a , только двумя стратегиями выбора: двуязычие или a -одноязычие. Условия (2) и (5) показывают переход к b -одноязычию для двуязычных и a -одноязычных соответственно, что противоречит предпосылкам нашей модели.

Если значение q^j становится слишком малым (близким к нулю), тогда двуязычные родители, отождествляющие себя с культурой большинства, прекращают обучать своих детей языку меньшинств. Это не означает, что двуязычная группа прекращает свое существование. Те двуязычные, кто отождествляет себя с культурой меньшинства, должны будут принять другое решение. Различие между условием (3) и (1) - наличие дополнительной переменной культурной полезности (или передачи культурной идентификации) $v(1)$, которая в основном подразумевает, что для передачи культурной идентификации родители, принадлежащие к меньшинству, видят пользу для своих детей так, как будто бы они жили в стране, где говорят только на одном языке меньшинства. Учитывая сильное желание передать свою культурную принадлежность, величина $v(1)$ должна быть достаточно большой, чтобы нейтрализовать отрицательные компоненты с тем, чтобы равенство оставалось истинным.

Другое условие для двуязычного меньшинства - это выполнение условия(4). Все же, если q^j - достаточно велико, Δu^{2b} будет достаточно большим, чтобы возместить дополнительные затраты $-C(\tau_{11}^a + \tau_{11}^b) + C(\tau_{11}^b)$ на изучение языка большинства. Кроме того, изучение языка большинства упрощается, когда удельный вес населения, говорящем на этом языке, достаточно высок.

Уравнения (5) и (7) показывают условие, при котором одноязычные родители не обучают своих детей второму языку при достаточно большом $v(1)$. Другими словами, если желание передать культурную принадлежность сильно, то (5) и (7) справедливы. С другой стороны, (5) и (7) почти всегда следуют из (6) и (8), если усилие, необходимое для изучения чьего-либо родного языка $\tau^{a,b}_{11}$, не высоко.

Если мы налагаем дополнительное предположение на функции социализации, то (8) подразумевает (6). Поскольку $\Delta u^{2b} \geq \Delta u^{2a}$ и $C(\tau^{b_{11}}) \geq C(\tau^{a_{11}})$, посмотрим на величины $\tau^{a_{11}} + \tau^{b_{01}}$ и $\tau^{b_{11}} + \tau^{a_{01}}$. Логично сделать следующее предположение: усилия, потраченные на обучение обоим языкам, когда родитель знает только один язык большинства, превышают усилия, потраченные на обучение двум языкам при родителе, знающем только один язык меньшинства. Тогда следует, что $C(\tau^{a_{11}} + \tau^{b_{01}}) \geq C(\tau^{b_{11}} + \tau^{a_{01}})$ и поэтому выполнение (8) влечет за собой справедливость (6).

Вывод. Одоязычные родители, принадлежащие меньшинству, не обучают своих детей языку большинства, если усилия, требуемые для этого, являются достаточно высокими.

Главными условиями, определяющими стабильность разноязычных групп населения, являются (1) и (8). Они управляют стабильностью распределения знаниями языка в обществе. Если начальная доля людей, говорящих на языке меньшинства, слишком низка, тогда все двуязычные, отождествляющие себя с культурой большинства, перестали бы говорить на языке меньшинства. Если это произойдет, то это не обязательно дестабилизирует другие языковые группы. Вопрос заключается в том, насколько большой первоначально была двуязычная группа населения в группе, отождествляющей себя с культурой большинства.

Если условия (1) и (8) верны и $\tau^{a_{11}} + \tau^{b_{01}} = \tau^{a_{01}} + \tau^{b_{11}}$, то существует равновесие Нэша в одношаговой игре и в обществе сохраняется начальная двуязычная среда.

Библиография

- [Волошин, 2006] Волошин А.Ф., Мащенко С.О., Теория принятия решений. Учебное пособие.–Киев: ИПЦ “Киевский университет”, 2006.–304 с. (на укр.яз).
- [Хартия, 1992] Европейская хартия региональных языков или языков меньшинств.– Страсбург, 5 ноября 1992 г.
- [Василенко, 2006] Василенко В. Какие языки в Украине нуждаются в особой защите. Коллизия между национальным законом и обязательствами государства // «Зеркало недели», № 10 (589), Суббота, 18 - 24 марта 2006 года.
- [Горицына, 2006] Горицына И.А., Глуценко А.А. Роль фактора культурной идентичности в двуязычном обществе.– В кн.: Proceeding of the XII International Conference “Knowledge-Dialog-Solution” – Varna, 2006.–Sofia: FOI Commerce, 2006.— P/158-164.

Информация об авторах

Ирина Горицына – старший научный сотрудник, e-mail: goritsyna@unicyb.kiev.ua

Александр Глуценко – заведующий лабораторией, e-mail: mmeed@unicyb.kiev.ua

Киевский национальный университет им. Т. Шевченко, факультет кибернетики; пр. Акад. Глушкова, 2, корп. 6, Киев, Украина.

H.4.2. Decision Support

СИСТЕМЫ ДИАГНОСТИРОВАНИЯ В МЕДИЦИНЕ, КАК ПЕРСОНАЛЬНЫЙ ИНТЕЛЛЕКТУАЛЬНЫЙ ИНСТРУМЕНТАРИЙ ВРАЧА

Алексей Волошин, Максим Запорожец, Павел Мулеса

***Аннотация:** Предлагаются принципы построения диагностических систем в медицине, основанных на моделировании «способа действия» эксперта в виде нечетких деревьев принятия решений с учетом критериев правдоподобия и полезности. Рассматриваются фрагменты «прикладных» деревьев при диагностировании инфекционных и урологических заболеваний. Обсуждаются возможности применения современного инструментария теории принятия решений при создании систем искусственного интеллекта.*

***Ключевые слова:** Теория принятия решений; деревья решений; правдоподобность; полезность; диагностические системы в медицине.*

Предисловие

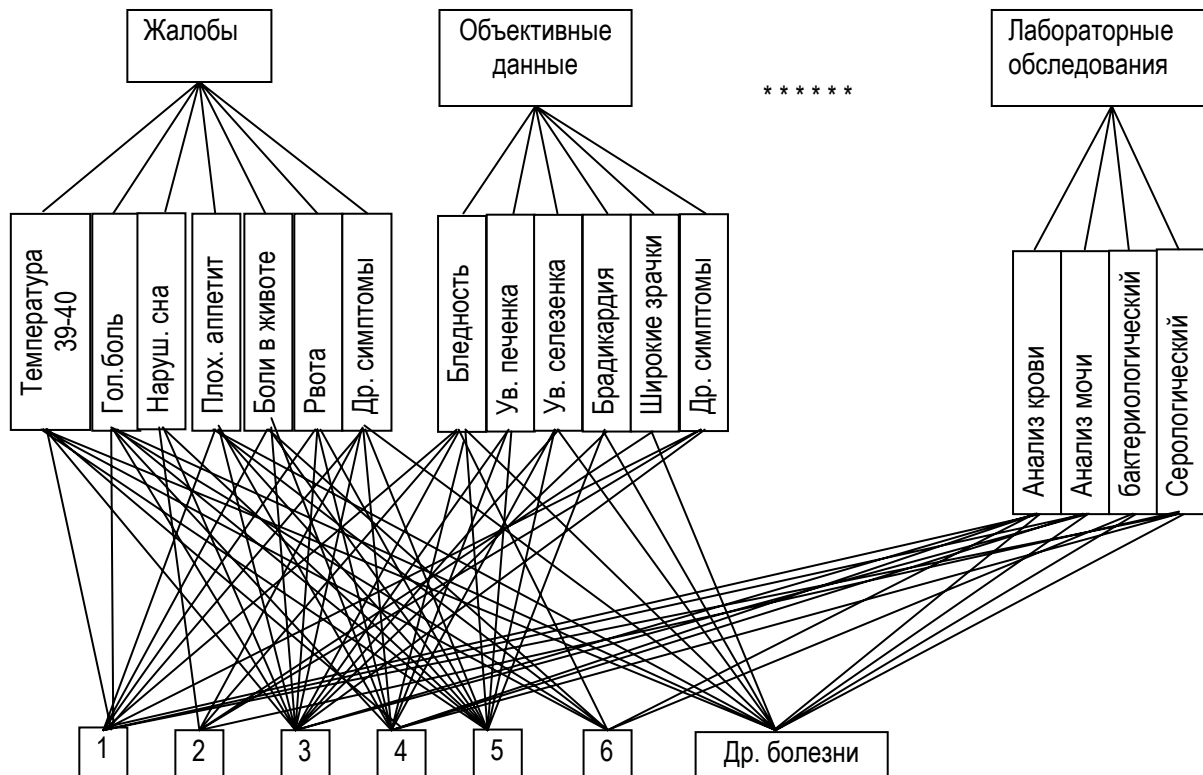
Одной из первых прикладных областей применения методов искусственного интеллекта явилась медицинская диагностика [Люгер, 2003], [Рассел, 2006]. Разработка экспертных систем диагностирования заболеваний насчитывает более 50 лет. Хотя теория принятия решений является стандартным инструментальным средством во многих проблемных областях таких, как бизнес, государственное управление, юриспруденция, военная стратегия, техническое проектирование и управление ресурсами, лишь немногие исследователи в области искусственного интеллекта [Рассел, 2006, с.810] приняли на вооружение инструментальные средства теории принятия решений в медицинской диагностике. Одной из основных причин ограниченного применения деревьев решения в медицине являются их «экспоненциальный размер» [Рассел, 2006]. Другим критерием, на наш взгляд (см. [Волошин, 2006]), является некорректность использования «усредненных» экспертных знаний. Во многих случаях диагностирования заболеваний отсутствуют «объективные» («общепринятые») критерии оценки интенсивности взаимовлияния факторов, определяющих заболевание. Процесс принятия решения врачом-диагностом до сих пор в определяющей степени является субъективным, в значительной степени зависящем от «интуиции», «опыта» и тому подобных слабо формализованных факторов. И хотя теперь, когда канонической программой искусственного интеллекта стал «интеллектуальный агент», предназначенный для помощи человеку [Рассел, 2006, с.1267], и о замене человека, принимающего решения, речи не идет, роль экспертной системы свелась к подобию медицинских учебников и справочников [Рассел, 2006, с.1269]. Врач обязан понять ход рассуждений, лежащих в основе любого решения системы. Иначе использование систем искусственного интеллекта может привести к тому, что люди станут более безответственными (кто будет нести юридическую ответственность, если диагноз окажется неправильным?). Поэтому в [Волошин, 2006] предлагалось перейти от концепции «разработки экспертной системы как «помощника» («интеллектуального усилителя») лица принимающего решения к концепции «персонального инструментария». А для этого необходимо положить в основу системы способ принятия решения именно пользователя системы. Разработчик же системы должен предоставить средства формализации этого способа, при этом для «объективизации» субъективных оценок лица,

принимающего решение, закладываемых в систему, необходимо учитывать психосоматические особенности лица, принимающего решения (в системах экономического прогнозирования [Волошин, 2005] «объективизировалась субъективность» экспертов-разработчиков системы). Более того, необходимо предоставить средства учета субъективных особенностей диагностируемого субъекта. Как отмечалось неоднократно в истории искусственного интеллекта, одной из основных сложностей разработки систем искусственного интеллекта является «извлечение» знаний у специалиста. Задача многократно усложняется, если необходимы не «стационарные» знания, а «динамические» («способ мышления»).

В данной работе отражен опыт авторов по общению со специалистами-диагностами при разработке диагностических систем как персонального интеллектуального инструментария, реализующих «способ внедренного действия» [Люгер, 2003] в виде нечетких деревьев решений [Волошин, 2003]. При этом в отличие от большинства диагностических систем, в которых результат диагностики основывается на критерии «правдоподобия» [Рассел, 2006], одновременно с оценкой дуг деревьев (взаимовлияния факторов), задаются веса (полезность) самых факторов. Последнее в определенной степени позволяет избежать одной из наиболее распространенных «ловушек» в диагностике – смешивание правдоподобия и важности [Рассел, 2006, с.804].

Построение деревьев решения

Метод дерева решения основан на построении иерархической структуры факторов, которые имеют прямое и не прямое влияние на диагноз. В верхней части дерева решений собраны основные, важнейшие факторы. Далее для этих факторов определяются подпроблемы, которые на них влияют. Таким самым образом обрабатываются выделенные подпроблемы и т.д. В листьях этого дерева собраны факторы, для которых подпроблемы уже не определяются. Для наглядной иллюстрации на рис.1 изображен фрагмент дерева решений для прогнозирования инфекционных болезней кишечной группы.



1 – сальмонеллез, 2 – ботулизм, 3 – гепатит А, 4 – лептоспироз, 5 – брюшной тиф, 6 – дизентерия.

Рис. 1.

Построение дерева решений сводится к выделению экспертами проблем и подпроблем (вершин дерева) и связей между ними (дуг дерева). Далее эксперты определяют важность (вероятности) переходов между вершинами. Допускается нечеткая оценка экспертов с помощью логических переменных, описанных значением функции принадлежности (векторами действительных чисел от 0 к 1). Каждый эксперт дает три оценки – оптимистичную, реалистичную и пессимистичную, скаляризация которых происходит с учетом психологического типа эксперта.

Дерево строится на основе коллективных оценок экспертов с использованием метода парных сравнений. Для построения «результатирующего» дерева используется алгебраические методы обработки экспертной информации.

Для определения оптимальных путей в дереве решений используется алгоритм последовательного анализа вариантов, что позволяют обрабатывать деревья с сотнями вершин [Voloshin, Pancheko, 2001].

Дерево решений задается таблицами. Каждая таблица – отдельный уровень дерева, каждая строка таблицы – отдельная вершина на этом уровне. Каждый элемент строчки – вероятность переходы с заданной вершины в вершину верхнего уровня. Эти вероятности задаются функциями принадлежности. Таблицы заполняются путем опроса экспертов.

Экспертным путем задаются матрицы – результат сравнения вариантов вершин, которые могут быть включены в дерево. На основе анализа матриц определяются вершины, которые включаются в дерево и вероятности, с которых возможен переход в них с вершин верхнего уровня.

Дерево решений рассматривается как граф (рис. 2), строится матрица инцидентности вершин графа, в которых находится информация о весах каждой вершины дерева решений.

На рис. 3 изображено дерево диагностики мочекаменной болезни.

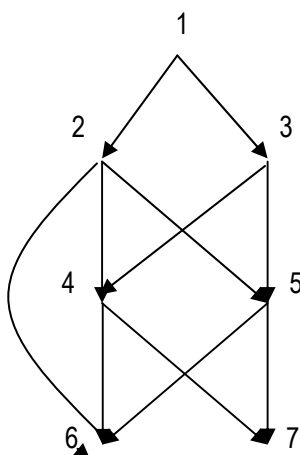


Рис. 2.

Описание системы

Во время принятия решения (установления диагноза) перед врачом ставится задача обработки большого массива за малый период времени, что является одним из составных успешного диагностирования и как следствие, успешного лечения.

В отличие от рассматриваемых ранее случаев [Волошин, Головня, 2005] дерево решений для диагностирования инфекционных заболеваний имеет многоуровневую структуру. Это связано с тем, что инфекционные заболевания делятся на группы, в группах собраны болезни, но некоторые болезни собраны в блоки.

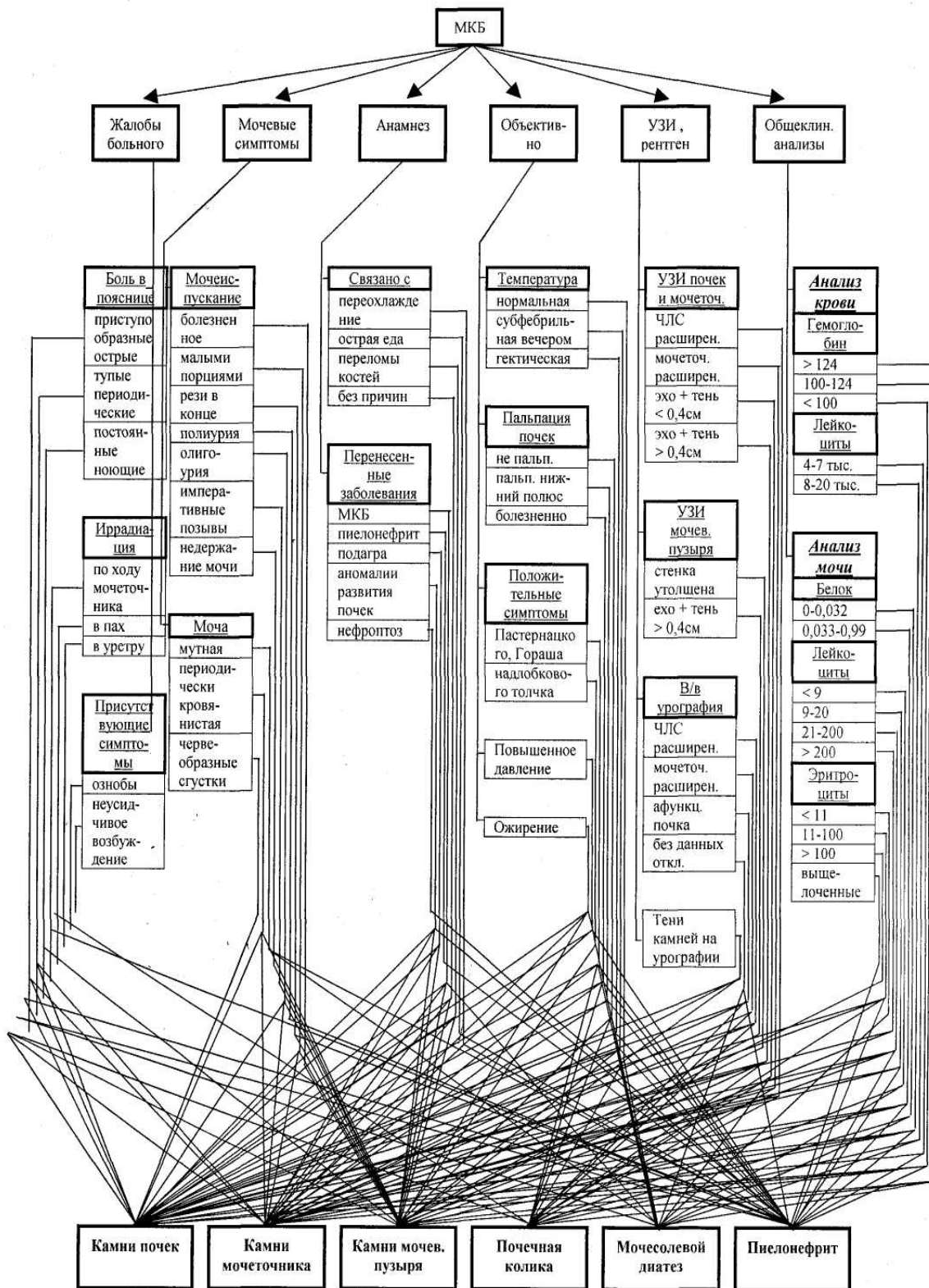
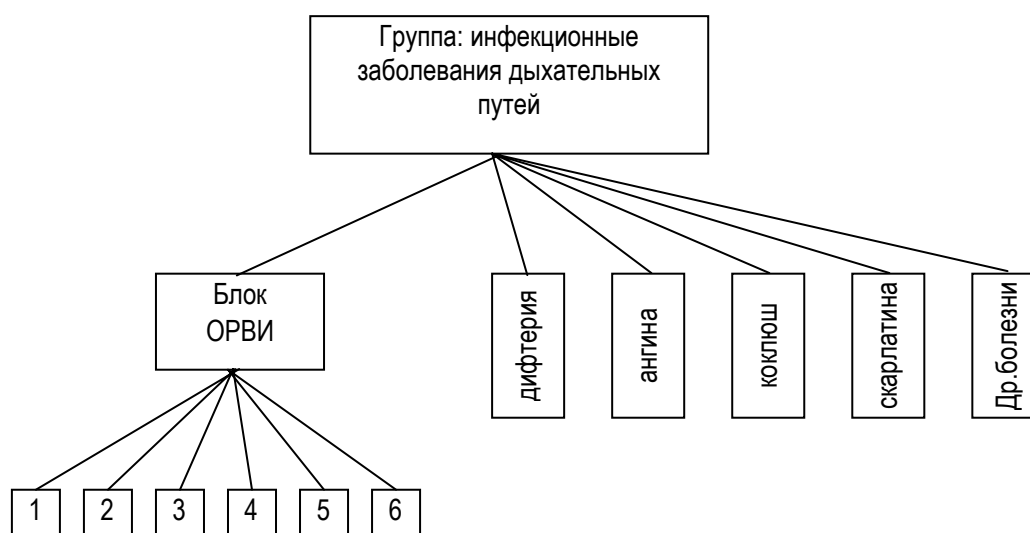


Рис. 3.

Инфекционные болезни делятся на пять групп: кишечные инфекционные болезни, инф. болезни дыхательных путей, кровяные (трансмиссивные) инф. болезни, инф. болезни наружных покровов и инф. болезни с парентальным механизмом передачи (инъекционным). В группах собраны диагнозы, но некоторые диагнозы собраны в блоки за подобием симптомов, методов распространения и подобностью клинической картины. Блок, например, ОРВИ (острые респираторные вирусные болезни) содержит в себе такие болезни, как: грипп, парагрипп, аденовирусная болезнь, риновирусная болезнь, респираторно-сенцитиальная болезнь, реовирусная болезнь (рис.4). Для каждой группы построено соответствующее дерево решений, основное дерево решений состоит из нескольких поддеревьев.

Такое распределение, в какой то мере - борьба с «проклятием размерности». Эта проблема возникает в процессе последовательного перебора вариантов, которые максимально будут соответствовать заданным, так как каждой болезни соответствует определенный набор симптомов. Эти симптомы могут частично, но не полностью, повторяться. Каждой болезни соответствует характерная черта, которая относит ее к соответственной группе инфекционных заболеваний, тем самым уменьшая объем поиска.



*) 1 - грипп, 2 - парагрипп, 3 - аденовирусная болезнь, 4 - риновирусная болезнь, 5 - респираторно-сенцитиальная болезнь, 6 - реовирусная болезнь

Рис. 4.

Как известно, в процессе становления диагноза врач на 90%, а иногда и на 100%, полагается на результаты лабораторных исследований.

Технический прогресс и научные достижения на сегодняшний день позволяют с помощью лабораторных исследований определить возбудитель болезни и саму болезнь на 100%. Но перед врачом стоит проблема назначить соответственное лабораторное обследование. Для этого ему необходимо проанализировать симптомы, найденные им во время осмотра, и жалобы на состояние здоровья пациента. Поэтому в дереве решений есть поддерево, где определяются назначения необходимых лабораторных обследований.

Для получения более точной информации, которая поступает от пациента в виде жалоб, врач должен быть еще и психологом. Он должен определить психологический тип пациента, например, правдивость и объем предоставляемой информации о жалобах на здоровье.

Алгоритм работы системы

В первую очередь врачу нужно определить метод лабораторного обследования, результаты которых вводятся в систему, устанавливается важность для каждого симптома. Программа анализирует дерево решения, которое создано для определения методов лабораторного обследования и выводит результаты. После анализа полученных результатов в систему вводятся данные с новыми важностями и критерием отбора. Критерием отбора в данном случае выступает сам возбудитель болезни, который определяет группу инфекционных заболеваний, который свойственен если не одной, то нескольким болезням одной группы или блока. Данный критерий обеспечивает уменьшение временных расходов уменьшением масштабов поиска. Система анализирует дерево решений данной группы, которая указана в критерии отбора и выводит результат. Если результатом есть блок (смотри рис.4), то система продолжает свою работу, анализируя только этот блок с самого начала. Но возможны и варианты, когда данные относятся к разным группам, в таком случае система анализирует соответственное дерево за данными одной группы, после этого - за данными, которые относятся к другой. Основной алгоритм работы системы изображен на рис. 5.

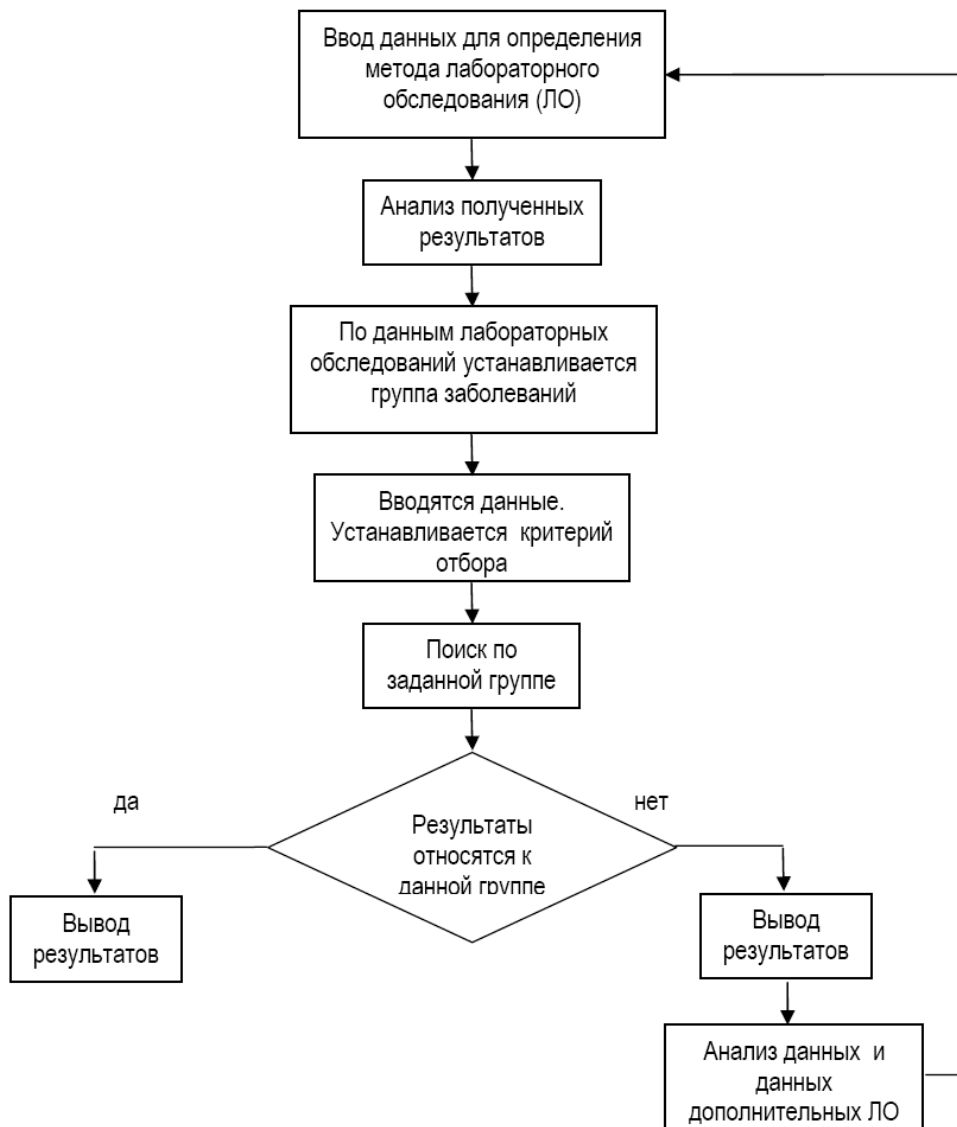


Рис. 5.

Метод отсеивания вариантов

Если при этом множество возможных вариантов сравнительно невелико, то искомым результатом система может получить путем прямого перебора. Однако с увеличением числа вариантов эта возможность практически исчезает даже при использовании ЭВМ.

Таким образом, возникает необходимость в применении общих методов целенаправленного перебора, позволяющих генерировать искомые данные за приемлемое время. Одним из подходов является метод последовательного анализа вариантов. В основу этого метода положена идея представления процесса решения в виде многоступенчатой структуры. Каждая ступень связана с проверкой наличия тех или иных свойств у подмножества вариантов (или отдельных вариантов) и ведет либо к непосредственному сокращению исходного множества вариантов, либо подготавливает возможность такого сокращения в будущем. На основе теоретического и практического анализа поставленной задачи необходимо формализовать те отличительные свойства, которыми должен обладать искомым вариант. Затем нужно выявить по возможности больше признаков, позволяющих установить, что данный вариант не является искомым. Среди этих признаков выбираются наиболее легко проверяемые и присущие наибольшему числу вариантов одновременно. После этого выбор численной схемы решения состоит в задании рационального порядка проверки признаков, позволяющих провести отсев неконкурентоспособных вариантов и найти оптимальный.

С точки зрения формальной логики схема последовательного анализа вариантов сводится к повторению следующей последовательности действий:

- разбиение множества вариантов решений задачи на семейство подмножеств, каждое из которых обладает дополнительными специфическими свойствами;
- использование этих специфических свойств для поиска логических противоречий в описании отдельных подмножеств;
- исключение из дальнейшего рассмотрения тех подмножеств вариантов, в описании которых имеются логические противоречия.

При этом методика последовательного анализа конструирования и отсеивания вариантов заключается в таком построении вариантов и выборе операторов их анализа, которые позволяют отсеивать бесперспективные части вариантов без их полного построения — по мере того как эту бесперспективность удается обнаружить. Поскольку при отсеивании бесперспективных частей вариантов отсеивается тем самым все множество его продолжений, то происходит значительная экономия в вычислительной процедуре, которая является тем более существенной, чем больше специфических свойств задачи использовано для построения операторов анализа и отсева [Поспелов, 1980], [Моисеев, 1971].

В начале основным правилом отсева бесперспективных вариантов был принцип монотонной рекурсивности, родственной критерию оптимальности динамического программирования [Беллман, 1960]. На основе этого принципа были построены алгоритмы пошагового конструирования вариантов для решения различных задач дискретной оптимизации [Михалевич, Шкурба, 1966].

Наряду с известными достоинствами алгоритмы пошагового конструирования решений обладают и определенными недостатками. Так, они, как правило, предъявляют чрезмерные требования к оперативной памяти ЭВМ и обуславливают с ростом числа ограничений задачи резкое увеличение объема вычислительной работы для поиска оптимального решения. Эти факты подтверждаются как вычислительными экспериментами, так и теоретическими оценками.

Вместе с тем сама методология последовательного анализа вариантов позволяет строить общие схемы решения дискретных оптимизационных задач, отличные от динамических (т. е. основанных на пошаговом конструировании решений). Отказ от идеи пошагового конструирования решений при решении задачи методом последовательного анализа и отсеивания вариантов приводит к необходимости организовать процедуры анализа подвариантов. Тем самым исчезает необходимость в выборе «начала» развития

частичных решений и устраняется «несимметричность» в анализе компонент решения. Устраняется также необходимость запоминания на каждом шаге множества «недоминируемых» частичных решений, подлежащих развитию на следующем шаге [Волкович, Волошин, 1984].

В основу методов анализа поддеревьев в системе заложен описанный выше декомпозиционный метод поиска альтернативных вариантов [Волошин, 1987].

Заключение

Авторам неизвестны факты применения современных достижений теории принятия решений в медицине, хотя за применение их в других областях, в первую очередь, в экономике, за последние годы получены две Нобелевские премии.

Реализуя концепцию взаимодействия «объект - способ действия - субъект» [Волошин, 2006] целесообразно рассматривать задачу установления диагноза как задачу коллективного принятия решений [Волошин, Мащенко, 2006], в которой «рациональными» агентами являются – «пациент – экспертная диагностическая система – врач». В некоторых случаях целесообразно, по-видимому, расширить число агентов, добавив, например, «фармацевта» (преследующего свои цели, зачастую антагонистические пациенту). А тут возможны уже совсем другие «принципы оптимальности», отличные от «максимизации полезности с учетом вероятности» [Рассел, 2006].

Не исключено, что компромисс в «жизни» достигается как реализация «принципов оптимальности Нэша и Парето» [Волошин, Мащенко, 2006] – выбирается решение (сознательно или неосознанно), от которого невыгодно отклоняться как всем агентам вместе, так и каждому в отдельности.

Благодарность

Авторы благодарны консультантам-экспертам: врачу-инфекционисту М. Б. Запорожец и врачу-урологу П. М. Мулесе.

Список литературы

- [Беллман, 1960] Беллман Р. Динамическое программирование. – М.: Изд-во иностр. лит., 1960. – 400 с.
- [Волкович, Волошин, 1984] Волкович В.Л., Волошин А.Ф., Горлова Т.М. и др. Методы и алгоритмы автоматизированного проектирования сложных систем управления. – К.: Наук. думка, 1984. – 216с.
- [Voloshin, Pancheko, 2001] Voloshin O.F., Pancheko M.V. the Forecasting of Stable Processes by a Tree Solution Method using a Pairwise Comparison Method for Analysis of Expert information. Труды международной конференции «KDS-2001», Том 1, Санкт-Петербург, 2001. -С.50-53 (англ.яз).
- [Волошин, 1987] Волошин А.Ф. Метод локализации области оптимума в задачах математического программирования // Доклады АН СССР.-1987.-Т. 293, №3.-С. 549-553.
- [Волошин, 2006] Волошин А.Ф. Системы поддержки принятия решений как персональный интеллектуальный инструментальный лица, принимающего решения. Труды конференции «KDS-2006», София, 2006. -С.149-153.
- [Волошин, Мащенко, 2006] Волошин А.Ф., Мащенко С.О. Теория принятия решений. – Киев: КНУ, 2006.-304с. (укр.яз).
- [Волошин, Панченко, 2002] Волошин А.Ф., Панченко М.В. Экспертная система качественного оценивания на основе многопараметрических зависимостей. «Проблемы математических машин и систем», 2002, №2.-С.83-89 (укр.яз).
- [Люгер, 2003] Люгер Ф.Дж. Искусственный интеллект. Стратегии и методы решения сложных проблем. Москва: «Вильямс», 2003. 264с.
- [Михалевич, Шкурба, 1966] Михалевич В.С., Шкурба В.В. Последовательные схемы оптимизации в задачах упорядочения выполнения работ. // Кибернетика, 1966, №2,-С. 34-40.
- [Моисеев, 1971] Моисеев Н.Н. Численные методы в теории оптимальных систем. – М.: Наука, 1971. – 434 с.
- [Поспелов, 1980] Поспелов Г.С. Системный анализ и искусственный интеллект. – М.: Изд. Вычисл. Центра, 1980. - 46с.
- [Рассел, 2006] Рассел Ст., Норвиг П. Искусственный интеллект: современный подход, 2-е изд.: Пер. с англ.-М.: «Вильямс», 2006.-1408с.

Сведения об авторах

Волошин Алексей Федорович – Киевский национальный университет имени Тараса Шевченко, факультет кибернетики, профессор. Киев, Украина. E-mail: ovoloshin@unicyb.kiev.ua

Запорожец Максим Владимирович - Киевский национальный университет имени Тараса Шевченко, факультет кибернетики, аспирант. Киев, Украина.

Мулеса Павел Павлович – Ужгородский национальный университет, математический факультет, аспирант. Ужгород, Украина. E-mail: ppmulesa@mail.ru

DECISION SUPPORT SYSTEM IN ULTRASOUND INVESTIGATIONS

Svetlana Cojocar, Constantin Gaidric

Abstract. A decision support system SonaRes intended to guide and help the ultrasound operators is proposed. The system is based on rules and images and can be used as a second opinion in the process of ultrasound examination.

Keywords: decision support systems, knowledge acquisition, image processing, ultrasound examination.

Introduction

We will describe a decision support system SonaRes, intended to support ultrasound diagnostics. The system plays a consultative role and offer to users its variants of diagnosis. System's solutions are motivated by presenting method(s) of its obtaining and the corresponding images that help to understand its reasons.

The problem of assuring an adequate medical assistance to the population depends both on training and qualification levels of medical personnel and on performance of the used diagnostic equipment. Nowadays it is impossible to offer medical services, even at the most modest level, without using of medical equipment, apparatuses, devices and technical complexes. Applications of medical equipment are very diverse and include diagnostics, treatment, supervision, compensation of a lesion or a handicap, etc.

There is no doubt that technical assistance in medicine is as important as in pharmaceutical assistance. The technical assistance in medical examinations is impossible without a strong qualification of a staff that exploits this equipment and interprets the obtained information. The quality of medical services also directly depends on how correctly and efficiently the medical diagnostic and treatment/recuperation equipment is used.

Ultrasound equipment is much cheaper than the MRI - Magnetic Resonance Imaging, CT – Computer Tomography, digital radiography etc, and its rational and efficient use could fill in many gaps in medical diagnosis. Technical progress of the last years in the field of ultrasound diagnostics allowed these methods to come to a leading position among imagistic procedures.

Echography has proved to be one of the most usable and beneficial paraclinic investigations as it is non-invasive and extremely efficient, has a great accuracy in its area of application, and is executed easily by a well-trained specialist. Despite the enthusiasm it meets, the echography has its real limitations like any other procedure. They are expressed sometimes through false-negative or false-positive images; sometimes these limitations are imposed by an examining physician's ability to obtain qualitative images or to interpret them.

In order to obtain quickly correct information on the specific case based just on images and their descriptions, it is important to create a unified system that will allow storing of the images and their annotations. Special techniques should be developed to annotate images. This collection of images and annotations will help ultrasound technicians to justify their final conclusions.

The primary use of the system might be as a 'second opinion' in difficult cases and in emergency; it does not replace physician who interprets echograms. Thus, SonaRes is destined to improve health care by providing a highly efficient diagnostics tool. The tool is well-suited to needs and current state of the medical equipment in hospitals and clinics.

A comparison with existing systems

Presently, there are only a few decision support systems for ultrasound diagnostic in the world, even though the attempts to elaborate systems of the kind have been started since 1970's. We can refer to publications in the mentioned field [Diez, 1997-Cabinet].

Some of the most known systems created for ultrasound investigation we can mention are DIAVAL [Diez, 1997] designated for echocardiography, ProtolSIS [Anderson, 1994, Kahn] (ultrasound and computerized tomography), SonoConsult [Huettig, 2004] et al.

Till now two basic approaches were applied to development of computer assisted ultrasound diagnostic systems:

a) Systems based on image analysis and classification.

Such systems pursue the purpose to make the decision on the basis of comparison of the initial images with those from a database (DB). The comparison helps to classify the available image according to classification existing in a database and/or give the user an opportunity to define a degree of its similarity with images available in a database. In the case of detection of similarity to any image (precedent) the decision based on already known decision for the existing precedent is given out.

b) Systems based on rules.

Such systems pursue the purpose to make the decision on the basis of the description obtained from the user and the data available in system and rules. More often such systems serve for information or training purpose with or without an additional diagnostic component.

In ProtolSIS the classification of images is made on the basis of 4 sets consisting from 25 precedents each. The probability of correct classification makes 72-84%.

Better results have been received in the system described in [Huo, 2001 - Drukker, 2004] concerning the domain of computerized detection and classification of cancer on breast ultrasound. A two-stage computerized method has been developed: the detection stage and classification stage. At the first (detection) stage the suspicious regions on ultrasound images are detected and subsequently distinguished among different lesion types. After the detection stage all candidate lesions are classified by a Bayesian neural net, based on computer-extracted lesion features. Two separate tasks are performed and evaluated at the classification stage. The first classification task is the distinction between all actual lesions and false-positive detection; the second classification task is the distinction between actual cancer and all other detected lesion candidates (including false-positive detection's).

First stage, gives the performance values of 94% and 91% for training and testing data sets respectively. Second stage, based on candidates' lesion classification, gives the performance values of 87% and 81% for training and testing data sets.

SonoConsult is a knowledge-based system which uses simple and complex rules to make a required decision, promotes completeness and carefulness of input of patient's state, and thus helps to minimize an opportunity of reaching the erroneous judgment.

The following table contains the main features to be incorporated in the developing system and points those features which reside in the existing systems.

	SonoConsult	Bayesian Network	ProtolSIS	LookInside	tCaUD	SonaRes
Use of both the images and their descriptions	–	X	–	X	–	X
The interactive interface for knowledge acquisition	–	–	–	–	–	X
The intelligent interface	–	–	–	–	–	X
Expertise reporting	X	–	–	–	X	X
Explanation of the decision	–	–	–	–	–	X
Possibility of adding to knowledge base on the basis of precedents	X	X	X	X	–	X
Examination of the organs interaction	–	–	–	–	–	X
Image processing	–	–	X	–	–	X
The standardized descriptions and decisions	X	–	X	X	X	X
Possibility to use the system in automated learning	X	X	X	–	–	X
Treating of patient's state in dynamics	–	–	–	–	X	–

One can compare five best and well-known existing systems using the data from the table:

- SonoConsult [Huetig, 2004], which is an expert system for structured and case-adequate documentation of sonographic findings with an additional diagnostic component;
- Bayesian Network [Haddawy, 1994] is a technique for reasoning under uncertainty, currently is being developed for application to medical decision making;
- ProtolSIS [Anderson, 1994, Kahn] is a decision support system, based on comparison of initial researched image with the images from the images' database for the purpose of similarity determination;
- "LookInside" [capabilities] is designed to operate with database of patients amenable to ultrasound examination;
- "The Cabinet of Ultrasound Diagnosis" (tCaUD) [cabinet] is also designed to operate with database of patients amenable to ultrasound examination.

From the table one can determine good characteristics and weak points of analyzed systems. We can see that no one system is taking into account interaction between organs. The system SonoConsult, which shows some performance results during its utilization, and is very appreciated by the specialists-physicians, has the most similar aim.

Main goal and system structure

Our goal is to develop an approach which includes interaction between organs and uses current and precedent similar images in decision making process. Special attention is paid to ergonomic user interface, which is generated dynamically by system according to the DB content and is adaptable to preferences and objectives (of investigation type) of the physician-echographer.

We will offer to specialist, even without wide experience, an access to a resource where the process of ultrasound examination is detailed and formalized and includes an enormous amount of useful information on anatomy, ultrasonic semeiology, differential diagnostic as well as condensed presentation of the main nosologic entities that should appear in the physician's mind at the moment of examination of each organ.

This system SonaRes help the specialist in ultrasonic analysis to draw the conclusion more correctly, especially, in emergency cases or in unspecific clinic/paraclinic cases, which do not seem to be included in any classical presentation; in cases where the obtained ultrasonic semeiology can provide a correct diagnosis without complicated and, often difficult of access, medical investigations.

SonaRes offer to a user a second opinion with necessary explanations and images that are similar to the examined case. Images can be processed and problem zone, if it is necessary for the user-physicist, can be marked out.

The main components of the system are the following:

- Knowledge acquisition
- Examination support
- Unified database (knowledge, images, annotations etc.)
- Image processing algorithms
- Reports generator

In order to develop these components we are elaborating and adapting:

- formalized descriptions of the abdominal organs, pathologies, anomalies;
- formalized descriptions of the ultrasound investigations methodology;
- unified, standardized disease descriptions;
- knowledge acquisitions methods based on ultrasound investigations characteristics;
- a diagnostics validation tool;
- a database model for the medical images, their annotations and fuzzy information storage;
- images clusterization and quick database searching algorithms;
- an ergonomic, dynamically generated and user friendly interface;
- reports' prototypes and their generator.

At the first stage we deal with abdominal zone investigation. The investigation process of this zone is especially difficult (more organs with additional interactions, higher level of confusion, etc.). We have approved our technique on gall bladder and extend it on other organs.

Knowledge structure modeling

As models of knowledge representation in the medicine domain a model based on rules or a semantic network usually are chosen. In both cases the problem is reduced to [Secrieru, 2007, Popcova, 2006] :

- determination of objects, concepts and their attributes which are used in the given problem area;
- definition of links between concepts;
- determination of metaconcepts and detailed elaboration of concepts;
- construction of the knowledge pyramid, being scale of metaconcepts ranks, rising on which means the deepening in understanding and increasing the level of metaconcepts generalization.
- validation of rules representing knowledge.

Knowledge validation

A set of trees representing decisions rules are constructed based on data tree structure and using knowledge about organ's pathologies and anomalies. These trees contain all the necessary factors which can help to

produce a conclusion. The validation goal is to evaluate the knowledge base correctness and completeness. One can obtain this information by testing the existing rules. The testing is performed by physicians, in order to do this work in more efficient way a validation tool was elaborated, which permits to simulate the examination process and evaluate the obtained conclusions.

The tool offers the possibility to obtain various conclusions modifying the attributes values. During the validation process the current session containing all the selected attributes values is saved. Thus to simulate a new examination which differ from the previous one just by one attribute it is sufficient to modify only one value. Also it is easy to establish the list of rules containing a specific combination of attribute-value [Jantuan, 2007].

Image processing

Non-using of images during decision-making process can lead to the loss of the valuable information and does not correspond to daily practice of the doctor. Therefore we use a collection of annotated images stored in Image DB, which can be used as illustrations in similar disease cases. Ultrasound images from DB will be preprocessed in interesting zones to facilitate their clusterization (this is necessary for quick and relevant search).

Image preprocessing will be performed using existing algorithms (image processing algorithms, adapted to ultrasound diagnostic methodology), as well as by using the new ones (based on heuristics, fuzzy, pattern-oriented filters etc.). The following image processing methods [Dzung, 1998 - Popcova, 2004] were applied: statistical treatment, noise reduction, contrast adjustment, borders and organ contours determination etc. and their combination to obtain more efficient result.

Examination support

The proposed method of acquisition (by means of expert shell) and storage of expert knowledge in Unified DB permits to effectuate a quick search of necessary information in two directions or modes. The first direction is from the concrete case description to determination of pathology and/or an anomaly; and the second one – from formulation of a hypothesis to its confirming or denying.

Following the first direction the user gives the necessary information describing a concrete case, and the system tries to determine if it is a pathology and/or an anomaly. To exclude at the early stage the input of inconsistent, erroneous or excessive information, this direction is followed step-by-step. If at any step the system can determine, on the basis of the entered information, a pathology and/or an anomaly, it informs the user.

Following the second direction, the user forms a hypothesis about presence in the concrete case of pathology and/or an anomaly. Then the system by means of additional questions tries to confirm or to deny this hypothesis.

Realization of both modes within the framework of unified support system of ultrasonic investigation process corresponds to the daily work of physicians. The first operating direction satisfies the requirements of the detailed patient examination; and the second direction corresponds to a simplified one, when it is necessary to confirm or to deny any diagnosis

A convenient dialog with user-physician (due to dynamic intelligent interface which includes a standardized explanation of the decision proposed by system) involving images in decision making process (based on visualization and comparison of ultrasound examined image with similar images from Image DB) permits to create a comfortable environment for physicians and helps him to prepare a standardized report, containing the examination results and, if necessary, the recommendations for additional investigation.

Conclusion

The proposed system does not intend to replace completely the physician; it just offers him a second opinion. In all cases user can receive all rules and judgments on the basis of which the decision was made. If the user doesn't agree with the decision, proposed by the system, his opinion will be sent to expert group for examination.

Acknowledgments

This work is supported by the STCU, project ref. 4035.

Bibliography

- [Diez, 1997] F.J.Diez, J.Mira, E.Iturraalde, S.Zubilaga. DIAVAL, A bayesian expert system for echocardiography. – Artif. Intell. Med. n.10, 1997, pp.59-73.
- [Anderson, 1994] Kahn C.E. Jr., Anderson G..M.Case-based reasoning and imaging procedure selection. Investigative Radiology 1994; 29:643-647., <http://www.mcw.edu/midas/old-papers.html>)
- [Kahn] Charles E.Kahn, Jr. Clinical Trial and Evaluation of a Prototype Case-Based System for Planning Medical Imaging Work-up Strategies. www.mcw.edu/midas/papers/ISIS.CBR94.ps
- [Huettig, 2004] M.Huettig, G.Buscher, T.Menzel, W.Scheppach, F.Puppe, H.P.Buscher. A Diagnostic Expert System for Structured Reports, Quality Assessment, and Training of Residents in Sonography. Medizinische Klinik, 2004, 99, nr.3, p117-122.
- [Huo, 2001] Huo Z, Giger M.L., Vyborny C.J.: Computerized analysis of multiple-mammographic views: Potential usefulness of special view mammograms in computer-aided diagnosis. IEEE Transactions on Medical Imaging 20: 1285-1292, 2001.
- [Horsch, 2004] Horsch K, Giger ML, Vyborny CJ, Luz V: Performance of CAD in the interpretation of lesions on breast sonography. Academic Radiology 11: 272-280, 2004.
- [Drukker, 2004] Drukker K, Giger M.L., Vyborny CJ, Mendelson EB: Computerized detection and classification on cancer on breast ultrasound. Academic Radiology 11: 526-535, 2004.
- [Haddawy, 1994] Haddawy P, Kahn CE Jr, Butarbutar M. Bayesian network model for radiological diagnosis. Medical Physics 1994; 21:1185-1192.
- [capabilities] <http://www.lins.ru/capabilities.shtml>
- [cabinet] <http://cabinet.fromru.com/>
- [Secieru, 2007] Iu.Secieru, O.Popcova, S.Puiu, D.Sologub. Knowledge Structure Modeling in Ultrasound Investigation Domain. Bit+. The 7th International Conference “Information Technologies 2007”, 24 -28 April, 2007, Chisinau, Republic of Moldova (to appear).
- [Popcova, 2006] O.Popcova, I. Secieru, D.Sologub, E.Jantuan, V.Papanaga. Decision Support System for Ultrasound Diagnostics // Proc. of the First International Conference of Yong Scientists Computer Science&Engeenering (CSE 2006) – Lviv, 2006 – p.30-31.
- [Jantuan, 2007] E.Jantuan. Knowledge validation tools. Bit+. The 7th International Conference “Information Technologies 2007”, 24 -28 April, 2007, Chisinau, Republic of Moldova (to appear).
- [Dzung, 1998] Palm Dzung L., Xu Chenyang, Prince Jerry L. A survey of current methods in medical image segmentation // Technical report. – Johns Hopkins University, Baltimore, 1998.
<http://www.csem.duke.edu/seminars/segmenation.pdf>
- [Jardim, 2002] Jardim Sandra Vilas Boas, Figueiredo Mário A. T. Automatic analysis of fetal echographic images // Proc. Portuguese Conf. on Pattern Recognition. – Aveiro: RecPad, 2002.
- [Dinggang, 2003] Dinggang Shen, Yiqiang Zhan, Davatzikos C. Segmentation of prostate boundaries from ultrasound images using statistical shape model // IEEE Transactions on Image Processing. – 2003. – vol. 22, № 4. – P. 539-551.
- [Popcova, 2004] O.Popcova, S.Cojocar, C.Gaindric. Image Processing in Ultrasound Diagnostic System. 8th International Symposium on Automatic Control and Computer Science, Iasi, Romania, October 22-23, 2004.

Authors' Information

Svetlana Cojocar – Dr.hab., Institute of Mathematics and Computer Science, Academy of Sciences of Moldova, Academiei str., 5, Chisinau, Republic of Moldova, e-mail: sveta@math.md

Constantin Gaindric - Dr.hab., Prof., Institute of Mathematics and Computer Science, Academy of Sciences of Moldova, Academiei str., 5, Chisinau, Republic of Moldova, e-mail: gaindric@math.md

ЭВОЛЮЦИОННОЕ МОДЕЛИРОВАНИЕ ПРОЦЕССА РАСПРОСТРАНЕНИЯ ПОЖАРА

Виталий Снитюк, Артем Быченко

Аннотация: В статье предложена эволюционная технология для определения времени и пути распространения пожара от точки его возникновения до особо опасного объекта. Она базируется на продукционных правилах, определяемых экспертным путем. Ядром такой технологии является использование методов *Soft Computing* для идентификации времени распространения и пути огня, а также оптимизации параметров соответствующих зависимостей. Рассмотрены аспекты решения указанной задачи с помощью нейронечеткой сети и эволюционных методов.

Ключевые слова: Эволюционное моделирование, пожары, особо опасные объекты.

Введение

Современные исследования процессов пожаротушения методами математического моделирования в основном посвящены исследованию типовых ситуаций. В частности – пожаротушению в жилом секторе, распространению пожаров в лесных массивах, разработке эффективных технических средств пожаротушения. В то же время исследованию динамики пожаров на особо опасных объектах не уделяется должного внимания, несмотря на то, что в последнее время промышленный рост сопровождается увеличением количества экологических и промышленных катастроф и аварий. Такое состояние можно объяснить исключительным характером чрезвычайных происшествий, стоимостью моделирования уникальных ситуаций, неоправданной уверенностью руководства и персонала предприятий в их безопасности. Важно заметить, что возможные потери от возникновения критических аварийных ситуаций многократно превышают стоимость информационно-аналитических систем, с помощью которых можно осуществлять прогнозирование динамики развития пожаров и объективизировать процессы принятия решений, базируясь на анализе субъективных суждений экспертов.

Определенные результаты в указанном направлении были получены в работах [Быченко, 2006], [Снитюк, 2006]. Так, в статье [Быченко, 2006] предложен первый шаг реализации информационно-аналитической технологии, заключающийся в разработке процедуры формирования модели времени распространения пожара, с помощью которой можно максимально близко моделировать процесс принятия решений человеком. Выполнена структурная идентификация и инициализация модели времени распространения пожара. В работе [Снитюк, 2006] предложен подход к формированию указанной модели, базирующийся на использовании нечетких баз знаний и соответствующих методов и направленный на реализацию информационно-консультативной геоинформационной системы, которая позволит осуществлять прогнозирование процесса развития пожара в пространственно-временной системе координат.

Очевидно, что мощность банка моделей процесса распространения пожара превышает единицу. Формирование, исследование и использование модели определяется двумя главными факторами: соответствующим критерием ее оптимальности и особенностями внешней среды. Заметим, что такой критерий носит субъективный характер и определяется экспертным путем либо лицом, принимающим решение (ЛПР).

Постановка задачи

Химические промышленные предприятия и предприятия энергетики имеют на своей территории особо опасные объекты, возникновение пожара на которых или достижение их огнем приведет к техногенным и

экологическим катастрофам. Поскольку процесс развития пожара достаточно стремителен, то прогнозирование его пути и времени является актуальной научной задачей, поскольку ее решение позволяет выполнить оптимальное размещение сил и средств пожаротушения и предотвратить катастрофу. Традиционно такая задача решается начальником пожарного расчета, исходя из личного опыта и интуиции, возможно с использованием консультативной помощи подчиненных. Такой субъективизм зачастую приводит к неправильным, не подкрепленным расчетами и информационно-аналитическим сопровождением решениям и, следовательно, к человеческим жертвам и другим негативным последствиям.

Формально постановка задачи определения наиболее возможного пути и времени развития пожара является такой. Предположим, что предприятие, имеющее опасное производство, находится в некоторой системе координат XY . Пусть (x_0, y_0) – точка возникновения пожара, (x, y) – координаты особо опасного объекта. Тогда возникает несколько задач, в частности необходимо определить:

- минимально возможное время и соответствующий путь достижения пожаром точки (x, y) ;
- наиболее возможный путь и соответствующее время достижения пожаром точки (x, y) ;
- среднее возможное время и соответствующие пути достижения пожаром точки (x, y) .

Исходными данными являются заключения m экспертов, в соответствии с пожеланиями которых выполнена дискретизация координатной сетки. Предположим, что путь огня может пройти по одному из p маршрутов: $\{S_1, S_2, \dots, S_p\}$. Каждый из маршрутов может состоять из разного количества участков. Так,

$$S_1 = \{s_1^1, s_1^2, \dots, s_1^{k_1}\}, S_2 = \{s_2^1, s_2^2, \dots, s_2^{k_2}\}, \dots, S_p = \{s_p^1, s_p^2, \dots, s_p^{k_p}\}, \quad (1)$$

где s_i^j – j -й участок i -го маршрута, k_i – количество участков i -го маршрута, $i = \overline{1, p}$. Очевидно, что время распространения пожара от точки (x_0, y_0) к точке (x, y) является суммой времен прохождения огнем указанных участков. Заметим, что участки в большинстве случаев являются отрезками прямой, но понятие участка маршрута включает в себя также технологические проемы, двери, окна и подобные конструкции.

Учитывая нормативную информацию, справочные материалы, опыт и интуицию, эксперт указывает время распространения огня по определенному участку в виде функции принадлежности. В большинстве случаев, если эксперт имеет наибольшую уверенность в некотором одном значении аргумента, то он определяет время через треугольную функцию принадлежности (ФП), имеющую два параметра [Згуровский, 1990]; если информационная энтропия является минимальной, ФП может иметь колоколообразную форму [Круглов, 2001]. Если у экспертов нет уверенности в точечном значении времени распространения огня и отсутствуют данные о подобных пожарах, то для решения нашей задачи рационально использовать трапециеподобные ФП, поскольку такие ФП позволяют использовать интервальное представление. Трапециеподобная ФП [Дюбуа, 1990] определяется пятью параметрами $(\underline{m}, \bar{m}, \alpha, \beta, h)$, где \underline{m} – нижнее модальное значение, \bar{m} – верхнее модальное значение, α – левый коэффициент скошенности, β – правый коэффициент скошенности, h – высота. Поскольку в эволюционном моделировании используются нейронные сети, то для использования для их обучения градиентных методов необходимо использовать гладкие ФП. Такими функциями являются колоколообразные (гауссовские) ФП с двумя параметрами.

Продукционные представления времени развития пожара

Для представления экспертных суждений используем продукционные правила [Рассел, 2005], [Джарратано, 2007]. Рассмотрим два способа их записи. В первом случае эксперт делает заключение о наиболее возможном с его точки зрения пути распространения огня. Другой способ более емкий и

базируется на суждениях экспертов о возможных путях развития пожара. Заметим, что в первом варианте пути распространения огня, указанные всеми экспертами должны совпадать, в противном случае задача сводится к другому варианту и эксперту предлагается оценивать варианты, предложенные другими экспертами. Таким образом, для наиболее возможного пути развития пожара имеем логическую схему:

$$\begin{aligned}
 & \text{If } t_1 \in A_1^1 \ \& \ t_2 \in A_2^1 \ \& \dots \ \& \ t_n \in A_n^1 \ \text{then } T \in B_1 \ \text{with weighth } w_1 \quad V \\
 & \text{If } t_1 \in A_1^2 \ \& \ t_2 \in A_2^2 \ \& \dots \ \& \ t_n \in A_n^2 \ \text{then } T \in B_2 \ \text{with weighth } w_2 \quad V \\
 & \dots \\
 & \text{If } t_1 \in A_1^m \ \& \ t_2 \in A_2^m \ \& \dots \ \& \ t_n \in A_n^m \ \text{then } T \in B_m \ \text{with weighth } w_m,
 \end{aligned} \tag{2}$$

где t_i – входные величины, указывающие на время прохождения огнем i -го участка пути, $i = \overline{1, n}$; n – количество участков пути; A_j^i – функция принадлежности, определенная j -м экспертом для времени t_i , $j = \overline{1, m}$; m – количество экспертов; T – время прохождения огнем от точки возникновения пожара (x_0, y_0) до точки (x, y) ; B_j – функция принадлежности, определенная j -м экспертом для времени T ; w_j – весовые коэффициенты, определяющие компетентность экспертов.

Иную форму представления имеет система продукций для варианта оценки всеми экспертами всех предложенных вариантов:

$$\begin{aligned}
 & \text{If } t_1^1 \in A_1^{11} \ \& \ t_2^1 \in A_2^{11} \ \& \dots \ \& \ t_{k_1}^1 \in A_{k_1}^{11} \ \text{then } T_1 \in B_1^1 \quad V \\
 & \text{if } t_1^2 \in A_1^{21} \ \& \ t_2^2 \in A_2^{21} \ \& \dots \ \& \ t_{k_2}^2 \in A_{k_2}^{21} \ \text{then } T_2 \in B_1^2 \quad V \\
 & \dots \\
 & \text{if } t_1^p \in A_1^{p1} \ \& \ t_2^p \in A_2^{p1} \ \& \dots \ \& \ t_{k_p}^p \in A_{k_p}^{p1} \ \text{then } T_p \in B_1^p \ \text{with weighth } w_1 \\
 & \text{or} \\
 & \text{If } t_1^1 \in A_1^{12} \ \& \ t_2^1 \in A_2^{12} \ \& \dots \ \& \ t_{k_1}^1 \in A_{k_1}^{12} \ \text{then } T_1 \in B_2^1 \quad V \\
 & \text{if } t_1^2 \in A_1^{22} \ \& \ t_2^2 \in A_2^{22} \ \& \dots \ \& \ t_{k_2}^2 \in A_{k_2}^{22} \ \text{then } T_2 \in B_2^2 \quad V \\
 & \dots \\
 & \text{if } t_1^p \in A_1^{p2} \ \& \ t_2^p \in A_2^{p2} \ \& \dots \ \& \ t_{k_p}^p \in A_{k_p}^{p2} \ \text{then } T_p \in B_2^p \ \text{with weighth } w_2 \\
 & \text{or} \\
 & \dots \\
 & \text{or} \\
 & \text{If } t_1^1 \in A_1^{1m} \ \& \ t_2^1 \in A_2^{1m} \ \& \dots \ \& \ t_{k_1}^1 \in A_{k_1}^{1m} \ \text{then } T_1 \in B_m^1 \quad V \\
 & \text{if } t_1^2 \in A_1^{2m} \ \& \ t_2^2 \in A_2^{2m} \ \& \dots \ \& \ t_{k_2}^2 \in A_{k_2}^{2m} \ \text{then } T_2 \in B_m^2 \quad V \\
 & \dots \\
 & \text{if } t_1^p \in A_1^{pm} \ \& \ t_2^p \in A_2^{pm} \ \& \dots \ \& \ t_{k_p}^p \in A_{k_p}^{pm} \ \text{then } T_p \in B_m^p \ \text{with weighth } w_m.
 \end{aligned} \tag{3}$$

В выражении (3) t_i^j – время прохождения огнем i -го участка в j -варианте развития пожара, $i = \overline{1, k_q}$; k_q – количество участков распространение огня в q -варианте развития пожара, $q = \overline{1, p}$; A_i^{jk} – функция принадлежности определенная k -м экспертом для времени t_i^j , $k = \overline{1, m}$; T_q – время прохождения огнем от точки (x_0, y_0) до точки (x, y) при q -м варианте развития пожара; B_k^q – функция принадлежности определенная k -м экспертом для q -го варианта развития пожара.

Определение компетентности экспертов на базе аксиомы несмещенности

Компетентность экспертов, выраженная в весовых коэффициентах их суждений, для ЛПР на момент проведения экспертизы и моделирования может быть неизвестна. Возможно, что возникнет необходимость проверки компетентности и подтверждения или опровержения оценки ЛПР. Предположим, что априорная информация об уровне компетентности экспертов отсутствует. Тогда рационально воспользоваться процедурой ее определения, разработанной в статье [Снитюк, 2000]. Такая процедура базируется на аксиоме несмещенности [Матвеевский, 1987], сущность которой заключается в том, что суждение большинства компетентно. Следствием из нее есть утверждение о том, что наиболее компетентным является тот эксперт, суждения которого максимально совпадают с суждениями других экспертов, т.е.

$$\gamma_{\max} = \max_k \gamma_k = \max_k \frac{\sum_{i=1}^n \sum_{j=1}^m w_i \mu(\Omega_{ij}, \Omega_{ik})}{\sum_{i=1}^n \sum_{j=1}^m \sum_{p>j}^m w_i \mu(\Omega_{ij}, \Omega_{ip})}, \quad (4)$$

где γ_k – компетентности экспертов, $k = \overline{1, m}$; w_i – весовые коэффициенты, указывающие на важность вопроса, $i = \overline{1, n}$; $\mu(\Omega_{ij}, \Omega_{ik})$ – мера близости ответов j -го и k -го экспертов на i -й вопрос.

Для определения меры близости суждений экспертов предложен алгоритм [Снитюк, 2000], базирующийся на их анкетировании, классификации вопросов анкеты в зависимости от типа ответов, соответствующих моделях и их композиции. Полученные значения компетентности экспертов используются в качестве исходных данных моделирования.

Эволюционная идентификация времени распространения пожара

Заключения экспертов в форме (2)-(3) имеют неточный характер, выраженный значениями соответствующих параметров функций принадлежности. Их настройка и оптимизация являются необходимым условием получения математических моделей и дальнейшего использования в практических задачах. Рассмотрим методы получения параметров (2).

Нейросетевое моделирование. Без ограничения общности в качестве модели выберем нейросеть TSK (Takagi, Sugeno, Kang'a) [Takagi, 1985]. Реализованная в ней система правил аналогична (2) и имеет вид:

$$\Pi_k : \text{если } (x_1 \in A_1^k \ \& \ x_2 \in A_2^k \ \& \ \dots \ \& \ x_n \in A_n^k), \text{ то } y = p_{k0} + \sum_{j=1}^n p_{kj} x_j, \text{ где } \mathcal{Y} \quad (5)$$

Нейронная сеть, реализующая соответствующий вывод, изображена на рис. 1.

В первом шаре нейронов выполняется фазсификация входных значений, т. е. для каждого значения каждой переменной и каждого значения терм-множества находят значения ФП $\mu_{A_j^k}(x_j)$, $k = \overline{1, m}$, $j = \overline{1, n}$. Количество нейронов второго шара соответствует количеству входов сети. В них выполняется агрегация ФП отдельных переменных и для каждого правила рассчитывается

$$w^k = \min_{i=1, n} \left\{ \mu_{A_i^k}(x_i) \right\}, \quad k = \overline{1, m}.$$

В нейронах третьего шара рассчитываются, собственно, значения функций TSK, которые умножаются на выходы нейронов предыдущего шара, т.е.

$$g_i = w^i \cdot (p_{i0} + \sum_{j=1}^n p_{ij} x_j).$$

Очевидно, что количество нейронов совпадает с количеством нейронов предыдущего шара. Четвертый шар образуют два нейрона. В первом из них рассчитывается сумма $f_1 = \sum_{i=1}^n g_i$, во втором - $f_2 = \sum_{i=1}^n w^i$. Результирующий единственный нейрон пятого шара выполняет деление $Y = \frac{f_1}{f_2}$.

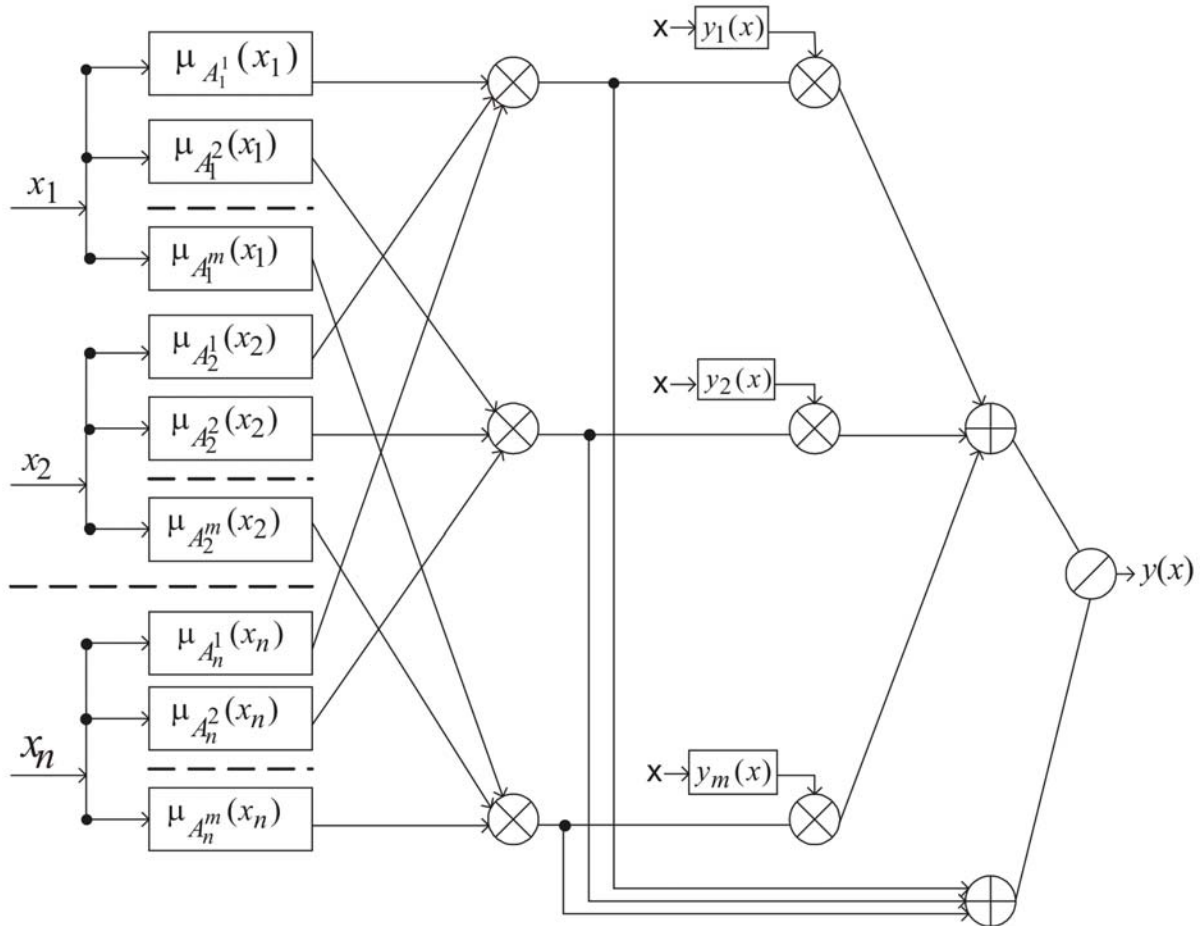


Рис. 1. Структура нечеткой нейронной сети TSK

Общее выражение функционирования сети TSK есть таким:

$$y(x) = \frac{1}{\sum_{k=1}^m \prod_{j=1}^n \mu_{A_j^k}(x_j)} \cdot \sum_{k=1}^m (p_{k0} + \sum_{j=1}^n p_{kj} x_j) \cdot \prod_{j=1}^n \mu_{A_j^k}(x). \quad (6)$$

Предположим, что нечеткой нейросетью реализуется неизвестное отображение

$$T = F(X). \quad (7)$$

Существует учебная выборка $\{(x^1, t^1), \dots, (x^n, t^n)\}$. Заметим, что векторы $x^i, i = \overline{1, n}$ являются действительными. Целевая функция для k -го образа будет такой:

$$E_k = \frac{1}{2} (\hat{Z}^k(\gamma) - Z^k)^2, k = \overline{1, n}, \quad (8)$$

де $Z^k(\gamma)$ – рассчитанные значения выхода нейросети, Z^k – значения, заданные таблично. Обучение сети осуществляется по градиентному методу. Фаззификация входов нейросети осуществляется ФП

$$\mu_{A_j^i}(x_i) = \frac{1}{1 + e^{b_j^i(x_i - a_j^i)}}, \quad i = \overline{1, n}, \quad j = \overline{1, m},$$

где m – количество правил. Тогда вектор параметров нейросети, которые подлежат оптимизации, будет таким:

$$\gamma = (a_1^1, b_1^1, a_1^2, b_1^2, \dots, a_1^n, b_1^n, a_2^1, b_2^1, \dots, a_2^n, b_2^n, \dots, a_m^1, b_m^1, p_{10}, p_{11}, p_{12}, \dots, p_{1n}, \dots, p_{m0}, p_{m1}, p_{m2}, \dots, p_{mn}).$$

Приведем основные выражения для обучения модели TSK:

$$a_i^j(t+1) = a_i^j(t) - \eta \frac{\partial E_k}{\partial a_i^j(t)}, \quad k = \overline{1, p}, \quad p - \text{количество обучающих образов};$$

$$\begin{aligned} \frac{\partial E_k}{a_i^j} &= \frac{\partial E_k}{\partial \hat{Z}^k(\gamma)} \cdot \frac{\partial \hat{Z}^k(\gamma)}{\partial \mu_{A_j^i}(x_i)} \cdot \frac{\partial \mu_{A_j^i}(x_i)}{\partial a_i^j} = (\hat{Z}_k(\gamma) - Z^k) \cdot \left[\frac{1}{\sum_{j=1}^m \prod_{i=1}^n \mu_{A_j^i}(x_i)} \times \right. \\ &\quad \left. \times (p_{j_0} + \sum_{i=1}^m p_{ji} x_i) \prod_{\substack{i=1 \\ i \neq j}}^n \mu_{A_j^i}(x_i) - \frac{\prod_{i=1}^n \mu_{A_j^i}(x_i)}{(\sum_{j=1}^m \prod_{i=1}^n \mu_{A_j^i}(x_i))^2} \cdot \sum_{j=1}^m (p_{j_0} + \sum_{i=1}^m p_{ji} x_i) \cdot \prod_{i=1}^n \mu_{A_j^i}(x_i) \right] \cdot \frac{b_j^i \cdot e^{b_j^i(x_i - a_j^i)}}{(1 + e^{b_j^i(x_i - a_j^i)})^2}. \\ b_i^j(t+1) &= b_i^j(t) - \eta \frac{\partial E_k}{\partial b_i^j(t)}, \quad k = \overline{1, p}. \end{aligned}$$

В этом случае выражение для вычисления производной отличается от предыдущего только последним множителем:

$$\frac{\partial \mu_{A_j^i}(x_i)}{\partial b_j^i} = \frac{(a_j^i - x_i) \cdot e^{b_j^i(x_i - a_j^i)}}{(1 + e^{b_j^i(x_i - a_j^i)})^2}.$$

Для совокупности параметров $p_{ij}, j = \overline{1, m}, i = \overline{0, n}$ коррекция осуществляется так:

$$p_{ij}(t+1) = p_{ij}(t) - \eta \frac{\partial E_k}{\partial p_{ij}},$$

$$\frac{\partial E_k}{\partial p_{ji}} = \frac{\partial E_k}{\partial \hat{Z}^k(\gamma)} \cdot \frac{\partial \hat{Z}^k(\gamma)}{\partial p_{ji}} = (\hat{Z}^k(\gamma) - Z^k) \cdot \frac{x_i \prod_{i=1}^n \mu_{A_j^i}(x_i)}{\sum_{j=1}^m \prod_{i=1}^n \mu_{A_j^i}(x_i)},$$

$$\frac{\partial E_k}{\partial p_{j_0}} = (\hat{Z}^k(\gamma) - Z^k) \cdot \frac{\prod_{i=1}^n \mu_{A_j^i}(x_i)}{\sum_{j=1}^m \prod_{i=1}^n \mu_{A_j^i}(x_i)}.$$

Очевидно, что количество параметров в зависимости (6) является очень большим для адекватного и быстрого обучения нейросети. Существуют методы ускорения процесса обучения, базирующиеся на разделении совокупности параметров на две группы. Параметры первой группы вычисляются в результате решения системы уравнений или считаются постоянными величинами, параметры второй группы получают в процессе обучения нечеткой сети.

Эволюционное моделирование. Известно, что эволюционное моделирование используется преимущественно для оптимизации дискретнозначных функций. Эволюционные методы имеют различия, но общим является наличие целевой функции или функции приспособленности. В нашей задаче такой функцией является сумма функций (8). Потенциальные решения представляют собой элементы вектора γ . Существуют два подхода к представлению потенциальных решений. В первом случае такие решения представляют как генотипы, т.е. соответствующие бинарные хромосомы, поскольку известно, что такое представление обладает максимальной информационной насыщенностью. Второй подход базируется на фенотипическом представлении, при котором решения имеют десятичный вид. Для него характерно получение новых решений с использованием нормально распределенных смещений и без рекомбинаций.

У нейросетевых моделей и эволюционных методов есть как преимущества, так и недостатки. В пользу эволюционного моделирования свидетельствует отсутствие требований к целевым функциям и предусмотренная внутри алгоритмов операция мутации, позволяющая минимизировать риск получения локальных оптимумов. Преимуществом нейросетевых технологий является определяемое алгоритмами монотонное стремление целевой функции к удовлетворительному значению. Эффективность использования той или другой технологии зависит от количества участков прохождения пожара, количества экспертов и процедуры использования обучающей и контрольной последовательностей.

Особенности критериальной оценки полученных результатов

На предыдущем шаге получены оценки параметров, использование и учет которых направлены на объективизацию субъективных суждений. Вместе с тем, проблема принятия оптимального или допустимого решения остается. Решение задачи идентификации времени пожара в случае учета производственных правил (2) и дальнейшего принятия решений не представляет трудности, поскольку путь распространения идентифицирован, а время является некоторым средневзвешенным экспертным показателем.

Иная ситуация имеет место, если эксперты указывают множество путей распространения пожара от исходной до конечной точки. Неопределенность, вызванная неединственностью особо опасных объектов на возможном пути распространения огня, и различные масштабы возможных катастроф определяют критерии принятия решений. Очевидно, что в роли таких критериев могут быть:

- наиболее возможный путь распространения огня и разработка соответствующей процедуры определения времени распространения огня по этому пути;
- путь, время распространения огня по которому является минимальным, и разработка процедуры определения такого пути из множества возможных;
- путь, время распространения по которому является средним, и разработка соответствующей процедуры его определения.

Такие критерии являются аналогами известных критериев Сэвиджа и Гурвица [Таха, 2005]. Предусловием выбора решения, исходя из перечисленных критериев, является построение и использование экспертной системы, базирующейся на нечеткой базе знаний с правилами типа (2) или (3) и оптимизированными параметрами.

Нерешенные задачи и перспективы исследований

Осуществляя моделирование и определяя путь и время распространения пожара от точки его возникновения до особо опасного объекта необходимо знать о последствиях, авариях, катастрофах, к которым может привести достижение его огнем. Таким образом, возникает задача идентификации

функции убытков от последствий пожара. Знание ее значений является необходимым условием принятия решений в критических условиях.

Применение эволюционного моделирования обосновывается значительной субъективностью исходной информации, отсутствием единственной математической модели и соответствующих математических методов. В то же время разработка экспертных систем для решения подобных задач на базе реализации геоинформационных технологий сталкивается с массой проблем, разрешение которых в условиях ресурсного дефицита невозможно. Однако возрастающие риски, связанные с развитием опасных производств, раньше или позже приведут к созданию и использованию подобных технологий.

Библиография

- [Быченко, 2006] А.А. Быченко. Модели распространения пожара на особо опасных объектах в условиях неопределенности // Искусственный интеллект. – 2006. – № 3. – С. 359-365
- [Снитюк, 2006] В.Е. Снитюк, А.О. Биченко. Аспекты нечеткости при моделировании процессов распространения пожара на особо опасных объектах // Автоматизированные системы управления и приборы автоматики. – 2006. – Вып. 134. – С.89-93
- [Згуровский, 1990] М.З. Згуровский. Интегрированные системы оптимального управления и проектирования. – Киев: Выща школа, 1990. – 351 с.
- [Круглов, 2001] В.В. Круглов. Нечеткая логика и искусственные нейронные сети. – Москва: Физматлит, 2001. – 224 с.
- [Дюбуа, 1990] Д. Дюбуа, А. Прад. Теория возможностей. – Москва: Радио и связь, 1990. – 286 с.
- [Хайкин, 2006] С. Хайкин. Нейронные сети: полный курс. – М.: “Вильямс”, 2006. – 1104 с.
- [Рассел, 2005] С. Рассел, П. Норвиг. Искусственный интеллект: Современный подход. – М.: Вильямс, 2005. – 1424 с.
- [Джарратано, 2007] Д. Джарратано, Г. Райли. Экспертные системы: Принципы разработки и программирования. – М.: “И.Д. Вильямс”, 2007. – 1152 с. [Снитюк, 2000] В.Е. Снитюк, Рифат Мохаммед Али. Модели и методы определения компетентности экспертов на базе аксиомы несмещенности // Вестник Черкасского государственного технологического университета. – 2000. – № 4. – С. 121-126.
- [Матвеевский, 1987] С.Ф. Матвеевский Основы системного проектирования комплексов летательных аппаратов. – Москва: Машиностроение, 1987. – 239 с.
- [Takagi, 1985] T. Takagi, M. Sugeno. Fuzzy identification of systems and ist application to modeling and control // IEEE Trans. Systems, Man, and Cybernetics. – 1985. – Vol. 15. – Pp. 116-132.
- [Таха, 2005] Таха Хемди А. Введение в исследование операций. – Москва: “Вильямс”, 2005. – 912 с.

Информация об авторах

Виталий Снитюк – Киевский национальный университет имени Тараса Шевченко; пр. Акад. Глушкова 2, стр. 6, Киев, Украина; e-mail: svit@artint.com.ua

Артем Быченко – Черкасский институт пожарной безопасности имени Героев Чернобыля; ул. Оноприенко, 8, Черкассы, Украина; e-mail: bichenko@ukr.net

MULTIDECISION-2: A MULTICRITERIA DECISION SUPPORT SYSTEM

Vassil Vassilev, Mariana Vassileva, Boris Staykov, Krassimira Genova,
Filip Andonov, Plamena Chongova

Abstract: *The paper presents a multicriteria decision support system, called MultiDecision-2, which consists of two independent parts - MKA-2 subsystem and MKO-2 subsystem. MultiDecision-2 software system supports the decision makers (DMs) in the solving process of different problems of multicriteria analysis and linear (continues and integer) problems of multicriteria optimization. The two subsystems MKA-2 and MKO-2 of of MultiDecision-2 are briefly described in the paper in the terms of the class of the problems being solved, the system structure, the operation with the interface modules for input data entry and the information about DM's local preferences, as well as the operation with the interface modules for visualization of the current and final solutions.*

Keywords: *multicriteria decision support systems, multicriteria analysis, multicriteria optimization.*

ACM Classification Keywords: *H.4.2 Information Systems Applications: Types of Systems: Decision Support.*

Introduction

The Multicriteria Decision Support System *MultiDecision-2* system is a successor of system *MultiDecision-1* (Vassilev et al. (2005a)) and the system *MultiDecision-2.1* (Vassilev et al. (2005b)) and it is designed to support DMs in solving different multicriteria analysis and multicriteria optimization problems. The multicriteria analysis problems can be divided into three types: problems of multicriteria choice, problems of multicriteria ranking and problems of multicriteria sorting. Many real life problems in management practice may be formulated as problems of choice, ranking and sorting of resources, strategies, projects, offers, policies, credits, products, innovations, designs, costs, profits, portfolios, etc. The multicriteria optimization problems are only problems of multicriteria choice. Many real life problems in planning, control and industrial production may be formulated as problems of multicriteria choice or linear (continues or integer problems) of multicriteria optimization.

In multicriteria analysis and multicriteria optimization problems several criteria are simultaneously optimized in the feasible set of alternatives. In the general case there does not exist one alternative, which optimizes all the criteria. There is a set of alternatives however, characterized by the following: each improvement in the value of one criterion leads to deterioration in the value of at least one other criterion. This set of alternatives is called a set of the non-dominating or Pareto optimal alternatives (solutions). Each alternative in this set could be a solution of the multicriteria problem. In order to select one alternative, it is necessary to have additional information set by the so-called decision maker (DM). The information that the DM provides reflects his/her global preferences with respect to the quality of the alternative sought.

The systems developed to support the solution of multicriteria analysis or multicriteria optimization problems can be classified in three groups: commercial, research or teaching and experimental (for testing of new methods). The software systems supporting the solution of multicriteria analysis or multicriteria optimization problems can be divided also in two classes – software systems with general purpose and problem-oriented software systems. The general-purpose software systems aid the solution of different multicriteria analysis or multicriteria optimization problems by different decision makers. One method or several methods from one and the same group are usually realized in this kind of systems for solving multicriteria analysis or multicriteria optimization problems. The problem-oriented software systems, which support the solving of multicriteria analysis or multicriteria optimization problems, are included in other information-control systems and serve to aid the solution of one or several types of specific multicriteria analysis or multicriteria optimization problems. In this connection problem-oriented user's interface is usually realized in this kind of systems and methods from different groups of multicriteria analysis or multicriteria optimization methods are included in some of these systems.

The following general-purpose software systems are developed to aid the solution of different multicriteria problems: VIMDA, Expert Choice, PROMCALC and GAIA, ELECTRE III-IV, MACBETH, VIP, Decision Lab, Web-HIPRE, MultiChoice and KnowCube (Weistroffer et al. (2005)). Four interesting problem-oriented software systems for supporting the solving of particular multicriteria analysis problems are the following systems: FINCLAS System - for financial classification problems, Agland Decision System – for agricultural property problems, the DESYRE System – for rehabilitation of contaminated sites, the MultCSync System – for incorporating multiple criteria in conservation planning.

Some well-known general-purpose software systems, which support the solving of multicriteria optimization problems, are the following systems: VIG, DIDAS, DINAS, MOLP-16, LBS, SOMMIX, MOIP, WWW-NIMBUS, MOLIP, NLPJOB and MOMILP (Weistroffer et al. (2005)). The Multicriteria Decision Support System for river water-quality planning and the ADELAIS System for portfolio selection are two attractive problem-oriented multicriteria optimization systems. In the class of multicriteria optimization software systems must also be included software systems, which implement different multicriteria evolutionary methods (algorithms). Four of them are the following ones: NSGM System (Srinivas and Deb (1994)), MOSES System (Coello and Christiansen (1999)), M-PAES System (Knowles and Corne (2000)) and the MOEA toolbox for MATLAB.

The paper presents some basic elements of the software system, called *MultiDecision-2*, which consist of two separate parts - the general-purpose software subsystem MKA-2, which is designed to support DMs in solving different multicriteria analysis problems and the general-purpose software subsystem MKO-2, which is designed to aid the solving of different multicriteria optimization problems. The subsystems MKA-2 and MKO-2 are described in the next two sections. Conclusions are given in the last section.

MKA-2 Subsystem

The MKA-2 subsystem, which is the first part of the *MultiDecision-2* system, is a successor of the software system MKA-1 (Genova et al. (2004)), developed in the Institute of Information Technologies – Bulgarian Academy of Sciences. The MKA-2 system operates under MS Windows operating system and it is designed to support DMs in solving different multicriteria analysis problems.

The multicriteria analysis problem may be described by a decision matrix A ($n \times k$), which can be defined as follows:

k_j	$k_1(.)$	$k_2(.)$...	$k_j(.)$...	$k_k(.)$
a_i						
a_1	a_{11}	a_{12}	...	a_{1j}	...	a_{1k}
a_2	a_{21}	a_{22}	...	a_{2j}	...	a_{2k}
...
a_i	a_{i1}	a_{i2}	...	a_{ij}	...	a_{ik}
..
a_n	a_{n1}	a_{n2}	...	a_{nj}	...	a_{nk}

Table 1. Decision Matrix

where a_i denotes an alternative with an index $i, i=1, \dots, n$; and $k_j(.)$ denotes a criterion with an index $j, j=1, \dots, k$.

The evaluation of the i -th alternative with respect to all the criteria is given by the row vector $(a_{i1}, a_{i2}, \dots, a_{ik})$. The evaluation of all the alternatives with respect to j -th criterion is given by the column vector $(a_{1j}, a_{2j}, \dots, a_{nj})^T$.

Different methods have been developed to solve multicriteria analysis problems. A great number of the methods proposed up to now, can be grouped in three separate classes (Vincke (1992)). The first class of methods (Dyer (2004)) includes the multiattribute utility (value) theory methods (such as Value Tradeoff Method, UTA Method,

MACBETH Method, Direct Weighting Method, AHP Weighting Methods). There are differences in the way in which the DM's global preferences are aggregated in the two subclasses of these methods. In the first one a generalized functional criterion is directly synthesized, whereas in the second subclass (weighting methods) it could be said that such a criterion (additive form) is indirectly synthesized. The two subclasses of methods are based on the assumption that there does not exist limited comparability among the alternatives. The second class of methods is called outranking methods (such as ELECTRE methods (Figueira et al. 2005)), PROMETHEE methods (Brans and Mareschal (2005)), etc.). They are based on the assumption that there exists limited comparability among the alternatives. In these methods one (or several outranking relation(s)) are first built to aggregate DM's global preferences, after which this outranking relation is used to assist the DM in solving the multiple criteria decision analysis problem. In most of the outranking methods it is assumed that the DM selects to specify some preference information about inter- and intra-criteria. While the inter-criteria information is expressed in the form of weights and veto thresholds, the intra-criteria information is usually expressed in the form of indifference and preference thresholds. The interactive algorithms (such as RNIM method (Narula et al. (2003), etc.) belong to the methods of the third group. They are "optimizationally motivated" and are oriented to solve multicriteria analysis problems with a large number of alternatives and a small number of criteria.

The MKA-2 system consists of internal-system modules, four solving modules and interface modules. It is realized in MS Windows environment, including the standard for this operating system user interface elements. The internal-system modules contain all global definitions of variables, functions and procedures of general purpose. The object possibilities of Visual Basic are utilized in MKA-2 system, creating several classes with respect to internal-system structures. They are the following: a class for messages, which encapsulates the output of error messages, dynamic context help information and logging events in the debug window, localization and identification of errors occurring during the system operation; a class matrix with some specific procedures, necessary for AHP method; a class for storing the information specific for the criteria in ELECTRE III and PROMETHEE II methods and a class for storing elements of the CBIM interactive method history. MKA-2 handles files with "*.mka" extension. Standard operations for creating, editing, loading and saving of files are implemented. The MKA-2 files contain input data and data related to the process and the results from solving multicriteria analysis problems.

The solving modules realize four methods - AHP Method, ELECTRE III Method, PROMETHEE II Method and CBIM Method and procedures for transformation of qualitative, ranking and weighting criteria into quantitative criteria. AHP Method is one of the most widely spread weighting methods. Pair-wise criteria comparison is used in this method to set DM's preferences. On this basis a pair-wise comparison matrix is constructed. The estimates of the weights can be found by normalizing the eigenvector corresponding to the largest eigenvalue of this matrix. ELECTRE III Method is one of the most often used outranking methods. It is based on an outranking relation, characterized by the definition of an outranking degree $S(a, b)$ associated with each ordered pair (a, b) of alternatives, representing the more or less great outranking credibility of a over b . There are two matrix needed to be evaluated: the concordance and the discordance matrix. The concordance matrix gives an assessment of agreement that one alternative is better than other one. It requires two type thresholds - indifference and preference thresholds. The discordance matrix gives an assessment of disagreement that one alternative is better than other one. That matrix requires additional threshold, called veto threshold, which allows the outranking relation to be rejected. In order to be obtained the degree of credibility of outranking, there follows the combining the two measures from concordance and discordance matrix. This degree is thus equal to the concordance index where no criterion is discordant or where no veto threshold is used, in the opposite case. The concordance index is lowered in function of the importance of the discordance. The obtained credibility matrix is essential for generating two distillation orders that show whether one alternative outranks the other or such an alternative is incomparable to the other. In order to be obtained final ranking the two orders are combined. PROMETHEE II Method is the second of the most often used outranking methods. In this method the intensity of the preference of one alternative over another alternative regarding each criterion is measured in terms of the so-called preference function. On the basis of two type thresholds - indifference and preference thresholds - six types of preference

functions are used in the method. The method provides a complete ranking of the alternatives through a pair-wise dominance comparison of net positive and net negative outranking flows. RNIM method (Narula et al. (2003)) is a representative of the interactive methods and is appropriate for solving multicriteria analysis problems with a large number of alternatives and a small number of criteria. The DM can provide desired or acceptable levels, directions and intervals of changes in the values of the criteria at any iteration. On the basis of this information, the method proposed enables the use of discrete optimization scalarizing problems, with the help of which the DM has the possibility for a more systematic and successful screening of the alternatives set.

The interface modules ensure the interaction between MKA-2 system, the DM and the operating system. This interaction includes the entry of the data for the multicriteria analysis problems; the entry of specific information for every method; the entry of information about DM's preferences; the visualization of the current results and the final result; the graphical presentation of the solutions; the printing out, reading and storing of files; the multi-language support, etc. The editing module enables entering, alteration and storing of quantitative, qualitative, ranking and weighting criteria. The interface preference modules aid the DM in the entry of criteria pair-wise comparison information, inter- and intra-criteria information and information about the desired or acceptable levels, directions and intervals of change in the values of the criteria. The current and final results and the parameters for the separate methods selected by the DM are presented digitally and graphically with the help of visual interface modules. The input/output interface modules enable the reading and storing in files, the printing of the current and final results obtained, as well as the printing of the information, given by the DM. The solution process of a multicriteria problem can be interrupted at any stage and activated from the place of its interruption at any time. MKA-2 system has comparatively rich printing functions – every piece of the data (entered or computed) may be printed. In this way, the entire process of decision making is documented – you can review the input data of the multicriteria problem, the DM's preferences entered, the current values obtained, and the final result also, which on its turn can be printed out in the form of values or graphics. The rest of the interface modules realize a dynamic help, multi-language maintenance, etc.

Fig. 1 shows a window with information about the pair-wise comparison of the criteria for one real multicriteria analysis problem, concerning the selection of an appropriate marketing action for advertising of bicycle manufacturing company products (Brans and Mareschal (2000)). This is information about DM's preferences in operation with AHP method. Fig. 2 presents a window with information about DM's preferences in operation with PROMETHEE II method.

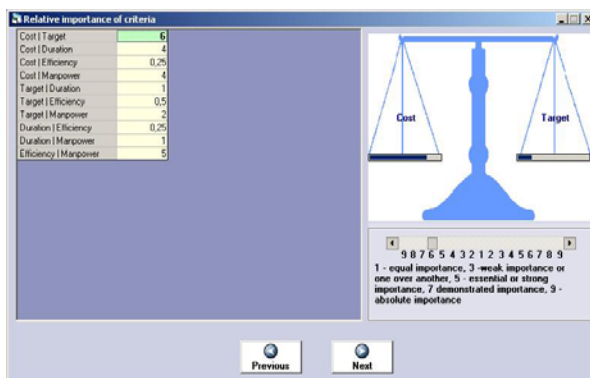


Fig. 1.

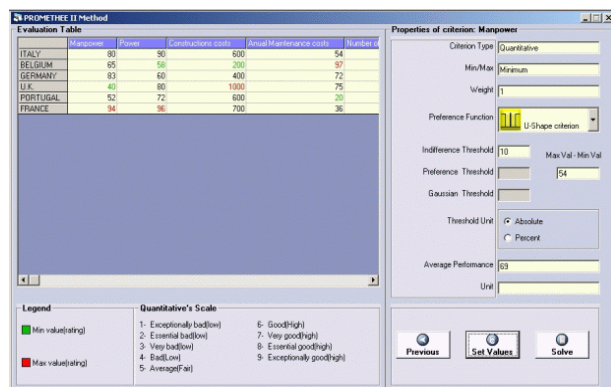


Fig. 2.

MKO-2 subsystem

The MKO-2 subsystem, the second part of the system MultiDecision-2, is a successor of the research software system MKO-1 (Vassilev et al. (2004)), developed in the Institute of Information Technologies – Bulgarian Academy of Sciences. The first version of the system MKO-2 software system is designed to aid the solution of

linear and linear integer problems for multicriteria optimization only and it is oriented towards operation under the control of MS Windows operating system.

The linear and linear integer multicriteria optimization problem may be described as follows:

To optimize simultaneously the criteria:

$$\{f_k(x), k \in K\}$$

subject to:

$$\sum_{j \in N} a_{ij} x_j \leq b_i, i \in M,$$

$$0 \leq x_j \leq d_j, j \in N,$$

$$x_j - \text{integers}, j \in N'; N' \subset N,$$

where:

$$f_k(x), k \in K \text{ are linear criteria of the type: } f_k(x) = \sum_{j \in N} c_j^k x_j;$$

$$x = (x_1, \dots, x_j, \dots, x_n)^T \text{ is the variables vector;}$$

$$f(x) = (f_1(x), \dots, f_k(x), \dots, f_p(x))^T \text{ is the vector of the criteria;}$$

$K = \{1, 2, \dots, p\}$, $M = \{1, 2, \dots, m\}$, $N = \{1, 2, \dots, n\}$ and $N' = \{1, 2, \dots, n' / n' \leq n\}$ are sets of the indices of the linear criteria, the linear constraints, the variables and the integer variables, respectively.

There are two main approaches in solving multicriteria optimization problems: a scalarizing approach (Miettinen (2003), Korhonen (2005)) and an approximation approach (Ehrgott and Wiecek (2005)). The major representatives of the scalarizing approach are the interactive algorithms. Multicriteria optimization problems is treated in these algorithms as a decision making problem and the emphasis is put on the real participation of the DM in the process of its solution. The interactive methods are the most developed and widespread due to their basic advantages – a small part of the Pareto optimal solutions must be generated and evaluated by the DM; in the process of solving the multicriteria problem, the DM is able to learn with respect to the problem; the DM can change his/her preferences in the process of problem solution; the DM feels more confident in his/her preferences concerning the final solution.

The interactive methods of the reference point (direction) and the classification-oriented interactive methods (Miettinen (1999)) are the most widely spread interactive methods solving multicriteria optimization problems. Though the interactive methods of the reference point are still dominating, the classification-oriented interactive methods enable the better solution of some chief problems in the dialogue with the DM, relating to his/her preferences defining, and also concerning the time of waiting for new non-dominated solutions that are evaluated and selected. The generalized interactive algorithm GENWS-IM is an interactive algorithm (Vassileva (2006)) with variable scalarization and parametrization. It is a generalization of a large part of the multicriteria optimization interactive algorithms developed up to the present moment. This generalization is with respect to the classes of the problems solved, the type of the defined preferences, the number and type of the applied scalarizing problems, the strategies used in the search for new Pareto optimal solutions. Starting from the current (weak) Pareto optimal solution, the generalized scalarizing (Vassileva (2006)) problem GENWS may be used. Altering some parameters of the generalized scalarizing problem GENWS the following known scalarizing problems can be obtained: the scalarizing problem of the weighted sum WS; the scalarizing problem of ϵ -constraints EO; the scalarizing problem STEM; the scalarizing problem STOM; the scalarizing problem of the reference point RP; the scalarizing problem GUESS; the scalarizing problem MRP; the external reference direction scalarizing problem RD3; the classification-oriented scalarizing problem NIMBUS; the classification-oriented scalarizing problem DALDI. On the basis of the generalized scalarizing problem GENWS, a generalized interactive (Vassileva (2006)) method GENWS-IM with variable scalarizations and parameterization could be designed, having the following

characteristics: the DM may set his/her preferences with the help of the criteria weights, ε -constraints, desired and acceptable levels of change of the criteria values, desired and acceptable levels, directions and intervals of alteration in the criteria values, etc.; during the process of the multicriteria problems solving, the DM may change the way of presenting his/her preferences. Starting from one and the same current Pareto optimal solution and applying different scalarizing problems (with respective alteration of GENWS), the DM may obtain different new Pareto optimal solutions at a given iteration, and this opportunity is especially useful in education and in comparison of different scalarizing problems.

A variety of methods to approximate the set of Pareto optimal solutions of different types have been proposed (Ehrgott and Wiecek (2005)). A big majority of methods are iterative and produce points or objects approximating this set. Some methods are exact equipped with theoretical proofs for correctness and optimality while some other methods are heuristic and often theoretically unsupported. The main representatives of the heuristic methods are the multicriteria genetic (evolutionary) methods (Deb (2001)). The multicriteria optimization problem is treated in these methods rather as a vector optimization problem, than as a decision making problem and the stress is placed on the determination of a subset of potential Pareto optimal solutions, which approximates well enough the whole Pareto optimal set. The solutions obtained with the help of the genetic methods, are near Pareto optimal solutions. Besides this, during the process of defining the approximating set, the DM is isolated and he/she is provided with a large set of solutions for evaluation and choice towards the end (this is a comparatively hard problem of multicriteria analysis

MKO-2 software system consists of three main groups of modules – a control program, optimization modules and interactive modules. The control program is integrated software environment for creation, processing and storing of files associated with MKO-2 system, as well as for linking and execution of different types of software modules. The basic functional possibilities of the control program may be separated in three groups. The first group includes the possibilities to use the applications, menus and system functions being standard for MS Windows – “File”, “Edit”, “View”, “Window”, “Help”, in the environment of MKO-2 system. The second group of functional possibilities encloses the control of the interactions between the modules realizing: creation, modification and storing of files, associated with MKO-2 system, which contain input data and data connected with the process of solution of linear and linear integer multicriteria optimization problems entered;

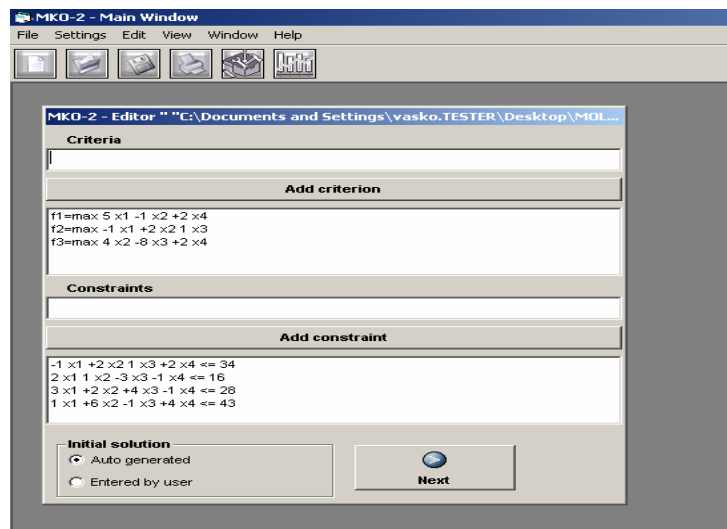


Fig.3.

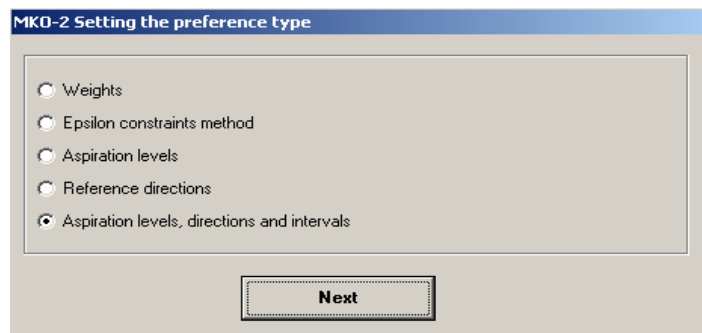


Fig.4.

localization and identification of the errors occurring during the process of operation with MKO-2 system. The third group of functional possibilities of the control program includes the possibilities for visualization of essential information about the DM and information of the system operation as a whole. The optimization modules realize the generalized interactive method GENWS-IM, two simplex algorithms solving continuous single-criterion problems (Vanderbei (1996)), an algorithm of "branches of bounds" type for exact solution of linear integer single-criterion problems (Wolsey (1998)) and an algorithm (Vassilev and Genova (1991)) for approximate solution of linear integer single-criterion problems.

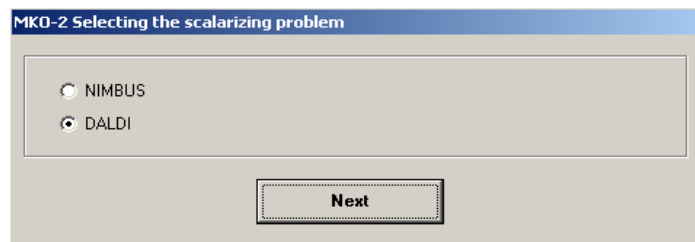


Fig.5.

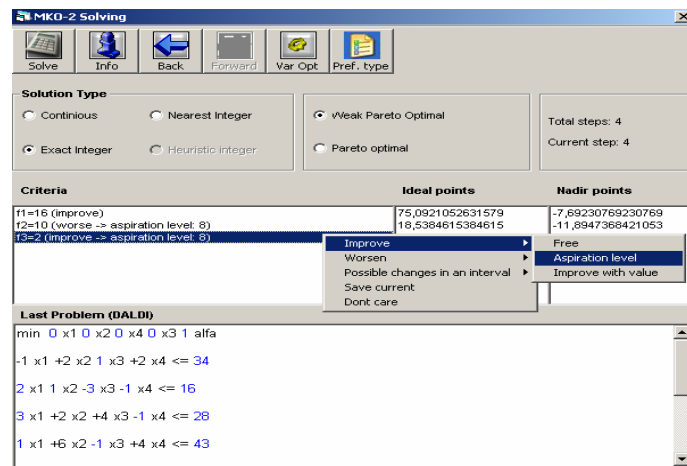


Fig.6.

The interface modules provide the dialogue between the DM and the system during the entry and correction of the input data of the multicriteria problems solved, during the interactive process of these problems solution and for dynamic numerical and graphical visualization of the main parameters of this process. With the help of an ending module the descriptions of the criteria and constraints are input, altered and stored, and also the type and limits of the variables alteration. Another interface module serves to supply two types of graphic presentation of the information about the values of the criteria at the different steps, as well as the possibilities for their comparison.

One of the main functions of MKO-2 system is to enable the extension of DM's possibilities to set his/her preferences with the help of criteria weights, ϵ -constraints, desired and acceptable directions of change of the criteria values, desired and acceptable levels, directions and intervals alteration of the criteria values. Twelve scalarizing problems are generated in MKO-2 system in order to realize these possibilities. Depending on DM's preferences, these scalarizing problems are automatically generated by the generalized scalarizing problem GENWS with the help of a change in their structure and their parameters.

MKO-2 system presents the DM different windows intended for entry and correction of the problem criteria and constraints, for setting his/her preferences. Fig.3, Fig.4, Fig.5 and Fig.6 show four of these windows. The window presented in Fig.3 is the basic window of the editor for input data entry – "MKO-2 Editor". The window presented in Fig.4, is designed to identify the type of DM's preferences. The DM may select among five types of preferences and let assume that he/she has selected to set the preferences by aspiration levels, directions and intervals. The window shown in Fig.5 is intended for selection of the scalarizing problem from the set of already known classification-oriented scalarizing problems. In order to enter the different types of DM's preferences, different windows are used. The window presented in Fig.6, is designed to enter DM's preferences with the help of desired or acceptable levels, directions and intervals of alteration in the criteria values, (operating with the classification-oriented scalarizing problem DALDI). and for solving of linear and linear integer multicriteria problems. The screen shows the setting of a new aspiration level for the value of the third criterion.

The solving of linear and linear integer multicriteria problems is realized with the help of 12 "MKO-2 Solving" windows, intended to work with the 12 interactive methods. Every windows "MKO-2 Solving" is divided into

several zones. Its upper part contains a band with buttons that realize the main functions of the process for interactive solution of multicriteria linear and linear integer problems. These are the following buttons: *Solve* - for starting the optimization module in order to find a new current solution of MKO-2, solving the scalarizing problem generated at this iteration; *Info* - for visualization of the variables values at the current solution in a separate window; *Back* and *Forward* - for navigation which allow the DM to go back to preceding steps and reconsider the solutions found. The next field of "MKO-2 Solving" window contains radio buttons for setup of the type of solution looked for: continuous, integer, approximate integer, the closest integer, as well as weak Pareto optimal or Pareto optimal. Below them information is found about the time of the system operation for the current problem in seconds, the number of the step being currently considered and the total number of the executed steps.

When solving multicriteria optimization problems, it is important to provide information not only about the last solution found, but also about the solutions found at previous iterations. It is important that the DM could "testify" how he/she has reached the last solution. Hence, the information about the interactive process of the multicriteria optimization problem solving, comprising not only the problem input data, the solutions obtained at each iteration, the preferences set by the DM for a new search and the scalarizing problems constructed, stored in "*.mlp" files associated with MKO-2 system serve not only to restart an interrupted solution process, but also for documentation.

Conclusion

MultiDecision-2 system is designed to support DMs in solving different multicriteria analysis and multicriteria optimization problems. MKA-2 system is designed to support the DMs in modeling and solving problems of multicriteria ranking and multicriteria choice. MKO-2 system is designed to model and solve linear and linear integer problems of MO. The user-friendly interface of MKA-2 and MKO-2 systems facilitates the operation of DMs with different qualification level relating to the multicriteria analysis and optimization methods and software tools. MKA-2 and MKO-2 systems can be used for education and for experimental and research problems solving as well. *MultiDecision-2* system is a local multicriteria decision support system and operates in two languages – Bulgarian and English. A number of Bulgarian universities use the system for the purposes of education and for experimental and research problems solving as well. A number of official organizations and companies use the system for solving real multicriteria decision making problems. The future development of the *MultiDecision-2* system will be realized in two directions. The first direction is connected with the addition of new methods. The second direction refers to web-based versions of the system, enabling distant decision making.

Bibliography

1. Brans, J., Mareschal, B. (2005). PROMETHEE Methods. In: Multiple Criteria Decision Analysis: State of the Art Surveys, (J. Figueira, S. Greco, and M. Ehrgott, Eds), Springer Verlag. London, 163-196.
2. Brans, J., Mareschal, B. (1990). The Promethee Methods for MCDM: the Promcale, Gaia and Bankadviser Software. Readings in Multiple Criteria Decision Aid (A. Carlos, C. Bana Costa, Eds.), Springer-Verlag, Berlin, 216-252
3. Coello, C., Christiansen, A. (1999). MOSES: A Multiobjective Optimization Tool for Engineering Design, Engineering Optimization, vol. 31(3), pp. 337-368.
4. Deb, K. (2001). Multi-Objective Optimization using Evolutionary Algorithms. Wiley-Interscience Series in Systems and Optimization. John Wiley & Sons, Chichester.
5. Dyer, J. (2004). MAUT-Multiattribute Utility Theory. (2005). . In: Multiple Criteria Decision Analysis: State of the Art Surveys, (J. Figueira, S. Greco, and M. Ehrgott, Eds), Springer Verlag. London, 265-297.
6. Ehrgott, M., Wiecek, M. (2005). Multiobjective Programming. In: Multiple Criteria Decision Analysis: State of the Art Surveys, (J. Figueira, S. Greco, and M. Ehrgott, Eds), Springer Verlag. London, 990-1018.
7. Figueira, J., Mousseau, V., Roy, B. (2005). ELECTRE Methods. In: Multiple Criteria Decision Analysis: State of the Art Surveys, (J. Figueira, S. Greco, and M. Ehrgott, Eds), Springer Verlag. London, 133-162.
8. Genova, K., Vassilev, V., Andonov, F., Vassileva, M., Konstantinova, S. (2004). A Multicriteria Analysis Decision Support System. Proceedings of International Conference "CompSysTech", (Rachev B., Smrikarov A., Eds.), Rouse, Bulgaria, IIIA.10-1 – IIIA.10-6.

9. Knowles, J.D., Corne, D.W. (2000). M-PAES: A Memetic Algorithm for Multiobjective Optimization. *Proceedings of the 2000 Congress on Evolutionary Computation*, vol. 1, pp. 325-332, Piscataway, New Jersey, IEEE Service Center.
10. Korhonen, P. (2005). Interactive Methods. In: *Multiple Criteria Decision Analysis: State of the Art Surveys*, (J. Figueira, S. Greco, and M. Ehrgott, Eds), Springer Verlag. London, 642-665.
11. Miettinen K. (2003). Interactive Nonlinear Multiobjective Procedures, In: *Multiple Criteria Optimization: State of the Art Annotated Bibliographic Surveys*(Eds. Matthias Ehrgott and Xavier Gandibleux), Springer New York, 227-276.
12. Narula S.C., Vassilev, V., Genova, K., Vassileva, M. (2003). A Partition-Based Interactive Method to Solve Discrete Multicriteria Choice Problems, *Cybernetics and Information Technologies*, 2, 55-66.
13. Saaty, T. B. (2005). The Analytic Hierarchy and Analytic Network Processes for the Measurement of Intangible Criteria and for Decision-Making. In: *Multiple Criteria Decision Analysis: State of the Art Surveys*, (J. Figueira, S. Greco, and M. Ehrgott, Eds), Springer Verlag. London, 345-407.
14. Srinivas, N., Deb, K. (1994). Multiobjective Optimization Using Nondominated Sorting Genetic Algorithms. *Evolutionary Computation*, vol. 2, No. 3, pp. 221-248.
15. Vanderbei, R. (1996). [Linear Programming: Foundations and Extensions](#). Kluwer Academic Publishers.
16. Vassilev, V., Genova, K., Vassileva, M. (2005). A Multicriteria Decision Support System MULTIDECISION-1. In: *Proceedings of the XI-th International Conference "Knowledge – Dialogue - Solution"* (V. Gladin et al., Eds.). Varna, Bulgaria, 279-286.
17. Vassilev, V., Genova, K., Vassileva, M., Staikov, B., Andonov, F. (2005). MultiDecision-2.1: A Software System for Multicriteria Analysis and Optimization. In: *Preliminary Proceedings of the 5th International Conference on Decision Support for Telecommunications and Information Society*. Warsaw, Poland, 175-187.
18. Vassilev V., Konstantinova, S. (2005). Multicriteria Macroeconomic Ranking of Two Groups of European Countries by Decision Support System MKA-2. *Cybernetics and Information Technologies*, 1, 85-99.
19. Vassilev V., Genova K., Vassileva M., Staykov B., Andonov F. (2004). Software Decision Support System MKO-1. Working Papers IIT/WP-184.
20. Vassilev, V., Genova, K. (1991). An Algorithm of Integer Feasible Directions for Linear Integer Programming. *European Journal of Operational Research*, 52, 203-214
21. Vassileva, M. (2006). Generalized Interactive Algorithm for Multicriteria Optimization. *Problems of Engineering Cybernetics and Robotics*, 56, 69-77.
22. Vassileva, M. (2005). Generalized Scalarizing Problem of Multicriteria Optimization. *Comptes rendus de l'Academie bulgare des Sciences*, 58, 5, 537-544.
23. Vincke, P. (1992). *Multicriteria Decision-Aid*, John Wiley & Sons, New York.
24. Weistroffer, H., Smith, C., Narula, S. (2005). Multiple Criteria Decision Support Software. In: *Multiple Criteria Decision Analysis: State of the Art Surveys*, (J. Figueira, S. Greco, and M. Ehrgott, Eds), Springer Verlag. London, 990-1018.
25. Wolsey, L.A. (1998). *Integer Programming*. Wiley-Interscience.

Authors' Information

Vassil Vassilev, PhD – Institute of Information Technologies, BAS, Acad. G. Bonchev St., bl. 29A, Sofia 1113, Bulgaria; e-mail: vvassilev@iinf.bas.bg

Mariana Vassileva-Ivanova, PhD – Institute of Information Technologies, BAS, Acad. G. Bonchev St., bl. 29A, Sofia 1113, Bulgaria; e-mail: mvassileva@iinf.bas.bg

Boris Staykov, PhD student – Institute of Information Technologies, BAS, Acad. G. Bonchev St., bl. 29A, Sofia 1113, Bulgaria; e-mail: bstaykov@iinf.bas.bg

Krasimira Genova, PhD – Institute of Information Technologies, BAS, Acad. G. Bonchev St., bl. 29A, Sofia 1113, Bulgaria; e-mail: kgenova@iinf.bas.bg

Filip Andonov, PhD student – Institute of Information Technologies, BAS, Acad. G. Bonchev St., bl. 29A, Sofia 1113, Bulgaria; e-mail: fandonov@iinf.bas.bg and Assistant Professor, Department of Informatics, New Bulgarian University, Montevideo St. 21, bl.2, office 610, Sofia 1618; e-mail: fandonov@nbu.bg.

Plamena Chongova, programmer – Institute of Information Technologies, BAS, Acad. G. Bonchev St., bl. 29A, Sofia 1113, Bulgaria; e-mail: pchongova@iinf.bas.bg

ПРОБЛЕМЫ ПРОГНОЗИРОВАНИЯ ЭКОНОМИЧЕСКИХ МАКРОПАРАМЕТРОВ

Алексей Волошин, Виктория Сатыр

Аннотация: Обсуждаются проблемы прогнозирования экономических макропараметров на основе принципа «субъективного мультидетерминизма» - экспертного учета максимального количества взаимосвязанных «объективных» и «субъективных» причин. Дается описание системы поддержки принятия решений для прогнозирования уровня инфляции и валового внутреннего продукта на основе метода дерева решений.

Ключевые слова: Метод дерева решений; индекс инфляции; валовый внутренний продукт.

Введение

В работе [Волошин, 2006] предложена технология прогнозирования (путем использованием системы качественного прогнозирования на основе многопараметрических зависимостей, представляемых деревом решений [Волошин, 2005]), реализующая концепция «субъективного мультидетерминизма» (в [Волошин, 2006] – «множественный субъективный детерминизм»). В основе этой концепции следствие определяется множеством взаимозависимых причин (объективных и субъективных, в частности, деятельностью субъекта, которая, в свою очередь, определяется его возможностями, волей, желаниями, предпочтениями и т.п.). Такой подход, безусловно, не оригинален – подобные взгляды высказывались, в частности, Нобелевским лауреатом В.В. Леонтьевым («К вопросу о плюралистической интерпретации истории и проблеме междисциплинарного сотрудничества», Гарвард, 1948), предпринявшим попытку разработать собственную методологию «постижения истории» - «ни экономический, ни антропологический, ни, скажем, географический анализ не могут при современном состоянии развития соответствующих наук привести к единственно правильному утверждению». Конечно, как с иронией отмечает В.В. Леонтьев, проще всего будет, когда все науки сольются в некую единую отрасль знаний. Но поскольку «золотой век» наук еще не наступил, следует отработать методы «междисциплинарной кооперации». В.В. Леонтьев предлагает использовать для объяснения (прогнозирования) того или иного явления сразу нескольких специалистов, которые работают независимо друг от друга, после чего «ведущий проекта» должен сопоставить представленные ими анализы и составить из них некое максимально полное и объективное исследование. В [Волошин, 2006] подобный поход интерпретируется как «субъективизация объективности» и предлагается сделать еще один шаг – «объективизировать субъективность» (учитывая психосоматические особенности специалистов-экспертов). И основное – если в 1948 г. можно было лишь декларировать подобный «плюралистический метод прогнозирования», то через 60 лет развития вычислительной техники, математических методов (в частности, подходов, объединенных названием «искусственный интеллект»), социологии, психологии и т.д. позволяет подобный подход в значительной мере реализовать ([Волошин, 1999, 2001, 2003, 2005]). В [Волошин, 2002, 2003] дается описание инструментария создания прикладных систем поддержки принятия решений в различных областях таких, например, как прогнозирование экономических параметров [Волошин, 1999, 2003], медицинской диагностики [Волошин, 2005]. В докладе [Волошин, 2006] в качестве примера приводились предварительные результаты использования созданного инструментария для прогнозирования индекса инфляции в Украине. В данной работе рассматривается проблема прогнозирования макроэкономических параметров, в частности, индекса инфляции и внутреннего валового продукта на основе инструментария, описанного в работах [Волошин, 2005, 2006].

Инструментарий создания прикладных систем прогнозирования

В работах [Волошин, Пихотник, 1999], [Voloshin, Panchenko, 2001], [Волошин, Панченко, 2002], [Voloshin, Panchenko, 2003], [Волошин, 2005], [Волошин, Головня, 2005], большинство которых представлялось на конференциях KDS, развивается концепция «качественного прогнозирования на основе многопараметрических зависимостей, представляемых деревом решений» [Волошин, Панченко, 2002]. Считается, что следствие определяется множеством взаимозависимых причин, степень влияния которых на следствие определяется «субъективно» (экспертным измерением). Чем больше параметров, которые «формируют» причину, тем лучше (для адекватности модели). Однако это приводит к сложностям в анализе модели (возникает «проклятие размерности», с которым необходимо бороться [Волошин, Панченко, 2002], в частности, и методами искусственного интеллекта).

Построение прикладной системы поддержки принятия решений (СППР) сводится к выделению экспертами проблем и подпроблем (вершин дерева) и связей между ними (дуг дерева). Экспертами определяются веса (вероятности) переходов между вершинами. Допускаются нечеткие оценки экспертов с помощью логических переменных, описываемых значениями функции принадлежности (векторами действительных чисел от 0 до 1). Каждый эксперт задает три оценки – оптимистическую, реалистическую и пессимистическую, скаляризация которых осуществляется с учетом психологического типа эксперта. Тип определяется на основании психологических тестов, заложенных в систему. На основе психологических тестов определяются также коэффициенты «правдивости», «независимости», «осторожности» и т.д.

Дерево строится на основе коллективных оценок экспертов с применением метода парных сравнений. Для построения результирующего дерева применяются алгебраические методы обработки экспертной информации, в качестве расстояния между ранжировками применяется метрика Хемминга и мера несовпадений рангов объектов. Результирующее дерево определяется как медиана Кемени-Снелла или как компромисс [Волошин, 2005]. В случае задания приоритетов в нечеткой форме элементы матрицы задаются через функции принадлежности.

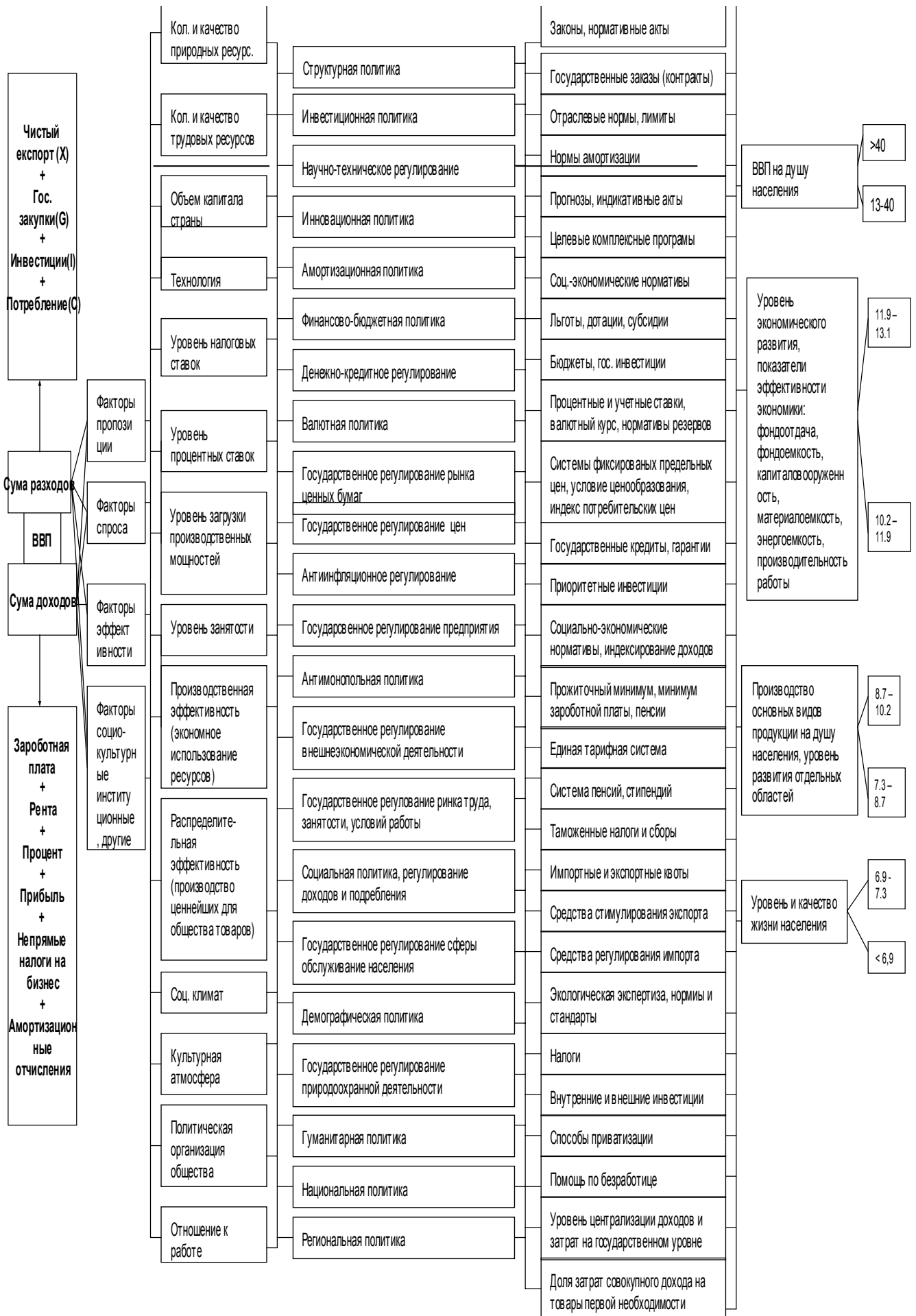
Для определения оптимальных путей в дереве предлагаются алгоритмы последовательного анализа вариантов [Волошин, Панченко, 2002], позволяющие обрабатывать деревья с сотнями вершин.

Дерево решений задается таблицами. Каждая таблица – это отдельный уровень дерева, каждая строка таблицы – отдельная вершина на этом уровне. Каждый элемент строки – это вероятность, с которой возможен переход из данной вершины в вершину нижнего уровня. Эти вероятности задаются функциями принадлежности, представляющие собой вектора действительных чисел от 0 до 1 любой длины. Таблица заполняется путем опроса экспертов. Существующие функции позволяют добавлять столбцы, строки, задавать словарь (который позволяет вербальным оценкам эксперта ставить в соответствие вероятности, путем задания определенных уровней), сохранять таблицы в файле, считывать таблицы из файла.

Экспертным путем задаются матрицы – результат сравнения вариантов вершин, которые могут быть включены в дерево. На основе анализа матриц определяются вершины, которые включаются в дерево и вероятности, с которыми возможен переход в них из вершин верхнего уровня. Если дерево решения декомпозируется на несколько поддеревьев, которые имеют одинаковые листья, вначале вычисляются вероятности этих листьев в каждом из них, а затем находятся вероятности для всего дерева в целом.

Прогнозирование индекса инфляции

Индекс инфляции является одним из основных макроэкономических параметров, в котором отражаются основные тенденции развития экономики. Универсальных и совершенных подходов к решению проблемы его прогнозирования сегодня не существует.



Методы количественного прогнозирования (временные ряды, регрессионный анализ, имитационное моделирование и т.п.), в основе которых лежит "продолжение прошлого", дают плохие результаты при прогнозировании "нестабильных" процессов, характеризующихся "нарушением монотонности", в основе которого лежат скачкообразные изменения, нехарактерные для развития процесса в прошлом. Проблема состоит в репрезентации будущего, которое не может интерпретироваться как обычное продолжение прошлого, поскольку будущее может приобретать принципиально новые формы. В основе такого прогнозирования ("качественного прогнозирования") лежит идея непосредственного использования знаний человека (эксперта). При этом, в первую очередь, следует учесть "нечеткость" экспертной информации, которая в свою очередь зависит от его профессиональных и психологических характеристик (компетентности, независимости, объективности, реализма, склонности к риску и т.д.). Поэтому для прогнозирования индекса инфляции применялся инструментарий, описанный в предыдущем разделе. Фрагмент дерева решений для прогнозирования индекса инфляции представлен ниже.

Для прогнозирования индекса инфляции выделяются следующие основные подпроблемы:

- Экономические – состояние промышленности, состояние аграрно-промышленного комплекса, состояние финансового рынка, торговля и т. д.;
- Политические – тенезация экономики, инвестиционная, валютная, антимонопольная политика и т.д.;
- Социально-демографическое состояние - безработица, социально-демографическая нагрузка, темп прироста населения и т.д.;
- Финансовые – финансово-бюджетная, валютная политика, государственное регулирование рынка ценных бумаг, государственное регулирование цен и т.д.

Прогнозное значение индекса инфляции в Украине на 2005 год, полученное нами в июне 2005 г., равнялось 12,8%. В бюджете Украины было заложено 9,8%, официальная статистика по результатам 2005 года дала 10,5%, международные эксперты – 12,5-13,0%. Высокая точность прогноза, на наш взгляд, обуславливается «объективной» причиной – учетом большого числа разнородных взаимосвязанных причин, влияющих на результат. Вторая причина, по мнению авторов, «субъективная» - это «узкоспециальная» экспертная оценка, эксперт зачастую и не догадывается, что он прогнозирует в конечном счете т. Прогноз значения индекса инфляции на 01.01.2008 г., полученное нами в апреле 2007 г., равнялось 17,3%. Интересно отметить, что Национальный банк Украины прогнозирует инфляцию в Украине в текущем году на уровне 7%, правительство – 8%, Президент Украины озвучил в марте 2007 г. уровень инфляции в 11-12%. Остается подождать KDS-2008 и сравнить прогнозы!

Определение показателя валового внутреннего продукта

Как известно, показатель ВВП можно определять тремя основными методами: по созданной продукции (производственный метод); по затратам (метод конечного использования); по доходам (распределительный метод).

В большинстве стран основными являются первые два метода. Выбор того или иного метода определяется наличием надежной информационной базы. В современных условиях в Украине «надежная информационная база» отсутствует, можно говорить не столько о точности статистических данных, сколько об их «полной неточности» (а, зачастую, и их отсутствии). Одним из основных, «искажающих» статистику факторов является теневой сектор экономики. Его «интегральная» экспертная оценка (от 40% до 60%) слишком неточна для использования в расчетах ВВП. Выход – в использовании принципа «косвенного расчета», используя (по В. В. Леонтьеву) метод «междисциплинарной кооперации» - длинные цепи событий естественно разбиваются на группы непосредственно связанных событий; каждая из них

изучается и объясняется отдельно, тип объяснения будет меняться при переходе от одной группы связей к другой; чисто «экономическая» интерпретация может оказаться наиболее подходящей для одной группы событий, чисто «политическая» или «социальная» – другой.

Для расчета (прогнозирования) ВВП было определено по возможности максимальное число групп-факторов («подпроблем», «поддеревьев» в дереве решений), прямо или косвенно влияющих на значение ВВП. В свою очередь, в поддеревьях определялось максимально возможное количество вершин, потенциально определяющих данную группу факторов. При принятии экспертного решения о включении конкретной вершины в поддерево, учитывалось, в частности, наличие (точность, достоверность, возможность получения) оценки данного фактора.

Учитывались следующие группы факторов и факторы в группе:

1. Реальный сектор (расходы за категориями расходов, расходы на конечное потребление домохозяйств, конечное потребление сектора общего государственного управления, индивидуальное конечное потребление, коллективное конечное потребление, валовое накопление основного капитала, изменение запасов материально-технических средств, экспорт товаров и услуг, импорт товаров и услуг, индекс промышленного производства, занятость, безработица, средняя заработная плата, индекс потребительских цен, индекс цен производителей);

2. Бюджетно-налоговый сектор (операции сектора общего государственного управления – общее сальдо, операции центральных органов государственного управления – общее сальдо, финансирование, за счет внешних источников, за счет внутренних источников, банковскими учреждениями, небанковскими учреждениями, поступление от приватизации государственного имущества, государственный и гарантированный государством внутренний долг, государственные внутренние долговые обязательства по начальным срокам погашения, краткосрочные государственные внутренние долговые обязательства (сроком до 1 года), среднесрочные и долгосрочные государственные внутренние долговые обязательства, государственные ценные бумаги, кредиты, долговые обязательства, гарантированные правительством Украины, ссуды международных финансовых организаций, официальные ссуды);

3. Финансовый сектор (аналитические счета банковского сектора по состоянию на конец периода, денежная масса, внутренний кредит, требования к Правительству, требования к другим секторам, чистые внешние активы, аналитические счета центрального банка, по состоянию на конец периода, денежная база, чистые требования к Правительству, обязательства перед правительством, требования к банкам, требования к другим секторам, внешние активы, внешние пассивы, процентные ставки, Национальный банк Украины, установлена учетная ставка, фактическая ставка за инструментами, банки Украины, за предоставленными кредитами на межбанковском рынке в национальной валюте, ставка за привлеченными депозитами в национальной валюте, ставка по предоставленным кредитам в национальной валюте).

4. Внешний сектор (платежный баланс по состоянию на конец периода, счет текущих операций, баланс товаров и услуг, доходы, текущие трансферты, счет операций с капиталом, финансовый счет, инвестиции из Украины (прямые + портфельные + другие), инвестиции в Украину (прямые + портфельные + другие), ошибки и упущения, международные резервы, по состоянию на конец периода, резервы в иностранной валюте, резервная позиция в МВФ, специальные права заимствования, золото, другие резервные активы, внешняя торговля товарами, валовой внешний долг, по состоянию на конец периода, валютный курс, курс доллара США относительно евро, средний за год).

5. Социальное состояние населения (средний размер месячной пенсии, средний размер месячной заработной платы, государственная помощь семьям с детьми, рынок труда, экономическая активность

населения, зарегистрирован рынок труда, безработица, демографическая ситуация, численность населения, естественное движение населения, миграционное движение населения, образование, дошкольные заведения, профессионально-технические учебные, высшие учебные заведения).

Наш расчет ВВП за 1-ый квартал 2007 г. дал оценку в 150 млрд. грн., официальная статистика (с учетом экспертной оценки по теневому сектору) – на 10% ниже. Этот факт можно интерпретировать как ошибку «интегральной официальной» экспертной оценки. Интересно отметить, что и прогнозные значения ВВП за 2006 год, выполненный в мае-июне 2006 года, дал те же 10% расхождения с официальной оценкой Государственного комитета по статистике. Безусловно, о точности нашего прогноза «истинному» значению тут речи не идет (хотя бы из-за отсутствия последнего). Было бы интересно провести исследования на «истинных» данных (хотя бы для одного сектора) для «обучения» системы – определения достоверных значений взаимовлияния факторов.

Заключение

Название и содержание конференции KDS полностью соответствует принципу «субъективного мультидетерминизма», используемого в работе для прогнозирования экономических макропараметров: ЗНАНИЯ (специалистов различных отраслей науки и практики) – ДИАЛОГ (между ними, с помощью специалистов по принятию решений и искусственному интеллекту) – РЕШЕНИЕ (проблем человечества).

Библиография

- [Волошин, 1999] Волошин А., Пихотник Е. Экспертная система прогнозирования курса гривны. "Искусственный интеллект", 1999, №2. С.354-359 (укр.).
- [Voloshin, 2001] Voloshin O.F., Panchenko M.V. The Forecasting of Stable Processes by a Tree Solution Method using a Pairwise Comparison Method for Analysis of Expert Information. Труды международной конференции «KDS-2001», Том 1, Санкт-Петербург, 2001. С.50-53 (англ.).
- [Волошин, 2002] Волошин А.Ф., Панченко М.В. Экспертная система качественного оценивания на основе многопараметрических зависимостей. "Проблемы математических машин и систем", 2002, №2. С.83-89 (укр.).
- [Voloshin, 2003] Voloshin O.F., Panchenko M.V. The System of Quality Prediction on the Basis of a Fuzzy Data and Psychography of the Experts. International Journal "Information Theories & Applications", 2003, Vol.10, №3. P.261-265.
- [Волошин, 2005] Волошин А.Ф. О проблемах принятия решений в социально-экономических системах. Труды конференции «KDS-2005», Том 1, София, 2005. С.205-212.
- [Волошин, Головня, 2005] Волошин А.Ф., Головня В.М. Система качественного прогнозирования на основе нечетких данных и психологии экспертов. Труды конференции «KDS-2005», Том 1, София, 2005. С.237-243.
- [Волошин, 2006] Волошин А.Ф. Системы поддержки принятия решений как персональный интеллектуальный инструментальный лица, принимающего решение Труды конференции «KDS-2006», София, 2006. С.149-153.

Информация об авторах

Волошин Алексей Федорович – Киевский национальный университет имени Тараса Шевченко, факультет кибернетики, профессор. Киев, Украина. E-mail: ovoloshin@unicyb.kiev.ua

Сатыр Виктория Валериевна – Киевский национальный университет имени Тараса Шевченко, факультет кибернетики, магистр. Киев, Украина. E-mail: Brili@ukr.net

ИНТЕЛЛЕКТУАЛЬНЫЕ ТЕХНОЛОГИИ В МАРКЕТИНГОВОМ АНАЛИЗЕ

Галина Сетлак

Аннотация: Целью данной работы является разработка процедур и алгоритмов использования методов искусственного интеллекта для обработки и анализа различных видов знаний. Предлагается интегрировать такие техники моделирования знаний и процессов принятия решений как нейронные сети и системы нечёткого вывода. Нейронная нечёткая система используется для решения задачи идентификации сложных процессов. Представлен пример применения к задаче классификации в анализе рынка.

Keywords: Data mining, классификация, нейронные сети, нечёткая логика, маркетинг, анализ рынка.

ACM Classification Keywords: I. Computing Methodologies, I.2 Artificial Intelligence

Введение

В последнем десятилетии в развитии средств и методов обработки информации появилась новая научная дисциплина, названная Data Mining, что переводится как „добыча” или „извлечение данных”, которая направлена прежде всего на автоматизацию обработки больших объёмов информации. Data Mining иначе ещё называется в литературе интеллектуальным анализом данных. Интеллектуальный анализ данных означает процесс обнаружения в „сырых данных” (raw data) ранее неизвестных, практически полезных и доступных интерпретации знаний, закономерностей, необходимых для принятия решений в различных сферах человеческой деятельности. При этом выделяются пять стандартных типов закономерностей, которые позволяют выявлять методы Data Mining: ассоциация, последовательность, классификация, кластеризация и прогнозирование. Основой для новой парадигмы современной технологии анализа информации являются последние достижения в области развития методов искусственного интеллекта.

В настоящих исследованиях для обработки и анализа информации предлагается интегрировать методы искусственного интеллекта, используемые в таких научных направлениях, как нейронные сети (искусственные аналоги человеческого мозга, моделирующие способность обучаться), теория нечётких множеств и логики – средства формализации естественно-языковых высказываний и нечёткого логического вывода, а также генетические алгоритмы. Достоинством таких гибридных нейронно-нечётких систем является гибкость, способность к адаптации за счет различных методов обучения, простота реализации, удобство при настройке, а также высокие качественные показатели.

Так как нечеткие системы работают со слабо структурированной качественной информацией, а нейронные сети используют только количественную информацию, объединение этих двух методов позволит использовать всю доступную информацию об объекте. Технология, разработанная на этой основе, объединяет соответствующим образом способность нейронных сетей к самообучению и способность нечетких систем обрабатывать качественную информацию. Несомненное достоинство гибридной нейронно-нечёткой технологии моделирования и обработки информации заключается в возможности просмотра сформированных правил и придания им содержательной, (лингвистической) интерпретации, что позволяет рассматривать аппарат нейронно-нечётких систем как средство извлечения знаний из экспериментальных баз данных.

Целью данной работы является оценка возможностей использования нечётких нейронных систем для предварительной обработки и анализа информации разного вида и решения задач классификации. Решается практическая задача маркетингового анализа рынка при помощи нейронной нечёткой системы.

Нейросетевые системы нечёткой классификации

Проблемы классификации являются одними из наиболее часто решаемых задач, как в повседневной хозяйственной деятельности, так и в экономическом анализе этой деятельности каждой организации. Решение задач классификации заключается в разработке, конструкции таких правил и закономерностей, которые позволили бы распознать определённые явления или объекты и определить их принадлежность к каким-то классам (группам, категориям). Задача значительно усложняется, если при определении классов необходимо брать во внимание очень большое количество характеристик исследуемых объектов и явлений, тем более, если эти характеристики слабо определены или трудно формализуемы. В последние годы для решения сложных задач классификации всё чаще используются искусственные нейронные сети.

В работах автора данной статьи ранее исследовались возможности использования нескольких видов искусственных нейронных сетей для решения практических задач классификации. Проанализированы и оценены такие нейронные сети как: рекуррентная сеть Хамминга [Сетлак Г., 2000], многослойный перцептрон, сеть радиальной базисной функции, вероятностные и сети Кохонена в работе [Сетлак Г., 2004]. Результаты анализа показали, что для корректной работы и реального применения нейронных сетей необходимо их обучать на достаточно обширных выборках входных данных, причём, чем больше входных анализируемых признаков, чем выше сложность сети и моделируемой функции, тем больше должна быть выборка. При решении задач классификации в менеджменте, где рассматриваемые экономические показатели характеризуются большим разнообразием и субъективной природой, необходимо также использовать качественные параметры, основанные на психологических особенностях восприятия, а не только численные или логические. Наиболее обещающим направлением в исследованиях и разработке таких интеллектуальных инструментов представляет собой объединение двух независимых интеллектуальных технологий: искусственных нейронных сетей и нечёткой логики [Jang J. S.R., Sun C.T., Mizutani E., 1997], [Grabmeier J., A. Rudolph, 2002], [Rutkowska D., 1997].

Теория нечётких множеств и нечеткой логики согласно с основной идеей Л.Заде в течение почти сорока лет широко используется как инструмент для моделирования и обработки нечёткой, лингвистической или иначе так называемой, качественной информации, для моделирования нечёткости мышления человека, его способности использовать приближённые оценки для описания сложных, плохо формализуемых процессов принятия решений в различных областях деятельности. Интеграция двух независимых интеллектуальных технологий: искусственных нейронных сетей и теории нечётких множеств и логики, таким образом, кажется наиболее натуральным процессом, который позволяет создать новую, более универсальную методологию обработки различных видов знаний и моделирования сложных процессов.

Постановка задачи

Принимаем, что известны следующие данные:

- входные переменные $X = (x_1, x_2, \dots, x_n)^T$, причём $X \in X_1 \times X_2 \times \dots \times X_n$ - являются терм-множествами переменной x_i , которые являются характерными признаками классифицируемых объектов.
- множество классов решений $S = \{S_1, S_2, \dots, S_m\}$ соответствующих выходной переменной y , $y \in Y$.
- функции принадлежности $\mu_{A_i^k}(x)$ конкретных значений входных переменных x_i к нечётким множествам A_i^k , ($i = 1, 2, \dots, n$, $k = 1, 2, \dots, N$),

- Нечёткие предикатные правила зависимостей, связывающих значения входных переменных с одним из возможных значений выходной переменной, разработанные экспертом или иначе так называемая нечеткая база знаний - логических высказываний типа „ЕСЛИ-ТО”.

Требуется разработать алгоритм, позволяющий определить для вектора входных переменных класс решений, к которому можем его отнести.

Рассмотрим нечётко-нейронную систему, структура которой представлена на рис.1.

Определение 1: нечётко-нейронной системой называем многослойную структуру, в состав которой входят следующие элементы:

- Система нечёткого вывода, на вход которой подаются лингвистические переменные (термы). В состав каждой нечеткой системы выводов входят следующие элементы [Zadeh L.A., Kasprzyk J., 1992] блок введения к нечеткости (fuzzification), база правил, содержащая совокупность нечетких правил, механизм выводов и блок приведения к четкости (defuzzification).
- Нечёткая система обучается оптимизационными методами: модифицированным алгоритмом Обратного распространения ошибок Backpropagation или при помощи гибридного метода основанного на генетическом алгоритме, используемых в нейронных сетях.
- В выходном слое нечётко-нейронной системы используется нейронная сеть для выполнения операции приведения к чёткости.

Определение 2. Нечеткая база знаний представляет собой совокупность правил „ЕСЛИ <входы> ТО <выход>”, которые отображают опыт эксперта (специалиста предметной области) и его понимание причинно-следственных связей, характерных для моделируемого объекта или процесса.

Механизм нечетких выводов рассматриваемой системы использует базу знаний в виде совокупности $R^{(k)}$, ($k=1,2,\dots,N$) нечетких предикатных правил вида:

$$R^{(k)}: \text{IF } x_1 \text{ is } A_1^k \text{ and } x_2 \text{ is } A_2^k \text{ and } \dots \text{ and } x_n \text{ is } A_n^k \text{ THEN } (y \text{ is } B^k), \quad (1)$$

в которых N - означает число нечётких правил, x_1, x_2, \dots, x_n, y - лингвистические переменные, A_i^k - нечёткие множества предпосылок правил ($i = 1, 2, \dots, n; k=1, 2, \dots, N$) и B^k – нечёткое множество заключений.

Обозначим также $A^k = A_1^k \times A_2^k \times \dots \times A_n^k$. Символами X_i и Y обозначаем пространства изменений входных и соответственно выходной переменных. Знания эксперта, содержащиеся в (1), отражают нечёткое отношение предпосылки и заключения. Поэтому его можем записать как нечёткую импликацию: $R^{(k)}: A^k \rightarrow B^k$, $k=1, 2, \dots, N$ и представить как нечёткое отношение, определённое на множестве $X \times Y$, т.е. $R^{(k)} \subseteq X \times Y$ является нечётким множеством с функцией принадлежности следующего вида:

$$\mu_{R^{(k)}}(x, y) = \mu_{A^k \rightarrow B^k}(x, y). \quad (2)$$

В рассматриваемой нечётко-нейронной системе (представленной на рис.1) выполняется операция введения нечёткости (fuzzyfication) типа синглетон (singleton), что означает нахождение степени истинности для предпосылок каждого правила и определение для этого функции принадлежности $\mu_i^k(x)$

нечёткого множества $A_i^k \subseteq X = X_1 \times X_2 \times \dots \times X_n$ на входе системы нечёткого вывода выполняется

по следующей формуле: $\mu_i^k(x) = \begin{cases} 1 & \text{если } x = x^* \\ 0 & \text{если } x \neq x^* \end{cases}$, где $x = (x_1, x_2, \dots, x_n)^T \in X$,

лингвистические переменные, а $X^* = (x_1^*, x_2^*, \dots, x_n^*)^T$ входные переменные нечёткой системы вывода.

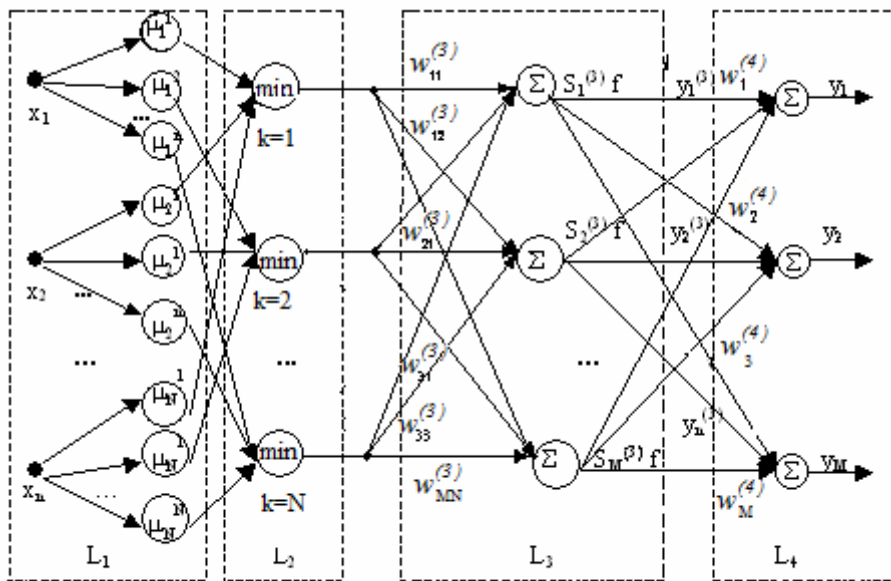


Рисунок 1. Нейронная нечёткая система

В системе элементы в слое L_1 определяют функции принадлежности (обозначенные μ_i^k) нечётких множеств $A_i^k, i=1,2,\dots,n, k=1,2,\dots,N$, где n - количество входов в системе, N – количество нечётких правил. Выходы в L_1 представляют собой функции принадлежности $\mu_{A_i^k}(x)$ для конкретных значений x_i^* поданных на i -тый вход. В виде функции принадлежности в данных исследованиях использовались функции вида гауссова, определяемые по формуле:

$$\mu_{A_i^k}(x) = \exp \left[- \left(\frac{x_i^* - \overline{x_i^k}}{\sigma_i^k} \right)^2 \right], \quad (3)$$

где $\overline{x_i^k}$ - является наиболее возможным значением переменной x , а величина σ_i^k представляет собой коэффициент концентрации или растяжения функции. Значения этих параметров корректируются в процессе обучения нечёткой нейронной системы.

В системе элементы второго слоя L_2 (обозначенные \min) реализуют нечёткий вывод согласно с методом нечёткого вывода Мамдани. В настоящее время в литературе кроме выше представленного алгоритма нечёткого вывода известны еще такие методы [Zadeh L.A., Kasprzyk J., 1992], [Rutkowska D., 1997]: алгоритм Larsena, алгоритм Sugeno, импликация Yager, импликация Zadeh, импликация Lukasiewicz, алгоритм TSK (Takagi-Sugeno-Kanga) и другие.

На выходах элементов второго слоя L_2 получаем степени истинности предпосылок каждого нечёткого правила $R^{(k)}$ в системе, которые вычисляются следующим образом:

$$\tau_k = \min_{1 \leq i \leq n} \left\{ \mu_{A_i^k}(x_i^*) \right\}. \quad (4)$$

Количество элементов в этом слое равно количеству нечётких правил N . Элементами слоёв L_3 и L_4 являются обычные нейроны. В этих слоях выполняется классификация, а затем приведение к чёткости (defuzzification). На вход слоя L_3 подаются степени истинности предпосылок нечётких правил $\tau_k, k=1,2,\dots,N$, вычисляемые по формуле (4). В нем имеется M -нейронов, где M – количество классов, которые

осуществляют взвешенное суммирование значений выходов нейронов предыдущего слоя. А их выходы формируются с использованием активационных функций. В качестве функции активации нейронов используется сигмоидальная функция следующего вида:

$$f(\alpha) = \frac{1}{1 + e^{-\alpha}}. \quad (5)$$

Выходы нейронов в этом слое интерпретируем как степени принадлежности к определённым классам. Следует отметить, что нечёткая классификация предполагает, что не всегда можно требовать принадлежность только к одной выбранной категории. Допускается возможность принадлежности данного образца в определённой степени к двум или трём классам. Необходимо в связи с этим разработать такой способ определения желаемых значений выходов во время обучения сети, учитывающий эту модификацию. В работе рассматривалась такая процедура, предложенная *Mitra i Pal*. Синаптические веса нейронов определяются в процессе обучения сети. Процедуры расчёта оптимальных значений синаптических весов выбираем в зависимости от того, какой алгоритм используем для обучения нечёткой нейронной сети. Если используется алгоритм Обратного Распространения Ошибок (*Back-propagation*), то расчёт оптимальных значений синаптических весов определяем следующим образом. Ошибку на выходе слоя L_4 определяем по формуле:

$$Q_r = \frac{1}{2}(y_r - d_r)^2 \quad \text{для } r=1,2,3,\dots,M, \quad (6)$$

где $y_r^{(4)}$ для слоя L_4 определяем как

$$y_r^{(4)} = f\left(\sum_{r=1}^M w_r^{(4)} y_r^{(3)} + w_0^{(4)}\right), \quad (7)$$

а также аналогично определяем для слоя L_3 . Полного вывода всех процедур и зависимостей в данной работе представлять нет возможности, поэтому рассмотрим только основные необходимые расчёты. Рекуррентные зависимости определяющие оптимальные значения синаптических весов в L_4 согласно с алгоритмом *Backpropagation* представляются такими формулами:

$$w_r^{(4)}(t+1) = w_r^{(4)}(t) - \eta Q_r^{(4)}(t) y_r^{(3)}(t), \quad (8)$$

где $\eta \in (0,1)$, а $Q_r^{(4)}$ рассчитывается по следующей формуле:

$$Q_r^{(4)}(t) = (y_r^{(4)} - d_r) f'(s^{(4)})|_t \quad (9)$$

и означает ошибку в L_4 слое. Аналогично рассчитываются оптимальные значения весов в слое L_3 . Следующим этапом вычислений является определение оптимальных значений параметров функций принадлежности $\mu_{A_i^k}$, $i=1,2,\dots,n$, $k=1,2,\dots,N$, реализуемых в первом слое L_1 .

Чтобы определить процедуру обучения параметров функций принадлежности для представленной нечёткой нейронной системы, выполняем обратный проход алгоритма обучения до первого слоя L_1 . Это приводит к необходимости расчёта производной для функции *min* (если это система нечёткого вывода по методу Mamdani, что выполнялось в настоящих исследованиях), которая, как известно, не дифференцируема. В таком случае воспользуемся операцией так называемой *softmin*, т.е. заменим определение *min* приближённой операцией, которая выполняется по следующей формуле:

$$\text{soft min}_{1 \leq i \leq n} = \frac{\sum_{i=1}^n \chi_i^k e^{-\varpi \chi_i^k}}{\sum_{i=1}^n e^{-\varpi \chi_i^k}}. \quad (10)$$

Тогда нетрудно доказать, что

$$\frac{\partial \chi_i^k}{\partial \chi_i^k} \left(\min_{1 \leq j \leq n} \{ \chi_j^k \} \right) = \frac{e^{-\varpi \chi_i^k}}{\sum_{j=1}^n e^{-\varpi \chi_j^k}}. \quad (11)$$

Как отмечается во многих публикациях, выше представленный алгоритм обучения нечёткой нейронной сети, основанный на алгоритме Back-propagation не гарантирует достижения глобального экстремума оценки качества. В исследованиях предлагается гибридный подход, разработанный на основе [7], применяющий генетический алгоритм для нахождения решений близких глобальному оптимуму, которые потом используются как начальная точка для алгоритма обратного распространения ошибок, в результате работы которого находится действительный глобальный экстремум.

Обучение нечёткой нейронной сети при помощи гибридного метода с использованием генетического алгоритма подробно описан в публикации автора данной статьи [Сетлак Г., 2004].

Результаты исследований

Выше представленная нейронная нечёткая система, которая является составным модулем интеллектуальной системы поддержки принятия решений IDSS (Intelligent Decision Support System) [Сетлак Г., 2004], использовалась для решения задач маркетингового анализа рынка и оценки положения фирмы на нём. В настоящей работе представим анализ рынка сбыта изделий домашнего хозяйства предприятия ХХХ. В качестве анализируемых наблюдений используем информацию о динамике реализации продукции (продаже) изделий, а именно пылесосов за 1999 -2001 года [таблица 1]. На основе исторических данных система обнаруживает сложные зависимости между затратами на рекламу, объемом продаж, ценой изделий, затратами на стимулирование сбыта. В результате оценки этих параметров на выходе системы классификации определяются причины уменьшения объема продаж и падение спроса на рынках.

Таблица 1. Анализируемые данные о реализации продукции предприятием ХХХ

Год/квартал	Объем продаж (штуки)	Затраты на стимулирование сбыта	Средняя цена (злотые)	Затраты на рекламу (в злотых)
1999 / I	227180	10000	385,65	12000
1999 / II	235090	6000	397,24	10000
1999 / III	217340	10000	452,20	10000
1999 / IV	261280	8000	478,92	12000
2000 / I	184380	5000	493,10	10000
2000 / II	147180	4000	526,35	8000
2000 / III	149300	3000	583,24	5000
2000 / IV	156520	4000	594,93	5000
2001 / I	121280	2000	620,70	5000
2001 / II	116530	0	634,56	10000
2001 / III	102160	0	663,20	2000
2001 / IV	112510	0	672,35	0

На вход нечёткой нейронной системы подаются следующие лингвистическая переменные:

X_1 – Цена (высокая, умеренная, низкая).

X_2 – Затраты на рекламу (высокие, средние, низкие).

X_3 – Эффективность рекламы (высокая, средняя, низкая), определяется как реакция изменения величины продажи к издержкам на рекламу, (возрастание или убывание).

X_4 – объём продаж (большой, но невозрастающий, средний, маленький).

X_5 – Затраты на стимулирование сбыта (высокие, средние, низкие).

Выходными параметрами классификации являются четыре класса, каждый из которых определяет одну из причин уменьшения объёма продаж и падение спроса на рынках:

Y_1 – Причина1 – (ошибочная маркетинговая стратегия), обозначено P1 .

Y_2 – Причина2 – (Мало эффективный вид рекламы), обозначено P2.

Y_3 – Причина3 – (Сезонное падение продаж), обозначено P3.

Y_4 – Причина4 – (Ошибочное ценообразование), обозначено P4.

Для моделирования функции принадлежности использовались гауссовы функции, показанные на рисунке 2, заданные формулой (3). На основе нечёткой базы правил, сформированной экспертами, строится нечёткая база знаний и разрабатывается система логического вывода. В качестве обучающей выборки для настройки нечёткой нейронной системы использовались результаты маркетингового анализа, выполненного для предприятия ХХХ фирмой экономического консалтинга в 2002 году, а также полученные в результате этого анализа оценки экспертов.

Исследовались следующие структуры нечётких нейронных систем:

- В этой структуре использовались гауссовы функции принадлежности вида (3), нечёткий вывод выполнялся методом Mamdani (т.е. импликация в форме нахождения минимум: $\min[\mu_{A_i^k}(x), \mu_{B^k}(y)]$), параметры функции принадлежности и нечёткая нейронная сеть обучались градиентным методом и гибридным, основанным на совместном использовании генетического алгоритма и Back-propagation .
- нечёткий вывод выполнялся методом Larsen'a (т.е. импликация в форме произведения $\mu_{A_i^k}(x) \cdot \mu_{B^k}(y)$), параметры функции принадлежности и нечёткая нейронная сеть обучались таким же образом, как и в выше описанной структуре. Общее число нечётких логических высказываний составляет $N = 26$.

Полученные результаты работы нечётко-нейронного классификатора представлены в таблице 2.

Практическая реализация и анализ выполнены с использованием программных пакетов NeuroSolutions CS и Genetic Library фирмы NeuroDimension [NeuroSolutions, 2002], а также нечёткая нейронная сеть моделировалась при помощи Fuzzy Logic Toolbox for MATLAB, программные модули интерфейса разработаны с использованием MS Visual C++.

Следует отметить, что в позициях 5,7,9,10, т.е. там, где определены две или больше причин падения цен, функции принадлежности к определённым классам были приблизительно равны. Возможность определения степени принадлежности некоторых переменных к двум или более классов является отличительной особенностью и преимуществом нечётко-нейронной классификации.

Таблица 2. Результаты работы нечётко-нейронного классификатора

N	X ₁	X ₂	X ₃	X ₄	X ₅	y
1	385,65	12000	Убывание	227180	10000	P1
2	397,24	10000	Незнач.рост	235090	6000	P1
3	452,20	10000	Убывание	217340	10000	P4
4	478,92	12000	Незнач.рост	261280	8000	P3
5	493,10	10000	Убывание	184380	5000	P3,P4
6	526,35	8000	Убывание	147180	4000	P2
7	583,24	5000	Убывание	149300	3000	P2,P3
8	594,93	5000	Незнач.рост	156520	4000	P1
9	620,70	5000	Убывание	121280	2000	P1,P2
10	634,56	10000	Убывание	116530	0	P2,P3

Выводы

В данной работе разработаны и представлены методы и алгоритмы обработки и анализа знаний разного вида (лингвистических, количественных, смешанных), основанные на нейронно-нечётких системах. Разработанная гибридная интеллектуальная система используется для решения задач классификации. Для обучения нечётких нейронных систем классификации используется Алгоритм обратного распространения ошибки (Back-propagation) и гибридный алгоритм, основанный на совместном использовании генетического алгоритма и метода Back-propagation.

Создан программный пакет для решения задач маркетингового анализа рынка и поддержки принятия решений в нечёткой среде. В результате оценки ретроспективной информации определяются причины понижения продажи изделий. Оценивается маркетинговая стратегия в фирме и политика ценообразования.

Важной особенностью представленных алгоритмов, входящих в состав нечёткой нейронной системы, является гибкость в отношении исходных данных для моделирования, способность использовать различные источники знаний. Предложенный подход к извлечению, представлению и обработке знаний, содержащих нечеткость, существенно повышает эффективность методологии автоматизированного построения интеллектуальных систем поддержки принятия решений. Все описанные модели, методы и процедуры программно реализованы и проходят этапы экспериментальной апробации, доработки и модификации.

Литература

- [Jang J. S.R., Sun C.T., Mizutani E.,1997] Jang J. S.R., Sun C.T., Mizutani E.: Neurofuzzy and Soft Computing, Prentice-Hall, Upper Saddle River 1997, p. 245.
- [Grabmeier J., A. Rudolph, 2002] Grabmeier J., A. Rudolph: Techniques of cluster algorithms in data mining, Data Mining and Knowledge Discovery, 2002, nr. 4, 303–360.
- [NeuroSolutions, 2002] NeuroSolutions, The Neural Network Simulation Environment, Copyright NeuroDimension Inc., Gainesville, 2002, 67 pp.
- [Rutkowska D., 1997] Rutkowska D.: Inteligentne systemy obliczeniowe, algorytmy genetyczne i Sieci neuronowe w systemach rozmytych, Akademicka Oficyna Wydawnicza PLJ, Warszawa, 1997r., 304 str.

[Сетлак Г., 2000] Сетлак Г.: „Нейронные сети в интеллектуальных системах управления производством”, Международный научно-технический журнал «Проблемы управления и информатики», N1, 2000, str.112-119, Киев.

[Сетлак Г., 2004] Сетлак Г.: Интеллектуальные системы поддержки принятия решений, Изд. ЛОГОС, Киев, 2004, стр.250.

[Zadeh L.A., Kacprzyk J., 1992] Zadeh L.A., Kacprzyk J.(ed.): Fuzzy logic for the Management of Uncertainty, Wiley, New York, 1992, p. 492.

Информация об авторе

Galina Setlak – Ph.D., D.Sc, Eng., Associate Professor, Rzeszow University of Technology, Department of Computer Science, W. Pola 2 Rzeszow 35-959, Poland, Phone: (48-17)- 86-51-433, gsetlak@prz.edu.pl

АНАЛИЗ И СРАВНЕНИЕ РЕЗУЛЬТАТОВ ОПТИМИЗАЦИИ ИНВЕСТИЦИОННОГО ПОРТФЕЛЯ ПРИ ПРИМЕНЕНИИ МОДЕЛИ МАРКОВИЦА И НЕЧЕТКО-МНОЖЕСТВЕННОГО МЕТОДА

Юрий Зайченко, Малихех Есфандиярфард

Аннотация: Рассмотрена задача определения оптимального инвестиционного портфеля в условиях неопределенности. Построена математическая модель задачи с использованием аппарата нечетких множеств. Предложен алгоритм оптимизации нечеткого инвестиционного портфеля. Проведены экспериментальные исследования предложенного метода оптимизации нечеткого инвестиционного портфеля и выполнен сравнительный анализ полученного решения с решением классической задачи Марковича.

Key words: portfolio optimization, fuzzy portfolio, Markovitz problem

Введение

Одной из важнейших задач управления финансовыми активами является задача оптимизации инвестиционного портфеля. Считается, что началом современной теории инвестиций стала вышедшая в 1952 г. статья Г. Марковитца «Выбор портфеля», в который впервые была предложена математическая модель формирования оптимального портфеля ценных бумаг и методы построения таких портфелей при определенных условиях на основе теоретико-вероятностной формализации понятия доходности и риска.

Однако прокатившиеся по всему миру рыночные кризисы 1997-1998 и 2000-2001 года, принесшие только американским инвесторам убытки в 10 триллионов долларов, показали, что существующие теории оптимизации фондовых портфелей и прогнозирования фондовых индексов себя исчерпали, и необходима существенная ревизия методов фондового менеджмента.

Таким образом, в свете явной недостаточности имеющихся научных методов для управления финансовыми активами, потребовалась разработка принципиально новой теории управления финансовыми системами, функционирующими в условиях существенной неопределенности. Большое содействие этой теории оказывает теория нечетких множеств, заложенная около полувека назад в фундаментальных работах Лотфи Заде. В случае применения нечетких чисел к прогнозу параметров от ЛПР требуется не формировать точечные вероятности оценки, а задавать *расчетный коридор* значений прогнозируемых параметров. Тогда ожидаемый эффект оценивается экспертом так же, как нечеткое число со своим расчетным разбросом (степенью нечеткости).

Целью настоящей работы является исследование и анализ качественно нового подхода к управлению фондовым портфелем, основанного на применении теории нечетких множеств, а также разработка реализующих данный подход алгоритмов и сравнение результатов их применения с результатами, полученными при использовании классических вероятностных методов.

Постановка задачи

Рассматривается фондовый портфель из N компонент и его ожидаемое поведение на интервале времени $[0, T]$. Каждая из компонент портфеля $i = \overline{1, \dots, N}$ характеризуется своей финансовой доходностью r_i (оцененной в точке T как относительное приращение цены актива за период).

Держатель фондового портфеля – частный вкладчик, инвестиционная компания, взаимный фонд – управляет своими инвестициями, руководствуясь определенными соображениями. С одной стороны, инвестор старается максимизировать свою доходность. С другой стороны, он фиксирует предельно допустимый риск неэффективности своих инвестиций.

Примем капитал инвестора равным 1. Задача оптимизации фондового портфеля заключается в нахождении долевого вектора ценового распределения бумаг в портфеле $x = \{x_i\} \quad i = \overline{1, N}$, максимизирующего доход инвестора при заданном уровне риска (очевидно, что $\sum_{i=1}^N x_i = 1$).

Постановленную задачу можно условно разделить на два этапа:

1. Разработка и реализация алгоритмов методов оптимизации фондового портфеля
 - а) для модели Марковица (вероятностный подход);
 - б) для детерминированной нечеткой модели на основе теории возможностей (нечетко-множественный подход).
2. Проведение сравнительного анализа методов и выявление их достоинств и недостатков для решения поставленной задачи.

Нечетко-множественный метод оптимизации фондового портфеля

Основные принципы и идея метода

Применение множественного подхода базируется на следующих положениях, где:

1. Риск портфеля - это не его волатильность, но возможность того, что ожидаемая доходность портфеля окажется ниже некоторой предустановленной плановой величины.
2. Корреляция активов в портфеле не рассматривается и не учитывается.
3. Доходность каждого актива – это неслучайное нечеткое число (например, треугольного вида или интервального вида). Аналогично, ограничение на предельно низкий уровень доходности может быть как обычным скалярным, так и нечетким числом произвольного вида. Таким образом, мы сводим два источника информации (средняя доходность и волатильность актива) в один (расчетный коридор доходности или цены) и тем самым объединяем два источника неопределенности в один.
4. Поэтому оптимизировать портфель в такой постановке может означать, в частном случае, требование максимизировать ожидаемую доходность портфеля в точке времени T при фиксированном уровне риска портфеля (по аналогии с тем, как это делается в [4] и [5]). Эффективная граница портфельного множества в этом случае - вогнутая линия в координатах «риск недопустимо низкой доходности портфеля - ожидаемая доходность портфеля». Каждой точке эффективной границы отвечает оптимальный портфель с четкими границами. Рассмотрим задачу на основе изложенной модели, в предположении самых широких допущений к виду её нечетких параметров.

Модель оптимизации портфеля на основе теории возможностей

Нечеткое число как модель доходности актива

Пусть имеется фондовый портфель из N активов на интервале $[0, T]$. Прогнозный перформанс каждой из компонент портфеля $i = 1, \dots, N$ на момент T характеризуется своей финальной расчетной доходностью r_i (оцененной в точке T как относительное приращение цены актива за период). Поскольку доход по ценной бумаге (ЦБ) случаен, его точное значение в будущем неизвестно, а вероятностное описание такого сорта случайности не вполне корректно, то в качестве описания доходности целесообразно использовать треугольные нечеткие числа, моделируя экспертное высказывание следующего вида: «Доходность ЦБ по завершении срока владения *ожидаемо равна* \bar{r} и находится в расчетном диапазоне $[r_1; r_2]$ ».

Таким образом, для i -ой ценной бумаги имеем:

\bar{r}_i – ожидаемая доходность по i -ой ценной бумаге; r_{i1} – нижняя граница доходности i -ой ценной бумаги; r_{i2} – верхняя граница доходности i -ой ценной бумаги.

$r_i = (r_{i1}, \bar{r}_i, r_{i2})$ – доходность по i -ой ценной бумаге, треугольное нечеткое число.

Тогда доходность по портфелю:

$$r = (r_{\min} = \sum_{i=1}^N x_i r_{i1}; \bar{r} = \sum_{i=1}^N x_i \bar{r}_i; r_{\max} = \sum_{i=1}^N x_i r_{i2}) \quad (9)$$

также является треугольным нечетким числом, где x_i – вес i -го актива в портфеле, причем

$$\sum_{i=1}^N x_i = 1, \quad 0 \leq x_i \leq 1 \quad (10)$$

Также определимся с критическим уровнем доходности портфеля на момент T . Это может быть нечеткое число треугольного вида $r^* = (r_1^*; \bar{r}^*; r_2^*)$. В вырожденном случае это обычный числовой норматив r^* , например, 10% годовых.

Оценка риска портфельных инвестиций

Перейдем к оценке собственно риска портфельных инвестиций. На рис.1 представлены функции принадлежности r и критериального значения r^* .

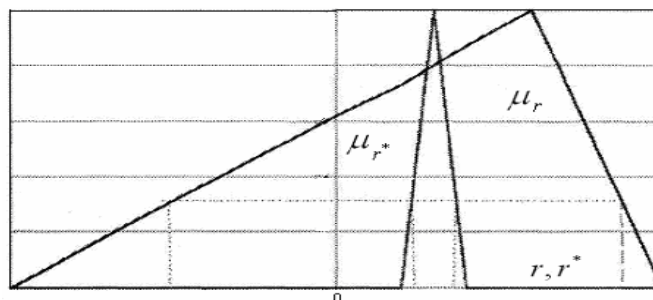


Рис.1 – Функции принадлежности r и r^*

Точкой пересечения этих двух функций принадлежности является точка с ординатой α_1 . Выберем произвольный уровень принадлежности α и определим соответствующие интервалы $[r_1, r_2]$ и $[r_1^*, r_2^*]$.

При $\alpha > \alpha_1$, $r_1 > r_2^*$, интервалы не пересекаются, и уверенность в том, что портфель эффективен, стопроцентная, поэтому степень риска неэффективности равна нулю. Уровень α_1 уместно назвать верхней границей зоны риска. При $0 \leq \alpha \leq \alpha_1$ интервалы пересекаются. На рис.2 показана заштрихованная зона неэффективного распределения активов в портфеле, ограниченная прямыми $r^* = r_1^*$, $r^* = r_2^*$, $r = r_1$, $r = r_2$ и биссектрисой координатного угла $r = r^*$.

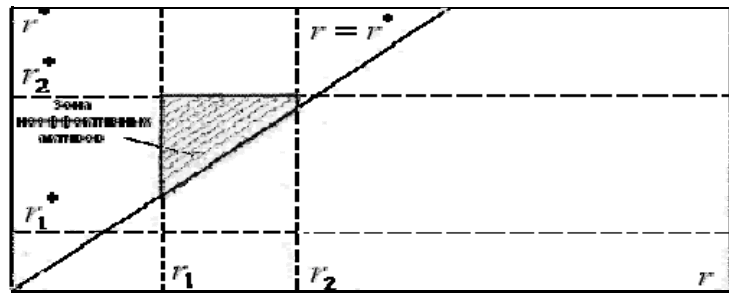


Рис.2 - Фазовое пространство (r, r^*)

Взаимные соотношения параметров $r_{1,2}^*$ и $r_{1,2}$ дают следующий расчет для площади заштрихованной плоской фигуры [2, 3]:

$$S_\alpha = \begin{cases} 0, & \text{при } r_1 \geq r_2^* \\ \frac{(r_2^* - r_1)^2}{2}, & \text{при } r_2^* > r_1 \geq r_1^*; r_2 \geq r_2^* \\ \frac{(r_1^* - r_1) + (r_2^* - r_1)}{2} \cdot (r_2^* - r_1^*), & \text{при } r_1 < r_1^*, r_2 > r_2^* \\ (r_2^* - r_1^*)(r_2 - r_1) - \frac{(r_2 - r_1)^2}{2}, & \text{при } r_1 < r_1^* \leq r_2; r_2 < r_2^* \\ (r_2^* - r_1^*)(r_2 - r_1), & \text{при } r_2 \geq r_1^* \end{cases} \quad (11)$$

Поскольку все реализации (r, r^*) при заданном уровне принадлежности α равновозможны, то степень риска, неэффективности $\varphi(\alpha)$ есть геометрическая вероятность события попадания точки (r, r^*) в зону неэффективного распределения капитала:

$$\varphi(\alpha) = \frac{S_\alpha}{(r_2^* - r_1^*) \cdot (r_2 - r_1)} \quad (12)$$

где S_α оценивается по формуле (12).

Тогда итоговое значение степени риска неэффективности портфеля:

$$\beta = \int_0^{\alpha_1} \varphi(\alpha) d\alpha \quad (13)$$

В важном частном случае (см. рис. 3), когда критерий эффективности определен четко уровнем r^* , то предельный переход при $r_2^* \rightarrow r_1^* \rightarrow r^*$ дает:

$$\varphi(\alpha) = \begin{cases} 0, & \text{при } r^* < r_1 \\ \frac{(r^* - r_1)}{(r_2 - r_1)}, & \text{при } r_1 \leq r^* \leq r_2; \alpha \in [0;1] \\ 1, & \text{при } r^* > r_2 \end{cases} \quad (14)$$

Для того чтобы собрать все необходимые исходные данные для оценки риска, нам потребуется два значения обратной функции $\mu_r^{-1}(\alpha_1)$. Первое значение есть r^* (по определению верхней границы зоны риска α_1), второе значение обозначим \tilde{r}^* . Аналогичным образом обозначим r_{\min} и r_{\max} – два значения обратной функции $\mu_r^{-1}(0)$.

Введем обозначение \tilde{r} - наиболее ожидаемое значение r . Тогда выражение для степени риска портфеля β , с учетом (14) и длинной цепи преобразований, имеет вид [3]:

$$\beta = \begin{cases} 0, & \text{при } r^* < r_{\min} \\ R \left(1 + \frac{1 - \alpha_1}{\alpha_1} \ln(1 - a_1) \right), & \text{при } r_{\min} \leq r^* \leq \tilde{r} \\ 1 - (1 - R) \left(1 + \frac{1 - \alpha_1}{\alpha_1} \ln(1 - a_1) \right), & \text{при } \tilde{r} \leq r^* < r_{\max} \\ 1, & \text{при } r^* \geq r_{\max} \end{cases} \quad (15)$$

где

$$R = \begin{cases} \frac{r^* - r_{\min}}{r_{\max} - r_{\min}}, & \text{при } r^* < r_{\max} \\ 1, & \text{при } r^* \geq r_{\max} \end{cases} \quad (16) \quad \alpha = \begin{cases} 0, & \text{при } r^* < r_{\max} \\ \frac{r^* - r_{\min}}{\tilde{r} - r_{\min}}, & \text{при } r_{\min} \leq r^* < \tilde{r} \\ 1, & \text{при } r^* = \tilde{r} \\ \frac{r_{\max} - r^*}{r_{\max} - \tilde{r}}, & \text{при } \tilde{r} < r^* < r_{\max} \\ 0, & \text{при } r^* \geq r_{\max} \end{cases} \quad (17)$$

Таким образом, степень риска β принимает значения от 0 до 1. Каждый инвестор, исходя из своих инвестиционных предпочтений, может классифицировать значения β , выделив для себя отрезок неприемлемых значений риска.

Модель управления доходностью портфеля

Для того, чтобы определить структуру портфеля, который обеспечит максимальную доходность при заданном уровне риска, требуется решить следующую задачу:

$$\{x_{opt}\} = \{x\} \mid r \rightarrow \max, \beta = const \quad (18)$$

где r и β определяются из формул (15)-(17), компоненты вектора x удовлетворяют (10).

Учитывая, что доходность портфеля:

$$r = (r_{\min} = \sum_{i=1}^N x_i r_{1i}; \tilde{r} = \sum_{i=1}^N x_i \tilde{r}_i; r_{\max} = \sum_{i=1}^N x_i r_{2i}) \quad (19)$$

где $(r_{1i}, \tilde{r}_i, r_{2i})$ – доходность i -ой ценной бумаги, получаем следующую задачу оптимизации (20)-(22):

$$\tilde{r} = \sum_{i=1}^N x_i \tilde{r}_i \rightarrow \max \quad (20)$$

$$\beta = const \quad (21)$$

$$\sum_{i=1}^N x_i = 1 \quad x_i \geq 0 \quad i = \overline{1, N} \quad (22)$$

При варьировании уровня риска β возможны 2 случая. Рассмотрим подробно каждый из них.

1. $\beta = 0$. Из (15) видно, что этот случай возможен когда $r^* < \sum_{i=1}^N x_i r_{i1}$.

Получаем следующую задачу линейного программирования (23)-(25):

$$\tilde{r} = \sum_{i=1}^N x_i \tilde{r}_i \rightarrow \max \quad (23)$$

$$\sum_{i=1}^N x_i r_{i1} > r^* \quad (24)$$

$$\sum_{i=1}^N x_i = 1 \quad x_i \geq 0 \quad i = \overline{1, N} \quad (25)$$

Найденный в результате решения задачи (23)-(25) вектор $x = \{x_i\} \quad i = \overline{1, N}$ и есть искомая структура оптимального для данного уровня риска портфеля.

2. $0 < \beta < 1$. Из (15) видно, что этот случай возможен когда $\sum_{i=1}^N x_i r_{i1} \leq r^* < \sum_{i=1}^N x_i \tilde{r}_i$, либо когда

$$\sum_{i=1}^N x_i \tilde{r}_i \leq r^* < \sum_{i=1}^N x_i r_{i2}.$$

а) Пусть $\sum_{i=1}^N x_i r_{i1} \leq r^* < \sum_{i=1}^N x_i \tilde{r}_i$. Тогда используя (15)-(17) задача (20)-(22) сводится к следующей задаче нелинейного программирования:

$$\tilde{r} = \sum_{i=1}^N x_i \tilde{r}_i \rightarrow \max \quad (26)$$

$$\frac{1}{\sum_{i=1}^N x_i r_{i2} - \sum_{i=1}^N x_i r_{i1}} \left(\left(r^* - \sum_{i=1}^N x_i r_{i1} \right) + \left(\sum_{i=1}^N x_i \tilde{r}_i - r^* \right) \cdot \ln \left(\frac{\sum_{i=1}^N x_i \tilde{r}_i - r^*}{\sum_{i=1}^N x_i \tilde{r}_i - \sum_{i=1}^N x_i r_{i1}} \right) \right) = \beta \quad (27)$$

$$\sum_{i=1}^N x_i r_{i1} \leq r^* \quad (28)$$

$$\sum_{i=1}^N x_i \tilde{r}_i > r^* \quad (29)$$

$$\sum_{i=1}^N x_i = 1 \quad x_i \geq 0 \quad i = \overline{1, N} \quad (30)$$

б) Пусть $\sum_{i=1}^N x_i \tilde{r}_i \leq r^* < \sum_{i=1}^N x_i r_{i2}$. Тогда задача (20)-(22) сводится к следующей задаче нелинейного программирования:

$$\tilde{r} = \sum_{i=1}^N x_i \tilde{r}_i \rightarrow \max \quad (31)$$

$$\frac{1}{\sum_{i=1}^N x_i r_{i2} - \sum_{i=1}^N x_i r_{i1}} \left(\left(r^* - \sum_{i=1}^N x_i r_{i1} \right) - \left(r^* - \sum_{i=1}^N x_i \tilde{r}_i \right) \cdot \ln \left(\frac{r^* - \sum_{i=1}^N x_i \tilde{r}_i}{\sum_{i=1}^N x_i r_{i2} - \sum_{i=1}^N x_i r_{i1}} \right) \right) = \beta \quad (32)$$

$$\sum_{i=1}^N x_i r_{i2} > r^* \quad (33)$$

$$\sum_{i=1}^N x_i \tilde{r}_i \leq r^* \quad (34)$$

$$\sum_{i=1}^N x_i = 1 \quad x_i \geq 0 \quad i = \overline{1, N} \quad (35)$$

Для решения задач (26)-(30) и (31)-(35) применен R-алгоритм минимизации недифференцируемых функций [9]. Пусть обе задачи: (26)-(30) и (31)-(35) разрешимы. Тогда структуре искомого оптимального портфеля будет отвечать вектор $x = \{x_i\} \quad i = \overline{1, N}$ – решение той из задач (26)-(30), (31)-(35)), значение целевой функции которой будет больше.

Анализ и сравнение результатов, полученных при применении моделей Марковитца и нечетко-множественного метода

Для сравнительного анализа исследуемых методов оптимизации фондового портфеля были использованы реальные данные по курсу акций компании ПАО «ЕЭС России» (EERS2) и ОАО Газпром (GASP), взятые за период с июля 2003 года по май 2006 года. Данные получены из архивов Московской фондовой биржи (МФБ).

Пусть, критическая доходность портфеля составляет 3,5% т.е. портфельные инвестиции, приносящие доход ниже 3% считаются неэффективными.

Нечетко-множественный метод дал следующие результаты (см. таблицу 1):

Таблица 1

ERS2	GASP	Доходность портфеля	Нижняя граница	Верхняя граница	Уровень риска
0	1	4,8	-4,1	5,7	0,5203
0,1	0,9	4,53	-3,59	5,52	0,5449
0,2	0,8	4,26	-3,08	5,34	0,5767
0,3	0,7	3,99	-2,74	5,16	0,6207
0,4	0,6	3,72	-2,06	4,98	0,6878
0,5	0,5	3,45	-1,55	4,9	0,8212
0,6	0,4	3,18	-1,04	4,62	0,8871
0,7	0,3	2,91	-0,53	4,44	0,9239
0,8	0,2	2,64	-0,02	4,26	0,949
0,9	0,1	2,37	0,48	4,08	0,9688
1	0	2,1	0,99	3,9	0,9833

Результаты, полученных с помощью модели Марковитца для аналогичных уровней риска таковы (Табл.2):

Таблица 2

ERS2	GASP	Доходность портфеля	Уровень риска
0,5	0,49	5.01	0,5
0,38	0,62	5.31	0,6
0,26	0,74	5.6	0,7
0,02	0,98	6.23	0,9

В нечетко-множественном методе, доходность каждого актива – это детерминированное нечеткое число. Её ожидаемое значение рассчитывается уже не из статистических данных, а исходя из состояния рынка в момент принятия инвестором решения. Таким образом, в рассматриваемом случае, ожидаемая доходность портфеля не слишком высока.

Структура оптимального портфеля, полученная в результате применения методов, для одних и тех же уровней риска тоже различна. Для того, что бы выяснить причину этого, рассмотрим следующие зависимости (рис.3).

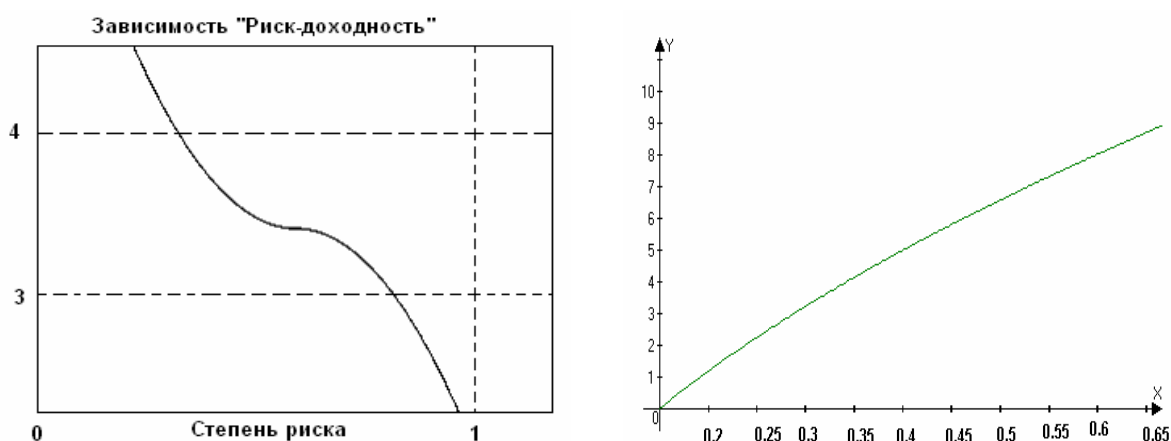


Рис. 3. Зависимость ожидаемой доходности от степени риска портфеля, полученного
а) нечетко-множественным методом. б) с помощью Марковитца

Зависимости ожидаемой доходности от степени риска портфеля, полученного указанными выше методами, практически противоположны. Причиной такого результата является различное понимание уровня риска портфеля.

В нечетко-множественном методе под риском понимается ситуация, когда ожидаемая доходность портфеля ниже заданного критического уровня. Со снижением ожидаемой доходности возрастает риск того, что доход от портфельных инвестиций окажется меньше критического значения.

В модели Марковича риск рассматривается как степень колеблемости ожидаемого дохода по портфелю, причем как в меньшую, так и в большую сторону, что противоречит здравому смыслу.

Различное понимание уровня риска портфеля является также причиной различия зависимостей степени риска от доли той или иной акции в портфеле, полученном разными методами приведенная на рис.4.

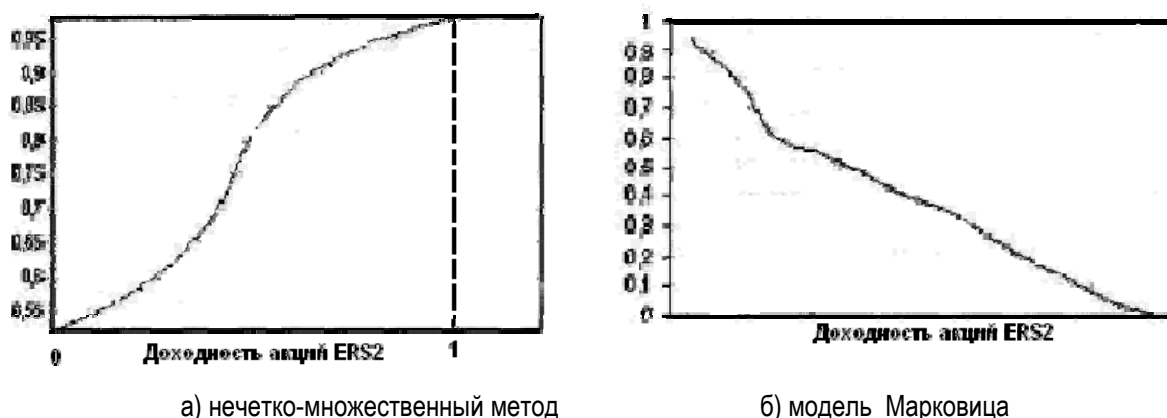


Рис. 4. Зависимость степени риска портфеля от доли акций ESP2

Вполне ясно, что с ростом доли низкодоходной бумаги в портфеле, даже несмотря на то, что расчетный коридор по EERS2 более узок, нежели расчетный коридор по GASP, падает ожидаемая доходность портфеля в целом - и, соответственно, растет риск неэффективности портфельного выбора.

Уровень же изменчивости ожидаемых доходов для акций EERS2 исходя из данных 2000-2006 гг. намного ниже, чем для акций GASP. Поэтому в модели Марковица, рассматривающей его как риск портфельных инвестиций, с увеличением доли акций EERS2 риск портфеля снижается.

Выводы

В данной работе проводилось исследование в области фондового менеджмента. Была рассмотрена относительно недавно возникший нечетко-множественный подход к портфельной оптимизации. В результате проведенного исследования была получена основанная на нечетко-множественном подходе математическая модель для нахождения структуры оптимального инвестиционного портфеля, лишенная большинства недостатков классических вероятностных моделей.

На основании теории нечетких множеств был разработан алгоритм оптимизации фондового портфеля.

В процессе исследования и сравнительного анализа модели Марковитца и нечетко-множественного метода определения оптимальной структуры фондового портфеля было выявлено следующее:

1. Структуры оптимального портфеля и показатели его ожидаемой доходности, получаемые с помощью модели Марковитца и нечетко-множественного метода кардинально отличаются.
2. С уменьшением объема выборки исходных данных по доходности активов модель Марковитца даёт более правдоподобные результаты. Однако слишком маленькая выборка также не допустима, т.к. не может дать полного представления рассматриваемых параметров.
3. Поскольку отклонение ожидаемой доходности в большую сторону, также как и в меньшую, рассматривается в модели Марковитца как риск, зависимости ожидаемой доходности от уровня риска портфеля, полученные с помощью упомянутой модели Марковитца и нечетко-множественного метода, практически противоположны.
4. По указанной выше причине также довольно часто доля высокодоходных активов в структуре портфеля, полученного с помощью модели Марковитца, неоправданно мала.

Литература

1. Недосекин А.О. Методологические основы моделирования финансовой деятельности с использованием нечетко-множественных описаний // Диссертация на соискание уч. ст. докт. экон. наук. СПб., 2003.

2. Система оптимизации фондового портфеля (Сименс Бизнес Сервисез Россия). – На сайте <http://www.sbs.ru/index.asp?objectID=1863&lang=rus>
3. Недосекин А.О. Система оптимизации фондового портфеля от Сименс Бизнес Сервисез // Банковские технологии. – 2003. – № 5. – Также на сайте: <http://www.finansy.ru/publ/fin/004.htm>
4. Недосекин А.О. Применение теории нечетких множеств к задачам управления финансами. Раздел 3 // Аудит и финансовый анализ. – 2000. – №2. - Также на сайте: <http://www.cfin.ru/press/afa/2000-2/08-3.shtml>
5. Недосекин А.О. Оптимизация бизнес-портфеля корпорации. - На сайте: http://sedok.narod.ru/s_files/2003/Art_070303.doc
6. International Conference on Fuzzy Sets and Soft Computing in Economics and Finance FSSCEF 2004 - На сайте http://www.ifel.ra/fsscef2006/2004/FSSCEF_I.pdf
7. Виленский П.Л., Лившиц В.Н., Орлова Е.Р., Смоляк С.А. Оценка эффективности инвестиционных проектов. М.: Дело, 1998.
8. Смоляк С. А. Учет специфики инвестиционных проектов при оценке их эффективности // Аудит и финансовый анализ. – 1999. – №3.
9. Шор Н.З. Задачі оптимального проектування надійних мереж // - К.: „Наукова думка”. – 2004р.

Информация об авторах

Зайченко Юрий Петрович, профессор, д.т.н., кафедра «Институт прикладного системного анализа». Киев, НТУУ «КПИ», ул. Политехническая 14. тел: +8(044)241-86-93, e-mail: zaych@i.com.ua

Малихех Есфандиярфард (Иран), аспирантка кафедры «Прикладная математика»; тел: +38(096)6915857, e-mail: fard_sem@yahoo.com

МОДЕЛЬ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ УПРАВЛЕНИЯ НЕНАДЕЖНЫМИ ЭЛЕМЕНТАМИ (ЛЮДЬМИ)

Тимофей Рябцев, Елена Антонова

Аннотация: В статье предложен возможный подход к созданию интеллектуальной системы управления людьми и коллективом при решении ими профессиональных задач, в которой бы учитывались личные особенности и психологическое состояние человека как «ненадежного» элемента. В статье описаны некоторые элементы модели интеллектуальной системы управления: модель профессиональной деятельности и модель «ненадежного» элемента – человека.

Keywords: *The theory of unreliable elements, The knowledge system, The intelligent control, KDS.*

Введение

Управление в широком понимании этого термина – непрерывный процесс воздействия на объект управления (личность, коллектив, технологический процесс, предприятие, государство) для достижения оптимальных результатов при наименьших затратах времени и ресурсов.

Управление существует более 7 тысяч лет, оно социально, так как началось с развития человеческих отношений. Тенденции развития управления – методы, отношение к объекту управления – в течение всего времени были направлены на улучшение социальной среды работников и более тонкое взаимодействие с каждым человеком с целью учета его индивидуальных особенностей и потребностей. Это вполне естественно, поскольку без людей нет организации. Без нужных людей ни одна организация не сможет достичь своих целей и выжить. А это крайне тяжело без эффективного управления трудовыми

ресурсами – одного из важнейших аспектов теории и практики управления. Для того чтобы предприятие работало эффективно, необходимо правильно организовать труд работников, при этом постоянно контролируя деятельность работников, используя различные методы управления.

Специфический вид управленческой деятельности, объектом которой является коллектив работников - персонал, получил название управления персоналом (кадрами). Специфический, потому что в управлении людьми необходимо учитывать такую неопределенную, ненадежную вещь, как «человеческий фактор».

Ненадежные элементы в контексте управления – это элементы, о состоянии которых нельзя получить достоверную информацию и поведение которых невозможно спрогнозировать с высокой степенью вероятности. В теории управления ненадежными элементами, прежде всего, являются люди.

Достижения новой информационной технологии сегодня уже позволяют рассматривать возможность автоматизации процессов управления людьми при решении ими профессиональных задач, являясь одним из основных резервов повышения эффективности управления. Десятилетия развития локальных сетей и сети Интернет предложили множество технологий для организации общения, как между людьми, так и между целыми предприятиями и программными комплексами, которые также могут быть использованы в контексте управления. Анализ существующих программных систем позволяет сделать вывод о том, что автоматизацию процесса управления понимают, прежде всего, как информационное обеспечение лиц, принимающих решения.

Для решения вышеописанных проблем предполагается разработка системы управления деятельностью коллективов. Основа системы будет состоять из модели интеллектуального управления людьми. Главной особенностью системы будет то, что она сама будет учитывать психологические особенности людей (человеческий фактор) в коллективе, что позволит учесть многие ситуации в работе с людьми. Зная психологию человека и особенности конкретного работника, можно влиять на мотивы его деятельности, т.е. управлять им с целью получения заданного результата.

На концептуальном уровне проектирования системы управления ненадежными элементами необходимо четкое определение общей аксиоматики описания процессов управления, т.е. фиксация исходных положений дальнейшего исследования, описание модели системы управления.

Модель системы управления – совокупность представлений о системе. Необходимо определиться с основными структурными элементами модели управления. В совокупности эти элементы должны образовывать полный и непротиворечивый набор моделей, характеризующий все базовые компоненты системы управления ненадежными элементами.

Применительно к задаче управления ненадежными элементами (людьми) при решении ими профессиональных задач набор элементов модели системы управления включает:

- 1) модель профессиональной деятельности, определяющую цели и задачи деятельности, план решения, ограничения и взаимосвязь ресурсных потоков;
- 2) модель «ненадежного элемента» (человека) как объекта и как субъекта управления, определяющая ролевые взаимоотношения в коллективе, характеристики ненадежных элементов;
- 3) модель управленческого процесса.

В данной статье рассмотрим модель профессиональной деятельности и модель «ненадежного» элемента (человека).

Модель профессиональной деятельности (МПД)

Цель разработки модели профессиональной деятельности: формализация способа описания и планирования деятельности, описание множества состояний выполняемой деятельности и взаимосвязей ее элементов, формализация алгоритма оценки эффективности деятельности на всех этапах по заданным критериям. На основе МПД формируется система агрегированного представления

планируемых бизнес-процессов, описание взаимосвязи параметров ресурсных потоков, а также представление отслеживаемых параметров выполнения деятельности для возможного вмешательства системы с целью обеспечения выполнения плана решения «ненадежными элементами».

Профессиональная деятельность - это последовательность имеющих начало и конец действий по решению задач, преследующая определенные цели и использующая соответствующие ресурсы.

Процесс построения модели конкретной профессиональной деятельности можно разбить на следующие процедуры:

Процедура 1. Формулирование профессиональной деятельности (ПД).

Процедура 2. Планирование ПД.

Процедура 3. Выполнение ПД.

При определении модели профессиональной деятельности необходимо учитывать следующие определяющие элементы этой ПД:

1. Цель профессиональной деятельности - конечный результат, выход, продукция, определяемая в терминах затрат, качества и времени реализации.
2. Сложность. Для достижения целей профессиональной деятельности необходимо решить множество подзадач. Отношения между подзадачами могут быть довольно сложными.
3. Ограниченность во времени. Подзадачи имеют начало и конец. Для их реализации необходима временная концентрация ресурсов. Если будет необходимо, то ресурсы используются на другие цели.
4. Жизненный цикл. По мере реализации подзадач, изменяется потребность в тех или иных ресурсах. Это изменение идет в определенной предсказуемой последовательности.

МПД системы управления должна обеспечить достижение целей проекта путем выполнения следующих мероприятий:

1) На этапе формулирования ПД:

1.1 Планирование профессиональной деятельности с определением подзадач.

1.2 Определение необходимых ресурсов.

2) На этапе планирования ПД:

2.1. Упорядочивание уровня сложности между подзадачами;

2.2. Распределение ресурсов между подзадачами.

3) На этапе выполнения ПД:

3.1 Выделение параметров контроля выполнения каждого этапа.

Приведем примеры элементов модели профессиональной деятельности:

Объект: Профессиональная деятельность

Описание: формальное общее представление среды выполняемой деятельности, в рамках которой будет осуществляться процесс управления.

Атрибуты: (критические точки: *результаты, ресурсы, сроки*)

Объект: Подзадачи

Описание: определяемая субъектом управления минимальная единица деятельности, являющаяся частью всей профессиональной деятельности.

Атрибуты: (критические точки: *результаты, ресурсы, сроки; должностная характеристика*)

Функция связи: Стратегия

Описание: представляет профессиональную деятельность в виде иерархии минимальных единиц (подзадач); результатом воздействия функции является иерархия подзадач в системе, представляющая отношения между ними. (Является рекурсивной: разбивает подзадачи до самого возможного низкого уровня иерархии).

Определение: *Стратегия (профессиональная деятельность) → {подзадачи} U {связи}*

Соглашение: Профессиональная деятельность == \sum {подзадачи}.

Функция связи: Цель

Описание: определяет результаты всей профессиональной деятельности, а также результаты каждой подзадачи; результатом воздействия функции является множество выходных продуктов деятельности, подтвержденных определенным документом.

Определение: *Цель (профессиональная деятельность) → {Результаты}*

Цель (подзадачи) → {Результаты}

Соглашение: {Профессиональная деятельность.Результаты} == \sum {Подзадачи.Результаты}.

Атрибут: Результаты

Описание: конкретное выражение итогов (определенного качества), достигаемых в результате деятельности.

ОДЗ: (документы, подтверждающие наличие продуктов деятельности)

Модель профессиональной деятельности тесно связана с моделью человека и его особенностями.

Модель человека - «ненадежного элемента» (МОЧ)

Существующие программные системы управления, элементами которых являются люди, обладают рядом особенностей, которые затрудняют процесс управления. Это напрямую связано с биологической, эмоциональной особенностью людей, что и делает их «ненадежными элементами». Ненадежность человека как объекта/субъекта управления отрицательно влияет на ход выполнения профессиональной деятельности.

Интеллектуальная система должна иметь знания о сотрудниках, содержащие:

- 1) Информацию о каждом сотруднике: тип поведения, его должностные обязанности – для определения методов мотивационного воздействия и определения видов поручаемых задач;
- 2) Знания об общих видах ненадежности людей при решении профессиональных задач – для применения знания об оправданных в данном случае методах мотивационного воздействия из «Банка Знаний системы о ненадежности и мотивации»;
- 3) Ответственность, определяемую для каждого сотрудника в отношении подзадач – для закрепления людского ресурса за подзадачами ПД;
- 4) Историю поведения: информацию о деятельности и реакции каждого сотрудника на виды мотивации – для ведения общедоступной статистики, как мотивационного инструмента повышающего активность сотрудников, представления отчетности для руководителей, а также определения наиболее успешных методов мотивации.

В модели человека выделим следующие понятия:

- 1) Субъект управления, который осуществляет управленческую и организационную работу, принимает решения и обеспечивает достижение поставленных целей. Субъект управления называют управляющей системой.

2) Объект управления, на который направлено управленческое воздействие в целях обеспечения функционирования и развития системы. Объект управления иногда называют управляемой системой.

Управление осуществляется некоторым субъектом по отношению к некоторому объекту (или объектам). Оправдано рассмотрение субъектов и объектов управления в иерархии управления, поскольку на разных ступенях иерархии один и тот же объект управления может быть как субъектом, так и объектом. Процессу управления всегда должно предшествовать определение субъекта и объекта управления в заданной иерархии отношений между ними, а также определение свойств и характеристик, отрицательно влияющих на выполнение профессиональной деятельности людьми.

Для разработки модели необходима формализация особенностей управляющей и управляемой систем (субъекта и объекта управления), определение возможных «блокирующих» воздействий (тормозящих осуществление деятельности по заданному плану решения), определение возможных и оправданных способов мотивации для предотвращения «блокирующих» воздействий данного «ненадежного элемента» [Ryabtsev, 2006].

На основе МОЧ формируется система представления «ненадежного элемента» (человека) как субъекта и объекта управления.

Приведем примеры элементов модели человека:

Объект: Коллектив

Описание: группа совместно работающих при решении профессиональных задач лиц, которые взаимодействуют друг с другом таким образом, что каждый может оказывать влияние (управлять) на другого, и одновременно находятся в зависимости друг от друга; то есть представляется иерархия субъектов и объектов управления.

Атрибуты: (ресурсы: люди)

Объект: Объект (управляемая система)

Описание: объект, на который направлено управленческое воздействие в целях обеспечения функционирования системы управления; осуществляет выполнение задач и обеспечивает достижение поставленных целей.

Элементы: (история поведения, ненадежность, тип поведения, должностная инструкция)

Объект: Субъект (управляющая система)

Описание: осуществляет планирование задач, распределение ресурсов, принимает решения и отслеживает достижение поставленных целей.

Элементы: (ненадежность, история поведения, тип поведения, должностная инструкция)

Элемент: Рольевые отношения

Описание: отношения между объектами и субъектами модели управления.

ОДЗ (область допустимых значений): (управляет, управляем)

Функция связи: Ответственность

Описание: признак отношения объектов и выполняемых задач.

ОДЗ: (есть, нет)

Определение: *Ответственность: {{задача}, {объект}} → {{есть, нет}}*

Элемент: Тип поведения

Описание: набор индивидуальных психологических особенностей объекта, представляющих его как личность.

ОДЗ: (высокая ответственность, нормальная ответственность, низкая ответственность)

Элемент: История поведения

Описание: накапливаемые характеристики объекта за жизненный цикл деятельности коллектива: успешность выполнения задач и реакции на мотивацию.

ОДЗ: {{успешно, неуспешно}} U {{успешно*, неуспешно*}}

Функция связи: Должностные обязанности*

Описание: определяет для объекта\субъекта требования должностных инструкций, необходимые для осуществления профессиональной деятельности по подзадам.

Определение:

Должностные обязанности (объект) → {должностная характеристика*}*

Должностные обязанности (субъект) → {должностная характеристика*}*

Все описанные объекты, атрибуты и их связи должны быть представлены в интеллектуальной системе управления. Архитектура разрабатываемой системы представлена на Рис 1.

Рассмотрим элементы архитектуры:

Субъект (объект) – люди (ненадежные элементы)

Инженер знаний – наполняет базу данных (задачи и люди), базу знаний методов управления и базу знаний процедур управления.

Подсистема субъекта управления – подсистема взаимодействия между субъектом управления и управляющей системой. Передает данные в базу данных о задачах и людях. Передает управление управляющей подсистеме. Получает отчет о результате выполнения задачи объектом.

Подсистема объекта управления – подсистема взаимодействия между субъектом управления и управляющей системой. Передает данные в базу данных о задачах и людях. Передает управление управляющей подсистеме. Получает действия необходимые для выполнения задачи.

Управляющая подсистема – система-посредник между подсистемами объекта и субъекта и подсистемами применения методов и процедур. Передает данные о задаче для подсистемы определения метода управления и получает данные о результате выполнения процедуры подсистемой применения процедур управления.

Подсистема определения метода управления - анализирует и выбирает методы наиболее эффективные и передает управление и данные – выбранный метод решения задачи подсистеме применения процедур управления. Получает исходные данные о решении задачи от управляющей системы, а также данные из базы о задачах и людях.

Подсистема редактирования знаний используется инженером знаний для наполнения базы данных о задачах и людях, наполнения базы знаний методов управления и базы знаний процедур управления.

База данных (задачи и люди) содержит информацию о задачах, разбитых на подзадачи. Также содержит информацию о сроках выполнения, критических точках, и людях, с их функциями и особенностями.

База знаний методов управления содержит описание известных методов управления для решения задачи по известному плану решения.

Использование всех компонентов интеллектуальной системы управления, позволит добиться такой работы системы, при которой будет снижаться «ненадежность» людей при решении профессиональных задач.

Главной особенностью системы управления будет то, что она сама будет учитывать психологические особенности людей (человеческий фактор) в коллективе, что позволит учесть многие ситуации в работе с людьми.

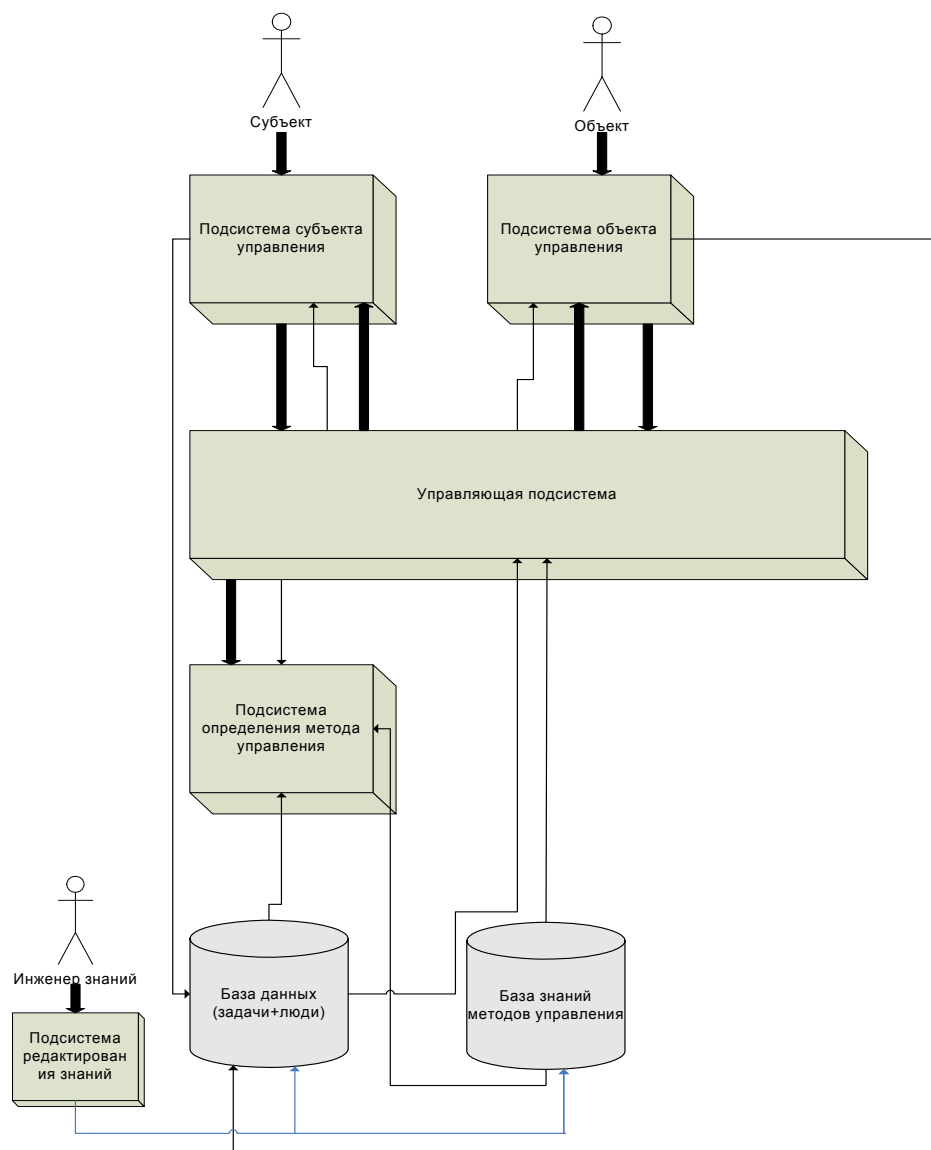


Рис. 1. Архитектура системы управления

Заключение

В статье предложен возможный подход к созданию интеллектуальной системы управления, в которой бы учитывались личные особенности и психологическое состояние человека как «ненадежного» элемента. В статье описаны составляющие модели интеллектуальной системы управления: модель профессиональной деятельности и модель «ненадежного» элемента – человека.

Существующие на сегодняшний день программные системы управления не в полной мере охватывают все характеристики и свойства человека как объекта или субъекта управления. Поэтому актуальной и практически значимой остается проблема создания такой интеллектуальной системы, в которой бы учитывались психологические особенности человека как личности при принятии управленческих решений.

Следует заметить, что предлагаемый в статье подход к решению задач интеллектуализации управления людьми с учетом их психологических особенностей и личностных характеристик может быть применен в тех предметных областях, где зависимость системы управления от особенностей человека значительна.

Библиография

[Ryabtsev, 2006] T.V. Ryabtsev, E.I. Antonova, R.V. Benger, "The Model of Intelligent system of activity people control in solving professional problems". In: "Artificial Intelligence 2006" magazine, Institute of Artificial Intelligence, Ukraine, Donetsk, 2006.

Информация об авторах

Timofey Ryabtsev, Elena Antonova – Far East State University, Institute of Mathematics and Computer Sciences, P.O. Box: 690003, Verhne-portovaya str. 64a - 58, Russia, Vladivostok; e-mail: nitro@farpost.com

ПОДХОД К РАЗРАБОТКЕ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ УПРАВЛЕНИЯ НЕНАДЕЖНЫМИ ЭЛЕМЕНТАМИ. ПРОЦЕДУРЫ УПРАВЛЕНИЯ.

Роман Бенгер, Елена Антонова

Аннотация: В статье представлен подход к проектированию системы интеллектуального управления людьми для повышения их надежности на основе применения интеллектуальных процедур управления [Benger, 2006]. Приведены результаты разрабатываемого подхода к проектированию интеллектуальной системы управления и базы знаний процедур управления.

Основным объектом модели системы служат процедуры управления, взаимодействие и деятельность которых осуществляет определенные виды воздействий на ненадежные элементы (людей) для достижения глобальной цели – повышения надежности людей при решении ими профессиональных задач. Предлагаемый в статье подход к разработке процедур управления позволит отслеживать эффективность применения управленческих процедур и соответствующие реакции людей на применяемые управленческие воздействия. Знания об эффективности и реакциях людей в дальнейшем позволят самообучаться системе, тем самым, улучшая возможности применения системы при дальнейшей ее эксплуатации.

Keywords: Management of people, Knowledge-based systems, KDS.

ACM Classification Keywords: H.1 Models and principles – General.

Введение

Управленческий процесс рассматривается как управление деятельностью по задаче: планирование и выполнение работ, координация действий специалистов и организаций, участвующих в решении задачи. Управленческий процесс можно разбить на управленческие процедуры, которые относятся к планированию, контролю выполнения задач и воздействию на исполнителей.

Интеллектуальная система управления на основе управленческих процедур должна обеспечивать эффективную реализацию плана решения задач профессиональной деятельности путем отслеживания критических точек, учета ненадежности сотрудников и применения целенаправленного воздействия на субъекты и объекты управления («ненадежные элементы») с целью обеспечения выполнения ими плана профессиональной деятельности.

Модель интеллектуальной системы управления

Цель разработки модели интеллектуальной системы управления - формализация способа описания и планирования деятельности, описания множества состояний выполняемой деятельности и взаимосвязей ее элементов, формализация алгоритма оценки эффективности деятельности на всех этапах по заданным критериям. Необходимо сформировать систему агрегированного представления планируемых бизнес-процессов, описание взаимосвязи параметров ресурсных потоков, а также представление отслеживаемых параметров выполнения деятельности для возможного вмешательства системы с целью обеспечения выполнения плана решения «ненадежными элементами».

В рассматриваемой модели интеллектуальной системы ресурсами являются люди. Поэтому необходимо учесть в модели «ненадежность» человека: формализовать особенности человека как субъекта и объекта управления, определить возможные «блокирующие» воздействия (тормозящие осуществление деятельности по заданному плану решения), определить возможные и оправданные способы мотивации для предотвращения этих «блокирующих» воздействий. При этом необходимо учитывать особенности конкретного «ненадежного элемента».

Закладываемые в систему знания об управлении «ненадежными элементами» (процедуры управления) должны быть представлены в виде наполняемой базы знаний. Используя эти знания, система будет принимать интеллектуальные управленческие решения. Таким образом, необходима формализация мероприятий по достижению показателей эффективности, а также формализация применения приемов защиты от «блокирующих» воздействий «ненадежных элементов» и применения способов мотивации. Необходимо определение алгоритма, позволяющего вмешаться в ход выполняемой деятельности и применить необходимые для данной ситуации и для данного «ненадежного элемента» управляющие процедуры, чтобы обеспечить выполнение плана решения задач профессиональной деятельности.

Таким образом, модель интеллектуальной системы управления характеризуется воздействием (использованием конкретного управленческого мероприятия – набора процедур управления) на функционирование «ненадежного элемента» и, в связи с этим, необходимостью пересмотра бизнес-процессов, описывающих план решения.

Приведем графическое представление части модели системы управления, содержащее основные понятия описываемой предметной области (Рис. 1.):

Данная модель описывает подход к проектированию системы управления. Определено понятие профессиональная деятельность как последовательность имеющих начало и конец бизнес-процессов, преследующая определенные цели и использующая соответствующие ресурсы («ненадежные элементы»).

При разработке подхода к описанию бизнес-процессов обеспечивается возможность вмешаться в проект в случае возможности его провала, чтобы скорректировать его, привести в соответствие с планом по заданным критериям (сроки, ресурсы, результаты). Для этого обеспечивается возможность представления текущих результатов работы для их оценивания и сравнения с планом профессиональной деятельности, то есть обеспечивается возможность в любой момент времени представить текущие отслеживаемые показатели проекта, по которым происходит мониторинг на предмет продвижения к целям профессиональной деятельности.

При описании элементов модели интеллектуальной системы управления описываются знания о «ненадежных элементах» (людях), содержащие:

- информацию о каждом сотруднике: тип поведения, его должностные обязанности – для определения возможности применения процедур управления, а также определения видов поручаемых задач;
- знания об общих видах ненадежности людей при решении профессиональных задач – для применения знания об оправданных в данном случае воздействиях – процедур управления;
- ответственность, определяемую для каждого сотрудника в отношении подзадач – для закрепления людского ресурса за подзадачами;
- историю поведения: информацию о деятельности и реакции каждого сотрудника на воздействия процедур управления – для ведения общедоступной статистики, как мотивационного инструмента повышающего активность сотрудников, представления отчетности для руководителей, а также определения наиболее успешных приемов воздействия.

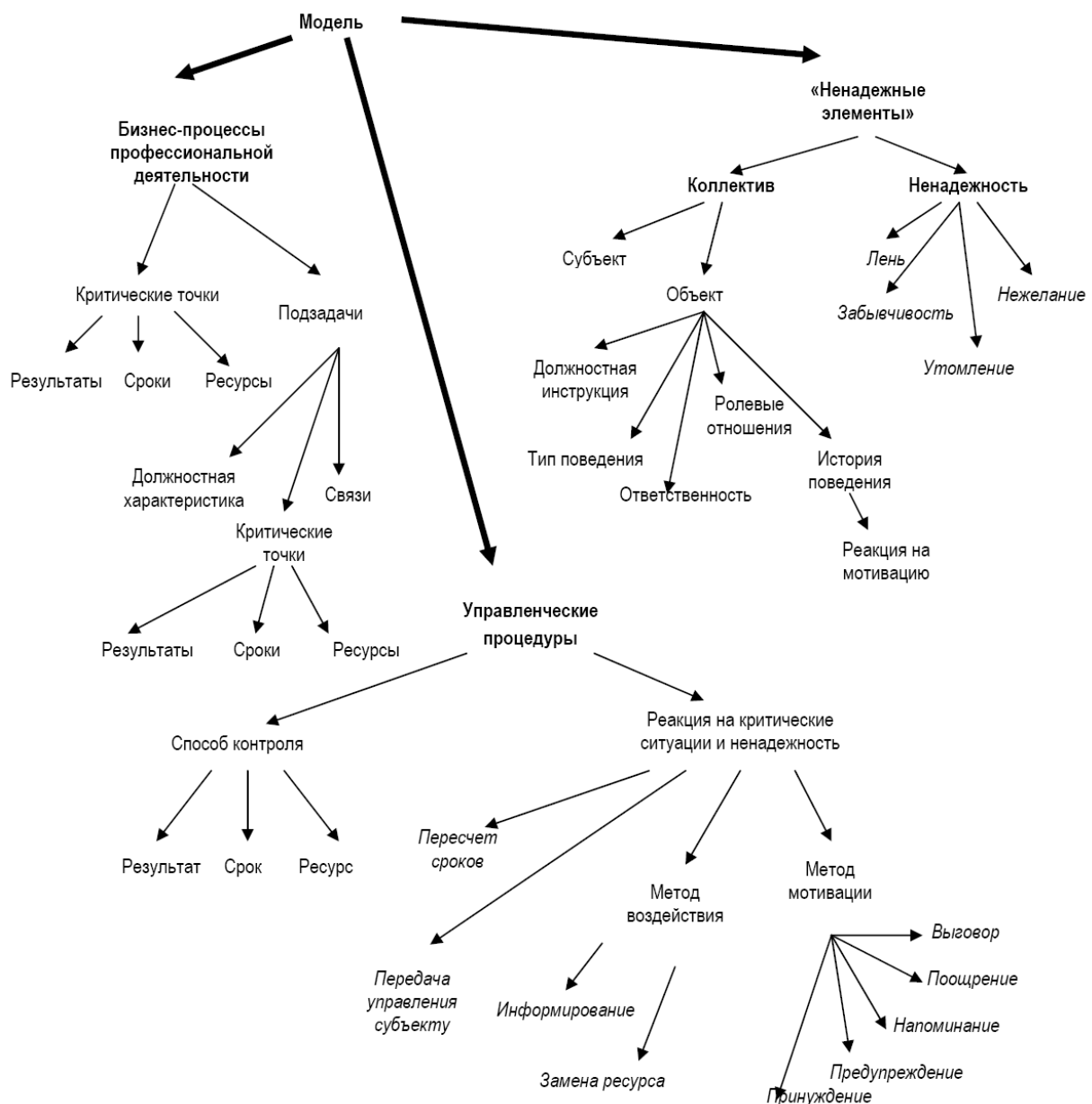


Рис.1. Элементы модели интеллектуальной системы управления

Таким образом, при разработке интеллектуальной системы управления «ненадежными элементами» в архитектуру системы необходимо закладывать механизмы, которые:

- отслеживают эффективность выполнения плана решения профессиональных задач, оперативно реагируют на изменения хода выполнения этого плана;
- используют различные процедуры управления «ненадежными элементами», затрагивающие различные методы управления персоналом (административные, экономические, социально-психологические), применительно к определенным людям (согласно истории их реакций на управляющие воздействия, типа личности и типа поведения).

Все возможные механизмы управления описываются в системе в виде отдельных процедур управления и составляют базу знаний процедур управления.

Описываемый в статье подход к представлению знаний и функционированию системы управления может быть выражен в рассматриваемой ниже архитектуре системы управления.

Архитектура интеллектуальной системы управления «ненадежными элементами». База знаний процедур управления.

Основные подсистемы архитектуры разрабатываемой интеллектуальной системы управления «ненадежными элементами» представлена на Рис. 2.

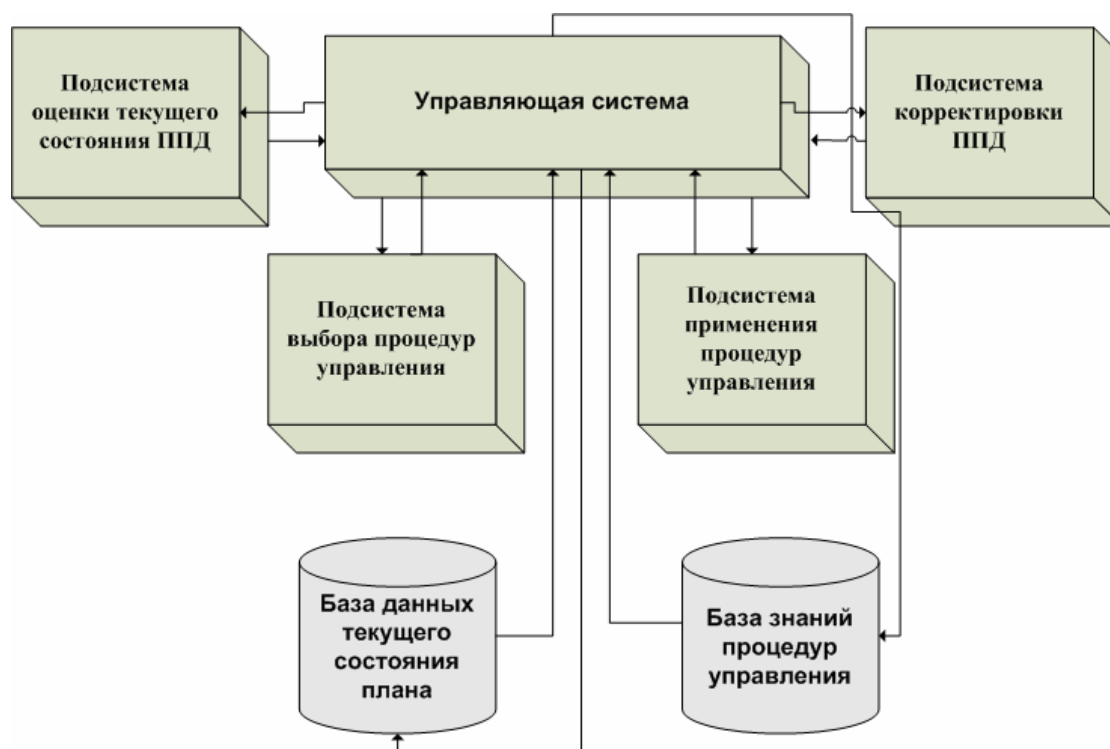


Рис.2.. Основные элементы архитектуры интеллектуальной системы управления

Рассмотрим элементы архитектуры:

Управляющая система – элемент архитектуры, обеспечивающий взаимодействие между подсистемами и информационными компонентами системы управления. Осуществляет передачу данных и управления между подсистемами.

Подсистема оценки текущего состояния плана профессиональной деятельности (ППД) – анализирует текущее состояние плана в **базе данных (БД) о текущем состоянии** и сравнивает его с заданным ППД для определения критических точек (обстоятельства, при которых возможен срыв плана).

Подсистема выбора процедур управления – анализирует получаемую информацию о текущем состоянии деятельности и информацию об эффективности применяемых процедур управления, и выбирает из **БЗ процедур управления** наиболее эффективный набор процедур управления, оправданный при данной ситуации.

Подсистема применения процедур управления – применяет процедуры управления, отслеживает реакции на их применение и сохраняет данные о реакциях на процедуры управления в **БЗ процедур управления**.

Подсистема корректировки ППД – переопределяет план профессиональной деятельности с учетом действий процедур управления и представляет текущее состояние плана в **БД о текущем состоянии**.

База данных текущего состояния плана профессиональной деятельности - содержит информацию о бизнес-процессах, представляющих план профессиональной деятельности: содержит информацию о сроках выполнения, критических точках, ресурсах.

База знаний процедур управления содержит знания о применяемых системой управления «ненадежными элементами» процедурах управления; знания о реакциях «ненадежных элементов» на применение этих процедур.

Планируется расширение данной архитектуры за счет добавления других подсистем (например, подсистема редактирования знаний, подсистема планирования деятельности и других) для создания полнофункционального программного средства.

Заключение

В статье изложен взгляд автора на подход автоматизированного интеллектуального управления деятельностью людей с учетом механизмов влияния на человека как «ненадежного элемента» в системе управления.

Описан подход к разработке модели на основе процедур управления. Выделены основные компоненты модели. Затрагиваются вопросы архитектуры системы, способной обеспечить контроль выполнения задач и повысить эффективность выполнения профессиональных задач людьми, используя процедуры управления.

Данный подход позволит создать интеллектуальное программное средство, которое в отличие от существующих на сегодняшний день программных систем управления, будет охватывать все характеристики и свойства человека и повысит надежность человека при решении им профессиональных задач.

Литература

[Benger, 2006] T.V. Ryabtsev, E.I. Antonova, R.V. Benger, "The Model of Intelligent system of activity people control in solving professional problems". In: "Artificial Intelligence 2006" magazine, Institute of Artificial Intelligence, Ukraine, Donetsk, 2006.

Информация об авторах

Roman Benger, Elena Antonova – Far East State University, Institute of Mathematics and Computer Sciences, P.O. Box: 690002, Okeanskiy avenue, 99 - 136, Russia, Vladivostok; e-mail: stainberg@mail.ru

МАТЕМАТИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ПРОГРАММНОГО КОМПЛЕКСА “ПЕРСОНАЛ” ПРИ ОПРЕДЕЛЕНИИ ТРУДОВОГО УЧАСТИЯ ЧЛЕНОВ КОЛЛЕКТИВА

Юрий Бондарчук, Григорий Гнатиенко

Аннотация: Рассматривается математическое и программное обеспечение автоматизированной системы “Персонал” при определении коэффициентов трудового участия членов коллектива. Приводятся возможности системы. Рассмотрен алгоритм определения коэффициентов трудового участия и компетентности руководителей структурных подразделений. Система “Персонал” внедрена на ряде предприятий пищевой промышленности Украины, где она успешно эксплуатируется на протяжении ряда лет.

Ключевые слова: обобщенное оценивание, коэффициент трудового участия, принятие решений, компетентность, эксперты, персонал.

Введение

В трудовых коллективах со сложной иерархической структурой чрезвычайно важную роль играют процедуры, описывающие различные способы мотивации. При наличии большого количества работников и структурных подразделений управление процедурами мотивации персонала невозможно без соответствующей автоматизации. Одной из функций автоматизированной системы персонал, описанной в работах [Бондарчук, 2002], [Бондарчук, 2004а], [Бондарчук, 2004б], [Бондарчук, 2006], является алгоритмизация определения, обеспечения анонимности и архивирования информации об экспертном определении меры трудового участия членов коллектива. Для адекватного оценивания экспертных данных при разработке подсистемы определения коэффициентов трудового участия был применен метод расчета оценок объектов и относительной компетентности экспертов в задаче обобщенного оценивания, описанный в работах [Волошин, 1993], [Гнатиенко, 2000а], [Гнатиенко, 2000б].

Постановка задачи обобщенного оценивания

Группа из k экспертов, которыми могут быть руководители подразделений разных уровней, с некоторыми начальными коэффициентами компетентности $\gamma_i \in [\gamma_i^H, \gamma_i^B]$, $i \in L = \{1, \dots, k\}$, где γ_i^H, γ_i^B — соответственно нижняя и верхняя границы шкалы коэффициентов, оценивает степень трудового участия членов коллектива со множеством индексов $I = \{1, \dots, n\}$. В результате экспертного оценивания порождается матрица оценок

$$C^{(0)} = (c_{ij}^{(0)}), \quad i \in L, \quad j \in I, \quad (1)$$

где $c_{ij}^{(0)} \in [c^H, c^B]$, $i \in L, j \in I$, — оценка i -м экспертом j -го работника, c^H, c^B — соответственно нижняя и верхняя границы шкалы оценок. Необходимо, учитывая весовые коэффициенты компетентности экспертов, на основании оценок (1) определить взвешенные результирующие (интегрированные, агрегированные) оценки каждого работника, рассчитать степень согласованности мнений экспертов о значимости каждого члена коллектива, оценить результирующую компетентность экспертов по их предыдущим оценкам согласованности с большинством экспертов. Понятно, что матрица

оценок (1) может быть неполной. В этом случае единственным требованием к матрице оценок $C^{(0)}$ будет наличие по каждому работнику оценки хотя бы одного эксперта.

Следует подчеркнуть, что уровень компетентности экспертов может зависеть от разных факторов: уровня руководителя в иерархической структуре предприятия, степени знакомства с работой оцениваемого коллектива, опыта и навыков адекватного оценивания, объективности или предвзятости к некоторым членам коллектива и т.п. Поэтому при разработке системы "Персонал" было принято решение применить процедуры коллективного оценивания с целью сглаживания возможных недостатков, приведенных выше.

При решении задач экспертными методами невозможно обойтись без применения дополнительной информации от экспертов или организаторов экспертизы. Введем эвристику для дальнейшего изложения задачи обобщенного оценивания, которую использовано для определения коэффициентов трудового участия.

Эвристика E1. Задана функцию изменения компетентности экспертов при оценивании каждого члена коллектива, т.е. матрица компетентности каждого эксперта при оценивании им любого работника в виде

$$\varphi_j(\gamma_i) = \gamma_{ij}^{(0)}, \quad i \in L, \quad j \in I.$$

Таким образом, коэффициенты компетентности экспертов являются переменными и зависят от работников, которых они оценивают. Т.е. допускается, что некоторых членов коллектива эксперт может оценивать более адекватно, чем другие эксперты и, соответственно, при оценке некоторых работников он имеет большие значения показателей компетентности, чем при оценке других членов коллектива. Будем также считать, что ряды элементов матрицы (1) являются нормированными, т.е.

$$\sum_{i \in L} c_{ij}^{(0)} = 1, \quad \forall j \in I.$$

Алгоритм определения коэффициентов трудового участия

Учитывая начальные коэффициенты компетентности экспертов для всех работников $\gamma_{ij}^{(0)}$, $i \in L$, $j \in I$, рассчитаем средние оценки каждого члена коллектива

$$c_j^{(1)} = \sum_{i \in L} \gamma_{ij}^{(0)} c_{ij}^{(0)}, \quad j \in I. \quad (2)$$

Рассчитаем теперь еще раз коэффициенты компетентности экспертов при оценивании членов коллектива с учетом полученного вектора $(c_1^{(1)}, \dots, c_n^{(1)})$ и осуществим их нормирование

$$\gamma_i^{(1)} = \sum_{j \in I} c_{ij}^{(0)} c_j^{(1)} / \sum_{l \in L} \sum_{q \in L} c_{lq}^{(0)} c_q^{(1)}, \quad i \in L. \quad (3)$$

Величины (3) формируют новый вектор относительных показателей компетентности экспертов. При такой процедуре осуществляется также уточнение значений функции смены компетентности экспертов в зависимости от оцениваемых структурных подразделений предприятия с учетом эвристики E1

$$\varphi_j(\gamma_i^{(1)}) = \gamma_{ij}^{(1)}, \quad i \in L, \quad j \in I. \quad (4)$$

Учитывая полученные оценки компетентности экспертов, выполним повторный расчет интегрированных оценок работников $c_j^{(2)} = \sum_{i \in L} \gamma_{ij}^{(1)} c_{ij}^{(1)}$, $j \in I$.

Теперь, учитывая аналогичным способом оценки работников $(c_1^{(2)}, \dots, c_n^{(2)})$, повторно рассчитываем относительные коэффициенты компетентности экспертов, вычисляя значения $\gamma_i^{(2)}, i \in L$, по формулам вида (3) и т.д. Эта процедура применяется до тех пор, пока гиперпараллелепипед весовых коэффициентов, состоящий из значений $\gamma_i^{(s)}, i \in L, s = 1, 2, \dots$, перестанет изменяться.

Понятно, что начальные нормированные весовые коэффициенты компетентности экспертов могут быть как равными, так и требующих применения процедуры (2)-(3). В последнем случае стабилизация предпочтений становится "направленной" в сторону мнений более "весомых" экспертов. Компетентность экспертов в этом случае также будет смещенной.

Если заданы явные функциональные зависимости (4) "весов" экспертов относительно членов коллектива, что оценивается, то они используются на следующих итерациях: $\gamma_{ij}^{(s)} = \varphi_j(\gamma_i^{(s)})$, $i \in L, j \in I$, где $\gamma_{ij}^{(s)}$ — величины, рассчитанные по формуле (4). Если же функции смены компетентности экспертов в зависимости от оцениваемых работников, заданы в табличном виде, то эта зависимость оценивается для последующих итераций на основании соотношений или закономерностей, которые можно выявить в заданной таблице (4).

Результатом решения задачи будет два массива данных: интервалы оценок работников $[c_j^H, c_j^B]$, $j \in I$, где $c_j^H = \min_{s=0,1,\dots} c_j^{(s)}$, $c_j^B = \max_{s=0,1,\dots} c_j^{(s)}$, $j \in I$, — соответственно наименьшее и наибольшее значения оценок для всех шагов процедуры; а также интервалы относительной компетентности экспертов $[\gamma_i^H, \gamma_i^B]$, $i \in L$, где $\gamma_i^H = \min_{s=0,1,\dots} \gamma_i^{(s)}$, $\gamma_i^B = \max_{s=0,1,\dots} \gamma_i^{(s)}$, $i \in L$.

В случае, когда матрица вида (1) является неполной, т.е. если допускаются ситуации, когда i — й эксперт не может оценить j — го работника, можно предложить обобщенное описание процедуры определения оценок членов коллектива и коэффициентов компетентности экспертов. Обозначим множество индексов экспертов, которые оценивают j — го работника, через L_i , $i \in I$, $L_i \subseteq L$, $i \in I$; а через $L_i^?$, $i \in I$, $L_i^? \subseteq L$, $i \in I$, — множество индексов экспертов, которые не в состоянии оценить j — го члена коллектива, $L_i \cup L_i^? = L$, $i \in I$.

Результирующие (средние, интегрированные, агрегированные) оценки каждого работника с учетом условия нормированности на каждом шаге $s = 0, 1, 2, \dots$, определяются таким образом:

$$c_j^{(s)} = \sum_{i \in L_i} \gamma_{ij}^{(s-1)} c_{ij}^{(s-1)} / \sum_{l \in L_i} \gamma_{lj}^{(s-1)}, \quad j \in I.$$

Коэффициенты компетентности экспертов на очередном шаге процедуры определяются формулой

$$\gamma_i^{(s)} = \sum_{\substack{j \in I \\ i \in L_j}} \gamma_{ij}^{(s-1)} c_j^{(s-1)} / \sum_{\substack{l \in I \\ i \in L_l}} \sum_{q \in I} \gamma_{lq}^{(s-1)} c_q^{(s-1)}, \quad i \in L.$$

Далее процедура обобщенного оценивания работников и определения коэффициентов компетентности экспертов, которые решают эту задачу, для неполных матриц вида (1) описывается по аналогии с процедурой, представленной формулами (1)-(4) для полных матриц.

Возможности подсистемы «Мотивация» программного комплекса «Персонал»

Описанная выше задача обобщенного оценивания реализована в подсистеме «Мотивация» программного комплекса «Персонал». В рамках этой подсистемы выполняются:

1. Поддержка работы с иерархической структурой предприятия. Система может работать в режиме *одного* (автономного) предприятия или *распределенной* системы. В последнем варианте каждое предприятие компании, работая со своей частью БД, все изменения с заданной периодичностью передает в центральную систему. Соответственно на предприятиях ведется текущее формирование данных, в центральной корпоративной части, кроме информационно-поисковых функций, решаются аналитические задачи, одной из которых есть "Мотивация". Данные, которыми обмениваются распределенные системы на предприятиях компании между собой и центральной системой, кодируются в специальном формате и в описываемой версии системы передаются через электронную почту.

2. Ведение электронного табеля использования рабочего времени с детализацией времени и качества труда. Можно констатировать имеющуюся сложность полного представления данных о работнике в электронном виде, например, часть работ короткое время смены выполняются по совместительству, есть масса форм совмещения должностей в пищевой и строительной промышленности. Паллиативным выходом всегда был вариант привязки оценки труда к ставке и/или объему продукции целого подразделения, например. Но тогда размывается вклад каждого работника. В системе «Персонал» используется электронный табель с любой степенью детализации выполненных операций и периодов времени.

Но использование табеля исчерпывающе приемлемо для технической работы на производстве с простыми операциями, когда учет объема и качества произведенного дает достаточно объективную оценку. И от эксперта для этого требуется минимум компетенции. И в этой части реализации большинства кадровых систем идентичны. Но кадровая система не есть бухгалтерской системой, хотя они и могут быть интегрированными. Поэтому для подсистемы "Мотивация" важна только экспертная оценка качества и количества трудового участия.

3. Ввод и коррекции экспертных оценок. Первоначально формируется база экспертов, часть которой автоматически составляют руководители подразделений. Экспертам присваиваются коэффициенты их компетентности в оценивании персонала. Система отслеживает то, чтобы по каждому из работников была хотя бы одна оценка.

4. Архивирование экспертных оценок. Информация в архиве есть основой данных для формирования кадрового резерва компании, что, в свою очередь, есть отдельной задачей.

Выводы

Рассмотрены основные задачи, которые возникают в службах управления персоналом большой компании в сфере мотивации. Предложена схема решения задачи материальной мотивации персонала на основе алгоритма обобщенного оценивания. Описанная схема позволяет службе управления персоналом осуществлять целенаправленный процесс мотивации трудовых ресурсов предприятия.

Библиография

[Бондарчук, 2002] Бондарчук Ю.В., Гнатієнко Г.М. Розподілена кадрова система підтримки прийняття рішень//Праці міжнародної школи-семінару "Теорія прийняття рішень", Ужгород, УжНУ, 2002.-С.12.

[Бондарчук, 2004а] Бондарчук Ю.В., Гнатієнко Г.М. Оцінка ризиків прийому кандидатів на роботу та ефективності роботи працівників//Праці II-ї міжнародної школи-семінару "Теорія прийняття рішень", Ужгород, УжНУ, 2004.-С.10.

- [Бондарчук, 2004b] Бондарчук Ю.В., Гнатієнко Г.М. Прийняття рішень щодо оцінки ефективності та ризиків прийому нових працівників до компанії //Сучасні проблеми математичного моделювання, прогнозування та оптимізації: Збірник наук. праць (за матер. Всеукр. наук.-метод. конф.). – Київ-Кам'янець-Подільський: Кам'янець-Подільський держ.унів., 2004, с.242-243.
- [Бондарчук, 2006] Бондарчук Ю.В., Гнатієнко Г.Н. Применение модели компетенций при решении задач управления персоналом //Proceeding of the XII-th International Conference "Knowledge-Dialogue-Solution", June 20-25, 2006, Varna (Bulgaria). – Sofia, FOI-COMMERCE - 2006. – p.165-172.
- [Волошин, 1993] Волошин О.Ф., Гнатієнко Г.М. Процедури визначення компетентності експертів// Вісн. Київ. ун-ту. Фіз.-мат. науки. – 1993. – № 3. – С. 102-111.
- [Гнатієнко, 2000а] Гнатієнко Г.М. Деякі математичні аспекти соціальної експертизи // Соціальна експертиза в Україні: методологія, методика, досвід впровадження; За ред. Ю.І.Саєнка. – К.: Ін-т соціології НАНУ, 2000. – 194 с.
- [Гнатієнко, 2000б] Гнатієнко Г.М. Методи оцінки компетентності спеціалістів. Математичні та інформаційні проблеми прогнозування наслідків техногенних та природних катастроф / Соціально-економічні наслідки техногенних та природних катастроф: експертне оцінювання; Відп.ред.: В.В.Дурдинець, Ю.І.Саєнка. – К.: "Стилос", 2000. – 260 с.

Информация об авторах

Юрій В. Бондарчук – *Национальный университет им. Т.Шевченко, факультет кибернетики, доцент, Киев, Украина, e-mail: byv@univ.kiev.ua*

Григорій Н. Гнатієнко – *Национальный университет им.Т.Шевченко, факультет кибернетики, докторант. Киев, Украина. e-mail: G.Gnatenko@veres.com.ua*

DESIGNING AND EVALUATING OF CREW ACTIVITY ALGORITHMS ON INITIAL STAGE OF DESIGNING OF ANTHROPOCENTRIC OBJECT

Boris Fedunov, Denis Vidruk

Abstract: *For reaching the goals of designing the system-forming kernel of modern anthropocentric object, algorithms of crew activity have to be designed already on the initial stage of its development. Technology of designing of crew activity algorithms and evaluating of time expenses for their realization is described.*

1. Introduction

Development of algorithms specifications of onboard digital computing systems (ODCS) and crew activity algorithms (CAA) of anthropocentric object (AO) includes following technical documents:

- An original-language technical document: "Logic of functioning of the "crew-onboard equipment" system". Text of this document is usually structured by typical situations (TS) of designed object functioning and their problematic sub situations (PsS);
- ODCS-algorithms and CAA specifications structured by TS and their PsS;
- Realizability estimation of designed onboard algorithms: ODCS-algorithms for onboard digital computing system and CAA for crew;
- Efficiency estimation of designed onboard algorithms.

2. Macro model of anthropocentric object for onboard algorithms specifications development

AO model consists of three global control levels (OGCL):

- I - Operative choice of purpose of current functioning (the purpose choice level).
- II - Choice of a rational way of achievement of appointed purpose.
- III - Realization of chosen actions.

The model of AO functioning includes: set of the general functioning purposes that is appointed for algorithmization; representation of each general purpose through a semantic network of TS; representation of each TS through a semantic network of PsS; OGCL level I and II problems are solved at system-forming kernel of AO.

2.1 Crew member (operator) model of onboard algorithms specifications development

Activity structure of operator of AO includes: decisions appointment processes; participation of the operator in processes of tracking as a part of tracking system; realization of a chosen decision. During development of CAA specifications the screens of information on the cabin indicators, structure of voice messages, lists of controls of an informational-control field of a cabin of crew are created. Decisions made by operator in a functioning process are divided in three types: π – perceptive-identificational decisions that do not demand any time to think over; ρ – speech-thought decisions that can be divided into several elementary acts of making decision; π - ρ – heuristic decisions.

Time expenses for making a π -decision are the time for reaction on signals of the cabin informational field: acoustical or light irritation, recognition of conventional signs on indicators etc. The corresponding average data about time of reaction on various irritations can be found in the literature on subject psychology or by experimental way.

For ρ -decisions it is necessary to take into account the time to think over the situation besides similar reactions expenses. Elementary acts of making decision (EAMD) are elementary logical operations which operator algorithm of making decisions, based on data from corresponding devices can be divided to. For example: simple arithmetic calculation, comparison of digital or literal signs, check of a logic condition and/or. The corresponding average data about time spent for EAMD can be found in the literature on subject psychology or by experimental way.

Time expenses for π - ρ -decisions can only be found by experimental way.

If there is a project of an operator cabin at a design stage of onboard algorithms specifications, the way of realization of the made decision (manual operations) and time spent for it are calculated on the basis of an arrangement and type of devices in a cabin. Manual operations are represented through elementary acts.

Operations of tracking are estimated on time expenses which are spent by the operator for elimination of a mistake of the tracking which have collected during the operator distraction from process of tracking. Discretely-

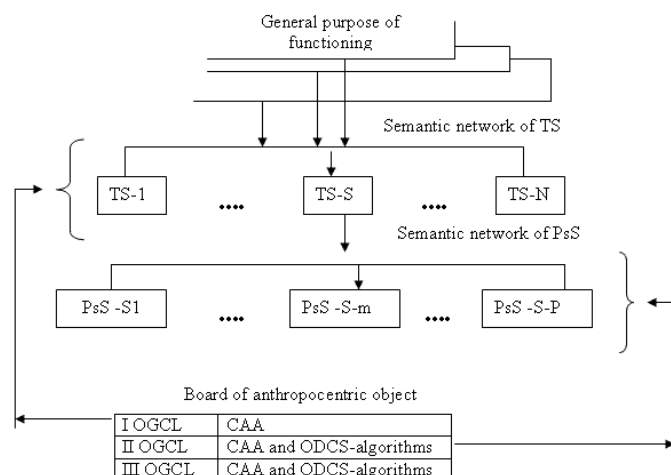


Figure 1. Macro model of anthropocentric object

continuous model of tracking is used. For calculation of time expenses for process of tracking, experimental distraction time dependence of mistake elimination time during a tracking process is determined. A possible arrangement of tracking updating points is placed on CAA earl by a design-engineer. For an estimation of CAA earl realizability optimum arrangement of tracking updating points gets out among all possible.

3. Specification development technology

For development of an original-language technical document "Function logic of the crew and onboard equipment system" and onboard algorithms specifications, the group of plenipotentiaries of organizations participating in development of AO is created. This document is structured by OGCL, TS, PsS and usually consists of several books: one book of "Function logic of the crew and onboard equipment system" of the 1'st level of OGCL and several books (in the number of TS) of "Function logic of the crew and onboard equipment system" for concrete TS.

Using the text of each book the generalized earl is designed as ordered by causally investigatory attitude the set of TS and the set of PsS.

Formal completeness of texts of these books is estimated by members of plenipotentiaries with the help of the "Logic-Text-Analysis" (LoTA) computer system. For each TS and its set of PsS the scheme of onboard algorithms (SOA) will be made. These algorithms are realized on ODCS. The operator decision earl (ODE) has to be made either. ODE is structured by causally investigatory attitude set of CAA that includes:

- decision making processes;
- participation of the operator in processes of tracking as a part of tracking system;
- realization of the chosen decision.

A possible arrangement of tracking updating points is placed on ODE by a design-engineer manually and the optimal arrangement is calculated. The design process of coordinated ODE and SOA is supported by "Bort" computer system.

3.1 CAA specifications design computer system "BORT"

The computer system "Bort-ADE" is crew activity algorithms specifications development computer system. It is a part of the "Bort" computer system named above. It is designed for plenipotentiaries group members and support development of ODE of selected PsS in appointed TS. The initial information for development includes:

- texts of "Function logic of the crew and onboard equipment system";
- set of recommended signs for information representation to the operator on the cabin displays;
- hardware-semantic shape of the crew cabin informational-control field including set and arrangement of control devices, types of operator inputs into ODCS, etc.

"Bort-ADE" system supports designers while constructing ODE of TS for selected PsS:

- finding fragments of selected PsS in texts of "Function logic of the crew and onboard equipment system" and in generalized earl;
- selecting next problem in this fragment;
- detalisation of operator accessible information of the informational-control field needed for solving selected problem of the current moment of AO functioning;
- decision type identification (π , ρ or $\pi-\rho$);

- introducing of accessible information using allowed symbols to the operator;
- verbal describing of operator solving decision of the selected problem;
- collecting the set and semantic features of ODCS-algorithms forming information for the operator to solve selected problem;
- representation of the CAA set of PsS as ODE;
- forming the possible arrangement of tracking updating points;
- determining the distraction time dependence of mistake elimination time during a tracking process using experimental data;
- uniting ODEs of each PsS into common TSS ODE.

The "Bort-ADE" system includes databases for: recommended symbols of information representation; hardware-semantic shape of the crew cabin informational-control field; crew activity time expenses in different TSs and PsSs (is filled while ODE constructing): normal, stress etc.

4. Realizability estimation of ts CAA set computer system "GRO-OTSENKA"

"GRO-otsenka" system is intended for automatization of ODE realization estimation process. On the basis of the data entered manually by the design-engineer or received from system "Bort", the program optimally arrangements tracking updating points, time expenses for each CAA, their mean-square deviations and similar parameters for each branch the ODE and any set of CAA inside ODE branch. The system also allows changing entered data operatively, automatically determining bottlenecks of ODE. The initial information includes: operator cabin informational-control field data, voice messages for the operator, etc.

4.1 Example of designing and realizability estimation of CAA

Fragment of "Function logic of the crew and onboard equipment system" texts: "Onboard measuring devices have located objects "B1", "B2", "B3" and are showing them at a screen using Cartesian system of coordinates. Operator location is represented as object A. Operator has to order this objects by distance while taking part at a tracking process keeping directory mark "F" inside moveless round "H"."

ODE constructed using this fragment consists of:

1. Perception of the onscreen objects parameters.
2. Keeping directory mark "F" inside moveless round "H" (tracking beginning)
3. Comparison of pieces A-B1 and A-B2
4. Probability of tracking mistake correction
5. Comparison of pieces A-B1 and A-B3
6. Probability of tracking mistake correction
7. Comparison of pieces A-B2 and A-B3
8. Probability of tracking mistake correction
9. Transferring ordering algorithms into ODCS (manual operations).
10. There is a probability of tracking mistake correction after each CAA.

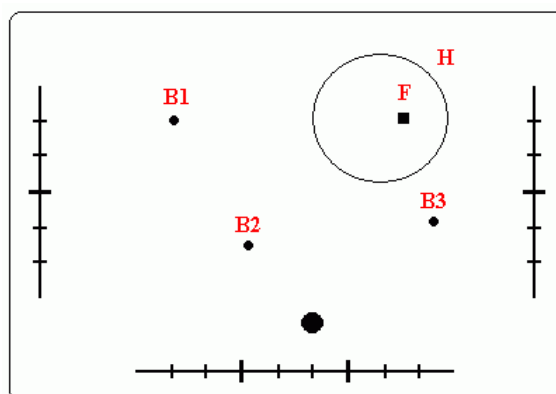


Figure 2. Onboard indicating screen example

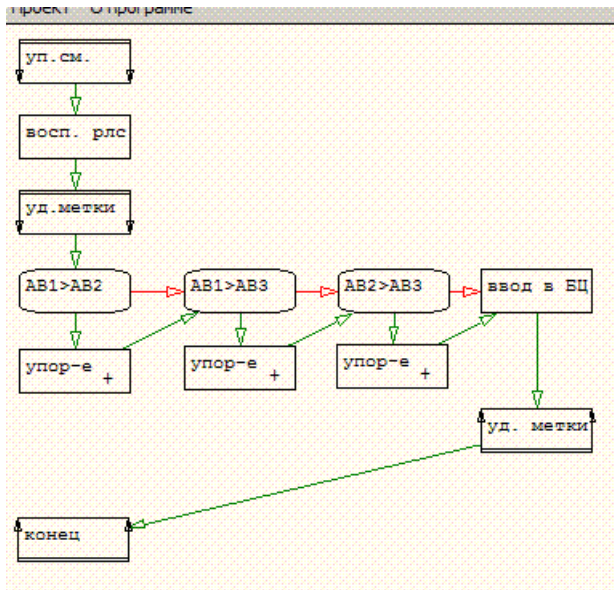


Figure 3. Example of ODE made with "Bort" system

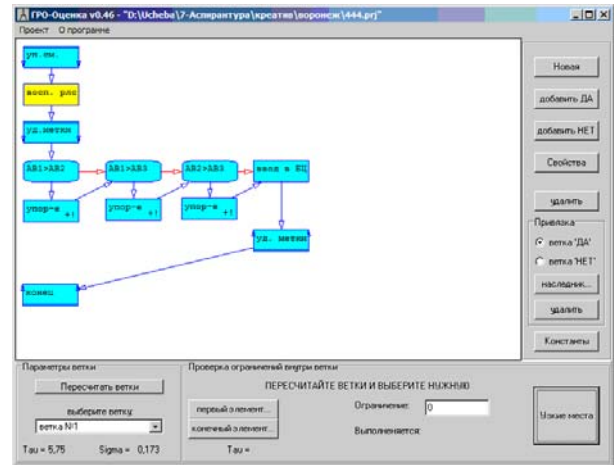


Figure 4. Example of "GRO-otsenka" results

4.2 ODE realization time expenses estimation

For realization time expenses estimation of this ODE fragment group of plenipotentiaries uses "GRO-otsenka" system. Time expenses and their mean-square deviations are calculated for each CAA and for whole ODE. Probable arrangement of tracking updating points and bottlenecks of ODE are determined automatically. The final result of the estimation stands out.

4.3 Example of estimation result report

I part	II part	III part
Branch №0 time: 5,75 deviation: 0,173 список CAA ветки:	CAA №4упор-е time= 0,7 deviation= 0,1 comment=	CAA №8упор-е time= 0 deviation= 0 comment=
CAA №1восп. рлс time= 3,51 deviation= 0,1 comment=	CAA №5AB1>AB3 time= 0 deviation= 0 comment=	CAA №9ввод в БЦ time= 0 deviation= 0 comment=
CAA №2уд.метки time= 0 deviation= 0 comment=	CAA №6упор-е time= 0,7 deviation= 0,1 comment=	CAA №10уд.метки time= 0 deviation= 0 comment=
CAA №3AB1>AB2 time= 0 deviation= 0 comment=	CAA №7AB2>AB3 time= 0 deviation= 0 comment=	CAA №11конец time= 0 deviation= 0

5. Conclusion

Design of crew activity algorithms set is necessary for solving problems of the system-forming kernel of anthropocentric object already on the initial stages of its development and directed on:

- determining of the information set shown to operator during typical problem situations;
- specification of control devices at the cabin;
- tentative estimation of loading of the operator problems decisions.

There are many inconsistencies between crew activity and onboard digital computing system algorithms development in the existing practice. Assistance of a computer support system covering all design stages, will raise quality of a product and will reduce expenses for its creation.

Bibliography

1. Fedunov B.E. 2002 "The Efficiency estimation of on-board intellect of the anthropocentric systems specifications algorithms", "Mechatronic, Automation, Control", Moscow (in Russian).
2. Fedunov B.E. 2002 "Methods of realizability estimation of anthropocentric system operator decision earl", State Research Institute of Aviation Systems, Moscow (in Russian).
3. Nevzorova O.A., Fedunov B.E. 2001 "Technical text analysis system: the main concepts and design", "Control Systems and Theory", Russian Science Academy (in Russian).
4. Kondrikova T.A., Fedunov B.E. 1999 "Bort - On-board artificial intellect specifications design process support computer system" "Control Systems and Theory", Russian Science Academy (in Russian).
5. Fedunov B.E., Vidruk D.G. 2006 "Computer system of a crew activity algorithms realizability estimation. Modeling Methodologies and Simulation", "Key Technologies in Academia and Industrial.", "Proceedings of 20th European Conference on Modeling and Simulation (ECMS2006) »

Authors' Information

Boris Fedunov – State Research Institute of Aviation Systems; Moscow Aviation Institute; Moscow, Russia; email: boris_fed@gosnias.ru

Denis Vidruk – Moscow Aviation Institute; Moscow, Russia; email: dinizzzz@mail.ru

E.4 Formal Models

UML: ИСТОРИЯ, СПЕЦИФИКАЦИЯ, БИБЛИОГРАФИЯ

Дмитрий Буй, Елена Шишацкая

Резюме: Сделан короткий экскурс в историю возникновения и развития современного универсального языка моделирования UML, приведен короткий обзор языка и предпринята попытка систематизировать научно-методическую литературу, посвященную этому языку.

Ключевые слова: UML, объектно-ориентированное проектирование, спецификация, семантика.

ACM Classification Keywords: E.4 Coding and information theory – Formal models of communication

Введение

В настоящее время все большую популярность приобретает язык моделирования UML (Unified Modeling Language), который стал “стандартом де-факто” в области разработки программного обеспечения и применяется для решения задач других предметных областей, например, задач бизнес-моделирования. В работе сделан короткий экскурс в историю возникновения и развития UML, приведен короткий обзор языковых структур и предпринята попытка систематизировать соответствующую библиографию.

Возникновение и назначение языка

В основе UML лежит несколько объектно-ориентированных методов, каждый из которых первоначально был ориентирован на поддержку отдельных этапов *объектно-ориентированного анализа и проектирования* (ООАП) [1]:

- метод Гради Буча (Grady Booch), условное название Booch (Booch’91, Booch Lite, Booch’93) – считался наиболее эффективным на этапах проектирования и разработки программных систем [1];
- метод Джеймса Румбаха (James Rumbaugh), Object Modeling Technique (OMT, позже OMT-2) – оптимально подходил для анализа процессов обработки данных в информационных системах [2];
- метод Айвара Джекобсона (Ivar Jacobson), Object-Oriented Software Engineering (OOSE) – содержал средства представления вариантов использования, имеющих существенное значение на этапе анализа требований в процессе проектирования бизнес-приложений [3].

История развития UML датируется 1994 г., когда началась интеграция/унификация вышеуказанных методов. Проект унифицированного метода (Unified Method) версии 0.8 был опубликован в октябре 1995 г.

Все вопросы разработки и сопровождения языка UML сконцентрированы в рамках консорциума OMG (Object Management Group) [4]. Хотя OMG был создан с целью разработки предложений по стандартизации объектных и компонентных технологий CORBA, язык UML приобрел статус второго стратегического направления в работе консорциума. В ноябре 1997 г. OMG объявил UML стандартным языком объектно-ориентированного моделирования и принял на себя обязанности по его последующему развитию. Группа специалистов обеспечивает публикацию описаний последующих версий языка UML и запросов предложений RFP (Request For Proposals) по его стандартизации. Статус языка UML определен как открытый для всех предложений по доработке и усовершенствованию. В 2003 г. в качестве результата

рассмотрения набора запросов RFP 2000 г. было опубликовано описание языка UML 2.0, включающее инфраструктуру UML [5], язык ограничений объектов (Object Constraint Language – OCL) [6], суперструктуру UML [7] и формат обмена диаграмм [8]. Основными инициативами консорциума OMG в рамках работы над проектом UML являются: (1) моделирование систем реального времени; (2) определение модели выполнения – точной спецификации поведения моделей, поддерживающихся UML; (3) обработка данных предприятия – определены так называемые профили, описывающие способы создания больших распределенных параллельных систем предприятия; (4) определение процесса разработки программного обеспечения – специфицированы структуры определения процессов разработки программного обеспечения; (5) стандарт хранения данных; (6) сопоставление технологии CORBA и языка UML; (7) формат XMI (Metadata Interchange) для обмена моделями UML в текстовом формате [9].

Существует консорциум партнеров UML (Digital Equipment Corp., HP, Intellicorp, IBM, ICON Computing, Microsoft, Oracle, Rational Software и другие), обеспечивающих уточнение нотации, усовершенствование и дополнение языка, а также сопровождение разработки инструментальных средств поддержки. Особое место занимает компания Rational Software Corporation [10], реализовавшая Rational Rose 98 – одно из первых инструментальных CASE-средств, в котором был поддержан язык UML [11, 12]. Необходимо отметить внимание компании Microsoft к технологии, базирующейся на языке UML, на основе которой создана информационная модель (UML Information Model), предназначенная для создания стандартного интерфейса между средствами разработки приложений и средствами визуального моделирования. На данный момент Rational Software Corporation официально входит в состав IBM (<http://www-306.ibm.com/software/rational/>).

Язык UML предназначен для решения следующих заданий [13]: (1) предоставить в распоряжение пользователей готовый к использованию выразительный мощный язык визуального моделирования, позволяющий разрабатывать осмысленные модели и обмениваться ими; (2) предусмотреть внутренние механизмы расширяемости и специализации базовых концепций языка; (3) обеспечить максимальную независимость проекта создания программного обеспечения от конкретных языков программирования и процессов разработки; (5) обеспечить формальную основу для однозначной интерпретации языка; (6) стимулировать расширение рынка объектно-ориентированных инструментальных средств создания программного обеспечения; (7) интегрировать лучший практический опыт использования языка и реализации программных средств его поддержки.

Принципы использования UML специфицированы в Rational Unified Process (RUP) – развитой методике создания программного обеспечения, оформленной в виде базы знаний, физически размещенной на web (<http://www-306.ibm.com/software/awdtools/rup/support/>) и оснащенной поисковой системой [14, 15]. Этот коммерческий продукт задает строгий регламент распределения задач и ответственности между исполнителями в процессе разработки программного обеспечения.

Основным документом по UML является [16], где описана *метамодель* UML и очень мало внимания уделяется семантике языковых конструкций. Описание текущей версии языка UML [5-8] и примеры разработки программных систем с использованием CASE-средства Rational Rose можно найти на веб-узле OMG (www.omg.org); модификации языка UML и его новейшие версии – на узле www.celigent.com/uml.

Так как официальная документация по UML весьма затруднительна для понимания, то выходит много книг, описывающих язык с разными акцентами. Например, базовая система обозначений UML популярно и доступно изложена в книге Мартина Фовлера (Martin Fowler) [17], которая считается одним из наилучших пособий по изучению языка. Следует отметить книги, написанные главными авторами UML – Г. Бучем, И. Джекобсоном, Д. Рембо (в другом переводе Румбахом или Рамбо): в [13] детальная информация об использовании UML (покрывает около 80% языка) проиллюстрирована применением языка на большом количестве примеров; в [15] рассматривается процесс объектно-ориентированной разработки

программного обеспечения; в справочнике по UML [18] охватывается весь язык и делается попытка раскрыть его содержательную семантику. Начинаящим стоит обратить внимание на издание [19], вышедшее в серии "... для чайников"; детальное и полное описание языка также приведено в [20]. Вопросам использования продукта Rational Rose посвящены [21, 22], но наиболее развернуто и доступно моделирование систем средствами UML с использованием Rational Rose описано и проиллюстрировано большим количеством примеров в [23] (переиздание [24] – расширено и адаптировано под UML 2.0). Также необходимо выделить серию книг под редакцией главных авторов языка: "Серия объектные технологии" издательства Addison Wesley посвящена применению языка [14, 25, 26, 27, 28]; применению UML и шаблонов проектирования посвящены также [29, 30].

UML создавался как язык моделирования общего назначения для применения в таких "дискретных" областях, как программное обеспечение, аппаратные средства и цифровая логика. Структуры UML позволяют фиксировать разнообразные решения по отображению (1) функциональности системы [28, 31, 32], (2) динамической и статической структуры системы, (3) организации элементов системы [33], (4) реализации системы. Популярность приобретает использование UML при проектировании баз данных [26, 34, 35]. Благодаря открытости (наличию в языке механизмов расширения) он предоставляет мощный инструментарий для решения задач других отраслей, например, бизнес-моделирования [36].

Общая структура языка

Семантика языка UML определяется для двух видов объектных моделей: *структурных* и *поведенческих*. Структурные (статические) модели описывают структуру сущностей или компонентов системы, включая их *классы, интерфейсы, атрибуты* и *связи*. Модели поведения (динамические) описывают поведение или функционирование объектов системы, включая их *методы, взаимодействие (сотрудничество)* между ними, а также процесс изменения состояний отдельных компонентов и системы в целом [1].

Формальное описание языка UML основывается на следующей общей иерархической структуре модельных представлений, состоящей из четырех уровней абстракции: (1) мета-метамодель, (2) метамодель, (3) модель, (4) объекты пользователя [5]. Уровень мета-метамодели образует базовую основу для всех метамодельных представлений и определяет язык для спецификации метамодели. Мета-метамодель определяет модель языка UML на наивысшем уровне абстракции (соответственно на самом низком уровне конкретизации) и является наиболее компактным его описанием. Метамодель – экземпляр или конкретизация мета-метамодели – определяет язык для спецификации моделей. Все основные понятия языка UML – это понятие уровня метамодели. Модель в контексте языка UML является экземпляром (конкретизацией) метамодели в том понимании, что любая (конкретная) модель системы должна использовать только понятия метамодели, конкретизировав их применительно соответствующей ситуации. Содержательно говоря, уровень модели предназначен для описания конкретной предметной области. Конкретизация понятий модели происходит на уровне объектов, которые являются экземплярами модели и содержат конкретную информацию о предметной области в соответствии с понятиями модели.

Основой представления UML на метамодельном уровне является описание трех его *логических блоков (пакетов)*: *основные элементы, элементы поведения* и *общие механизмы* [7].

Концептуальная модель языка

Концептуальная модель языка включает основные строительные блоки, правила их сочетания и общие механизмы [13, 17, 18].

Словарь языка UML содержит *сущности* (абстракции, являющиеся основными элементами модели) и *отношения* (основные соединительные строительные блоки), Сущности и отношения по определенным правилам соединяются в конструкции – *диаграммы*.

В UML определено четыре типа сущностей [13]:

- *структурные сущности*, распадающиеся на *основные* (класс (Class), интерфейс (Interface), кооперация (Collaboration), прецедент (Use case), активный класс (Active class), компонент (Component), узел (Node)), *разновидности* основных (актер (Actor), сигнал (Signal), утилита (Utility, вид классов), процесс (Process), нить (Thread, вид активных классов)) и *остальные* (приложения (Application), документ (Document), файл (File), библиотека (Library), страница (Page), таблица (Table, вид компонентов));
- *сущности поведения* (Behavioral things) – *взаимодействие* (Interaction) и *автомат* (State machine);
- *группирующие сущности* – *пакет* (Packages);
- *анотационные сущности* – *примечание* (Note).

Основными типами отношений в UML являются отношения: *зависимости* (Dependency), *ассоциации* (Association) (разновидностью ассоциации является отношение *агрегации* (Aggregation)), *обобщения* (Generalization) и *реализации* (Realization). Существуют также их вариации, например, *уточнение*, *трассировка*, *включение* и *расширение* (для отношений зависимости).

Для построения корректно оформленной модели в UML определены правила, позволяющие корректно и однозначно определять: (1) имена сущностей, отношений и диаграмм, (2) область действия имен (контекст, в котором имя имеет некоторое значение), (3) видимость имен (для использования другими элементами), (4) целостность (правильность и согласованность соотношения элементов), (5) выполнение модели [13].

Эффективность и упрощение применения языка обеспечивается использованием определенных соглашений, так называемых, *общих механизмов*: *спецификаций* (Specifications), *дополнений* (Adornments), *принятых распределений* (Common divisions) и *механизмов расширения* (Extensibility mechanisms) [13, 17, 18].

Каждый элемент нотации UML имеет уникальное графическое обозначение и спецификацию – текстовое представление синтаксиса и содержательной семантики соответствующего строительного блока. Практически все строительные блоки характеризуются дихотомией “класс / объект” и “интерфейс / реализация”. Это основные подходы деления реальности при объектно-ориентированном моделировании систем.

UML допускает контролируемые расширения для адаптации языка к конкретным потребностям. Наличие внутренних механизмов расширения принципиально отличает UML от таких средств моделирования как IDEF0, IDEF1X, IDEF3, DFD и ERM [37], являющихся замкнутыми и не допускающими расширения средствами самого языка. К *механизмам расширения* UML относятся: (1) *стереотип* (Stereotype), расширяющий словарь языка (позволяет создавать из существующих блоков новые, специфические для конкретной решаемой задачи); (2) *тэг-значение* (Tagged value), расширяющее свойства строительного блока (дает возможность включать новую информацию в *спецификацию* элемента); (3) *ограничение* (Constraint), расширяющее семантику строительного блока (позволяет добавлять новые или модифицировать существующие правила посредством семантических ограничений, заданных на естественном языке либо на формальном языке OCL). Некоторые расширения приобрели такую популярность, что были внесены в стандарт текущей версии UML [13, 17, 18].

Графические конструкции языка

В нотации UML все представления о моделях сложной системы фиксируются в виде специальных графических конструкций, получивших название *диаграмм* [6]. Диаграмма в UML – это графическое представление набора элементов, изображаемое, как правило, в виде связанного графа с вершинами (сущностями) и ребрами (отношениями). Теоретически диаграммы могут содержать любые комбинации сущностей и отношений. Однако на практике применяется небольшое количество типовых комбинаций. В языке UML определенные следующие виды диаграмм: (1) *диаграмма вариантов использования* (Use

case diagram); (2) *диаграмма классов* (Class diagram); (3) *диаграммы поведения* (Behavior diagrams), к которым относятся *диаграмма состояний* (Statechart diagram), *диаграмма деятельности* (Activity diagram), *диаграммы взаимодействия* (Interaction diagrams); в свою очередь к диаграммам взаимодействия относятся *диаграмма последовательности* (Sequence diagram) и *диаграмма кооперации* (Collaboration diagram); (4) *диаграммы реализации* (Implementation diagrams), к которым относятся *диаграмма компонентов* (Component diagram) и *диаграмма развертывания* (Deployment diagram).

Перечень этих диаграмм и их названий является каноническим, они – неотъемлемая часть графической нотации языка UML. Процесс ООАП неразрывно связан с процессом построения этих диаграмм, при этом совокупность построенных диаграмм является самодостаточной в том понимании, что в них содержится вся информация, необходимая для реализации проекта сложной системы.

Каждая из диаграмм детализирует и конкретизирует разные представления о моделях сложной системы в терминах языка UML. При этом диаграмма вариантов использования является наиболее общей концептуальной моделью сложной системы, исходной для построения всех других диаграмм. Диаграмма классов есть, по сути, *логической моделью*, отображающей *статические аспекты структурного построения* сложной системы. Диаграммы поведения также являются разновидностями логической модели, которые отображают *динамические аспекты функционирования* сложной системы. И, наконец, диаграммы реализации служат для представления физических компонентов сложной системы и поэтому относятся к ее *физической модели*. Таким образом, интегрированная модель сложной системы в нотации UML представляется в виде совокупности отмеченных выше диаграмм.

Следует обратить внимание, что при моделировании системы часто строят *диаграмму объектов* (Object diagram), представляющую множество объектов и отношений между ними в заданный момент времени (“снимок системы”); не являясь канонической диаграммой UML, эта диаграмма полезна в детализации системы.

Подходы к спецификации семантики языка

Начальной целью развития UML было обеспечение более точного описания языка моделирования – поддержка формальной основы для понимания языка моделирования. Однако до настоящего времени формальная семантика не является частью стандарта. Обзор нескольких подходов, касающихся определения такой семантики, приведен в [38], где рассмотрены *теоретико-множественный, трансляционный, метамодельный* подходы и предложена так называемая “*свободная семантика*”. Следует отметить, что в самой спецификации UML существует много огрехов и противоречий: так, в работе [39] рассматриваются отношения включения и расширения для вариантов использования, в [40] анализируются стандартные стереотипы метамодели UML. В работе [41] предлагается использовать RM-ODP (Reference Model Open Distributed Processing [42]) для решения проблем языка. Отметим, что согласно спецификации UML модель RM-ODP выступает структурой, которая непосредственно влияет на архитектуру метамодели языка (раздел “Preface: Relationships to Other Models” [7]). Кроме того RM-ODP используется в MOF (Meta-Object Facility) для управления типами. В [41] идентифицируются три проблемы метамодели UML и предлагается их решение на базе RM-ODP:

- структурный хаос семантики языка – “высокая техника, лаконичность и сложность для понимания новичками”; решение: использование структуры RM-ODP и теории типов Б. Расела;
- отсутствие декларативной семантики, противоречивость семантики языка при описании (1) отношений между моделями, построенными с использованием языка, и (2) непосредственно субъектов моделирования; решение: реализация базовой концепции моделирования (Basic Modeling Concept);
- недостаточное теоретическое обоснование используемой метамодели языка UML; решение: предлагаемая в статье концепция моделирования на основе RM-ODP, теории типов Б. Расела и подхода А. Тарского к интерпретации формальных исчислений считается полностью обоснованной.

В многих работах, посвященных формализации модели и метамодели языка UML, рассматривается не сам язык, а некоторый его подязык, формальный и строго структурированный. Так, в [43] рассматриваются BON (Business Object Notation, объектно-ориентированный язык моделирования, по существу совпадающий с подязыком диаграмм классов UML [44]) и PVS (Prototype Verification System, язык спецификаций, разработанный для автоматического анализа метамоделей языков моделирования [45]). Результатом этой работы является полная формальная спецификация метамодели объектно-ориентированного языка моделирования в форме, пригодной для автоматического анализа. Однако BON по сравнению с UML более формализованный и “подогнанный” под условия решаемой задачи. Аналогичный подход использован в [46], где в качестве платформы для формализации выбран формализм Boon, состоящий из метамодели и языка формальных спецификаций Odal – простого строго типизированного языка, семантика которого задана в терминах так называемого λ -исчисления. В [47] на основе [48] рассматривается формализация языка MML (Metamodelling Language), являющегося подмножеством UML; эта формализация предложена авторами в качестве базы для всего UML 2.0. Наконец, в работе [49] демонстрируется применимость алгебраического подхода для формального описания ER-диаграмм (Entity-Relationship diagrams), являющихся аналогами диаграмм классов UML.

Выводы

UML является мощным, гибким средством моделирования, описание стандарта которого является открытым для последующего совершенствования. Неоднозначность как некоторых конструкций самого языка, так и подходов к его формальной семантике, наличие в спецификации неформальных описаний требует дальнейшего развития формальной основы для полной и непротиворечивой интерпретации языка.

Литература

- Booch G. Object-oriented analysis and design with applications. Second edition. – The Benjamin/Cummings Publishing Company Inc., 1994. – 589 p.
- Rumbaugh J., Blacha M., Premerlani W., Eddy F., Lorensen W. Object-Oriented Modeling and Design. – Prentice-Hall Inc., 1991. – 432 p.
- Jacobson I. Object-Oriented Software Engineering. A Use Case Driven Approach. – Addison-Wesley Publishing Company, 1993. – 547 p.
- Object Management Group, 2003. OMG Unified Modeling Language Specification / www.omg.org.
- www.omg.org / 04-10-14pdf.
- www.omg.org / 05-06-06pdf.
- www.omg.org / 05-07-04pdf.
- www.omg.org / 05-06-04pdf.
- Рамбо Д. Тенденции в развитии языка UML и разработки ПО / <http://www.interface.ru/fset.asp?Url=/rational/umltend.htm>.
- <http://www.rational.com/uml>.
- Кумсков М. Rational Rose 98 – CASE-продукт нового поколения / <http://www.interface.ru/public/rose98/rose98.htm>.
- Новичков А. Rational Rose для разработчиков / <http://www.interface.ru/fset.asp?Url=/rational/xros.htm>.
- Booch G., Rumbaugh J., Jacobson I. The Unified Modeling Language User Guide. – MA.: Addison-Wesley Publishing Co., 1999. – 512 p.
- Kruchten P. The Rational Unified Process – An Introduction. – MA.: Addison-Wesley, 2000. – 234 p.
- Jacobson I., Booch G., Rumbaugh J. The Unified Software Development Process. – MA.: Addison-Wesley Publishing Co., 1999. – 512 p.
- Booch G., Jacobson I, Rumbaugh J. The UML specification documents. – Santa Clara, CA.: Rational Software Corp., 1997. (Спецификация размещена на www.rational.com.)
- Fowler M., Scott K. UML Distilled. – MA.: Addison-Wesley, 2000. – 472 p.

- Rumbaugh J., Jacobson I., Booch G. Unified Modeling Language Reference Manual. – MA.: Addison-Wesley Publishing Co., 1999. – 576 p.
- Chonoles M. J., Schardt J.A. UML 2 for Dummies. – Hungry Minds, 2003. – 412 p.
- Kendal S. Fast Track UML 2.0. – Apress, 2004., 416 p.
- Трофимов С. А. Практическая работа в Ration Rose / <http://oad.asf.ru/Files.aspx>
- Вендров А. М.. Объектно-ориентированный анализ и проектирование с использованием языка UML и Rational Rose. Практикум / <http://oad.asf.ru/Files.aspx>
- Boggs W., Boggs M. Mastering UML with Rational Rose. – Sybex Inc., 1999. – 579 p.
- Boggs W., Boggs M. UML with Rational Rose 2.0. – Sybex Inc., 2002. – 702 p.
- Кватрани Т. Визуальное моделирование с помощью Rational Rose и UML: Пер. с англ. – Москва: Издательский дом "Вильямс", 2003. – 192 с.
- Нейбург Э., Максимчук Р. Проектирование баз данных с помощью UML.: Пер. с англ. – Москва: Издательский дом "Вильямс", 2002. – 288 с.
- Гома Х. UML. Проектирование систем реального времени, распределенных и параллельных приложений.: Пер. с англ. – Москва: ДМК, 2002. – 704 с.
- Розенберг Д., Скотт К. Применение объектно-ориентированного моделирования с использованием UML и анализ прецедентов.: Пер. с англ. – Москва: ДМК, 2002. – 160 с.
- Ларман К. Применение UML и шаблонов проектирования. 2-е издание.: Пер.с англ. – Москва: Издательский дом "Вильямс". 2004. – 604 с.
- Gamma E., Helm R., Jonson R., Vissides J. Element of Reusable Object-Orient Software. – Addison Wesley Longman Inc., 1994. – 372 p.
- Коберн А. Современные методы описания функциональных требований к системам.: Пер. с англ. – Москва: ЛОРИ, 2002. – 262 с.
- Лэффингуэлл Д., Уидриг Д. Принципы работы с требованиями к программному обеспечению. Унифицированный подход.: Пер. с англ. – Москва: "Вильямс", 2002. – 448 с.
- Duffy Daniel J. Domain architectures: models and architectures for UML applications. – John Wiley & Sons Ltd, 2004. – 390 p.
- Мюллер Р. Базы данных и UML. Проектирование.: Пер. с англ. – Москва: ЛОРИ, 2002. – 420 с.
- Nock C. Data Access Patterns: Database Interactions in Object-Oriented Applications. – Addison Wesley, 2003. – 512 p.
- Eriksson H.-Erik, Penker M. Business Modeling with UML: Business Patterns at Work. – Wiley Computer Publishing, 2000. – 274 p.
- Маклаков С. В. Создание информационных систем с AllFusion Modeling Suite. – Москва: Диалог-МИФИ, 2005. – 432 с.
- Husman H. Loose Semantics for UML/OCL // Society for Design and Process Science, 2002. – P. 32-39.
- Genova G., Llorens J., Quintana V. Digging into Use Case Relationships // Lect. Notes Comput. Sci. – 2002. – V. 2460. – P. 115–127.
- Gogolla M., Henderson-Sellera B. Analysis of UML Stereotypes within the UML Metamodel // Lect. Notes Comput. Sci. – 2002. – V. 2460. – P. 84-99.
- Naumenko A., Wegmann A. A Metamodel for the Unified Modeling Language // Lect. Notes Comput. Sci. – 2002. – V. 2460. – P. 2-17.
- RM-ODP Open Distributed Processing – Reference Model - ISO, ITU.: ISO/IEC 10746-1, 2, 3, 4 | ITU-T Recommendation X.901, X.902, X.903, X.904, 1995-1998.
- Paige R., Ostroff J. Metamodelling and Conformance Checking with PVS // Lect. Notes Comp. Sci. –2001. –V.2029. –P.2-16.
- Walden K., Nerson J.-M. Seamless Object-Oriented Software Development. – Prentice-Hall, 1995.
- Owre S., Shankar N., Rushby J., Stringer-Calvert D. The PVS Language Reference Version 2.3 – September, 1999. – (Tech. Rep. / SRI International Technical Report).
- Overgaard G. Formal Specification of OO Modeling // Lect. Notes Comput. Sci. – 2000. – V. 1783. – P. 193-207.
- Clark T., Evans A., Kent S. The Metamodelling Language Calculus: Foundation Semantics for UML // Lect. Notes Comput. Sci. – 2001. – V. 2029. – P. 17-31.
- Cardeli L, Abadi M. A theory of Objects. – Springer-Verlag, 1996.
- Lellahi K. Conceptual Data Modeling: An Algebraic Viewpoint // Lect. Notes Comput. Sci. – 2001. – V. 2244. – P. 336-348.

Информация об авторах

Буй Дмитрий Борисович – заведующий лабораторией проблем программирования, Киевский национальный университет имени Тараса Шевченко, факультет кибернетики: Украина, Киев, 03680, пр. Глушкова 2, корп.6; e-mail: buy@unicyb.kiev.ua

Шишацкая Елена Владимировна – инженер-программист лаборатории проблем программирования, Киевский национальный университет имени Тараса Шевченко, факультет кибернетики: Украина, Киев, 03680, пр. Глушкова 2, корп.6; e-mail: shyshatskaja@unicyb.kiev.ua

МОДЕЛЬ “СУЩНОСТЬ-СВЯЗЬ”: РОЛИ, СИЛЬНЫЕ И СЛАБЫЕ ТИПЫ СУЩНОСТЕЙ И ТИПЫ СВЯЗЕЙ

Дмитрий Буй, Людмила Сильвейструк

Аннотация: Рассматриваются и формализуются такие понятия модели “сущность-связь”: роли, сильные и слабые типы сущностей, сильные и слабые типы связей. Раскрыты принципы построения диаграммы сущностей и связей.

Ключевые слова: модель “сущность-связь”, сильный тип сущности, слабый тип сущности, сильный тип связи, слабый тип связи, роль, диаграмма сущностей и связей.

ACM Classification Keywords: E.4 Coding and information theory – Formal models of communication

Введение

В работе продолжается рассмотрение и формализация модели “сущность-связь” (entity-relationship model). В работах [Буй Д.Б., Сильвейструк Л.М., 2006 а; Буй Д.Б., Сильвейструк Л.М., 2006 б] были уточнены сущности, связи и структурные ограничения связей. В этой статье рассматриваются такие понятия: роли, сильные и слабые типы сущностей, сильные и слабые типы связей. Также рассматривается графическая нотация Чена диаграммы сущностей и связей (entity-relationship diagram).

Роли

Семантическое значение вхождения типа сущности в тип связи называется ее ролью. Вхождениям типов сущностей в тип связи можно присывать ролевые имена – для уточнения назначения каждого вхождения (более неформально: для уточнения назначения каждой сущности-участницы в связи). Здесь надо говорить именно о вхождении, так как n -арный тип связи, где $n = 2, 3, \dots$, задается на n необязательно разных типах сущностей. Особое значение имеют ролевые имена в типах связей на одном типе сущности. В этом случае нет других средств кроме ролевых имен для идентификации разных вхождений одного типа сущности в тип связи. Таким образом, смысл ролевых имен заключается в синтаксическом различении различных вхождений типа сущности в тип связи.

Серьезно говорить о семантическом аспекте ролевых имен, то есть об определении функций участников связи ролевыми именами, не приходится, поскольку вся “семантика” заключается в выборе “мнемонических” ролевых имен, которые и должны отображать семантику.

Можно сказать, что понятие роли часто вспомогательное, оно существует для удобства пользователя; поэтому роли (ролевые имена) можно даже не вводить, особенно в случае, когда явное введение этих имен ничего не прибавляет к пониманию модели.

Ролевые имена, которые будем обозначать через v_1, v_2, \dots , дают возможность уточнить тип связи (с минимальной интерпретацией).

Типом связи арности n между типами сущностей (который соединяет типы сущностей) E_1, \dots, E_k , где $k \leq n$, и типы сущностей E_1, \dots, E_k попарно разные, назовем сюръективное отображение вида $R: \{v_1, \dots, v_n\} \rightarrow \{E_1, \dots, E_k\}$, где ролевые имена v_1, \dots, v_n попарно разные.

Ролевые имена дают возможность сделать следующий шаг в уточнении типа связи, интерпретируя его не как отношение [Буй Д.Б., Сильвейструк Л.М., 2006 а], а как таблицу [Редько В.Н., Брона Ю.И., Буй Д.Б., Поляков С.А., 2001], схема которой совпадает с множеством ролевых имен типа связи.

Слабые и сильные типы сущностей, слабые и сильные типы связей

Иногда сам факт существования сущностей некоторого типа сущности зависит от особой связи этих сущностей с сущностями других типов сущностей. То есть можно сказать, что тип связи вводит тип сущности. Дадим содержательное объяснение.

Слабым типом сущности (weak entity type) называют такой тип сущности, существование которого зависит от других типов сущностей. Тип связи, который соединяет данные типы сущностей, называется *слабым типом связи (weak relationship type)*, то есть это такой тип связи, который вводит в рассмотрение слабый тип сущности.

Вообще любой тип связи можно использовать в качестве слабого типа связи [Чен П., 1995], но мы наложим на (n -арний) слабый тип связи такие два ограничения:

- в слабый тип связи может входить только один слабый тип сущности, который вводится этим слабым типом связи (и детализируется другими типами сущностей, которые входят в этот слабый тип связи);
- степень участия слабого типа сущности, который вводится слабым типом связи, полная (то есть сущности слабого типа сущности могут рассматриваться, только находясь в связи).

Уточняя первое ограничение скажем, что *слабый тип сущности E вводится слабым типом связи R* , если R имеет вид $R: \{v_1, \dots, v_n\} \rightarrow \{E_1, \dots, E_k\}$, где $E \in \{E_1, \dots, E_k\}$, причем $|R^{-1}(\{E\})| = 1$.

Сильным типом сущности (strong entity type) называют такой тип сущности, существование которого не зависит от других типов сущностей.

Тип связи, который не является слабым типом связи, называется *сильным типом связи (strong relationship type)*.

Сильный тип связи еще называют регулярным отношением сущности (regular entity relation) [Чен П., 1995], а слабый тип связи – поддерживающей связью (supporting relationships) [Гарсиа-Молина Г., Ульман Дж., Уидом Дж., 2004, гл. 2, с. 82-83].

Сильный тип сущности называют еще родительским (parent), сущностью-владельцем (owner), доминантным (dominant) или регулярным (regular); слабый тип сущности – дочерним (child), зависимым (dependent) или подчиненным (subordinate) [Коннолли Т., Бегг К., Страчан А., 2000, часть II, гл. 5, с. 185].

Отметим, что понятие слабого типа сущности имеет несколько интерпретаций (например, [6, гл. 3]). Это обусловлено тем, что некоторые проектировщики баз данных под слабым типом сущности понимают тип сущности, присутствие которого в базе данных зависит от других типов сущностей. Другие же считают, что слабым типом сущности является тип сущности, который логически зависит от других типов сущностей (то есть, вводится некоторым типом связи). Понятно, что вторая трактовка, которого мы придерживаемся, является более узкой.

Чтобы проиллюстрировать отличие этих двух подходов, рассмотрим два примера [Крэнке Д., 2003, гл. 3, с. 90]. Первый пример – это два типа сущностей Руководитель и Студент, причем по правилам учебного

заведения каждый студент должен иметь руководителя. Следуя первому подходу (к определению слабого типа сущности) тип сущности Студент – слабый тип сущности, но согласно второго подхода, если сущность типа сущности Студент может существовать без связи с некоторой сущностью типа сущности Руководитель, то тип сущности Студент – также сильный тип сущности.

Второй пример – три типа сущностей Пациент, Врач и Рецепт. Естественно, рецепт выдается конкретному пациенту некоторым существующим врачом, то есть тип сущности Рецепт вводится определенным слабым типом связи, который включает сильные типы сущностей Пациент и Врач. Итак, тип сущности Рецепт является слабым типом сущности как при первом, так и при втором подходах.

Есть и третий содержательный подход, который отражает ситуацию, когда сущности некоторого типа сущности невозможно уникально идентифицировать значениями их собственных атрибутов; значит, для их идентификации надо использовать слабый тип связи, который связывает данный тип сущности с другими типами сущностей (который вводит этот слабый тип сущности); причем атрибуты этих типов сущностей (сильных типов в данном контексте) будут входить в состав первичного ключа слабого типа сущности, который вводится.

Этот тип сущности является особым видом слабых типов сущностей и называется *идентификационно-зависимым типом сущности (ID-dependent entity type)*; то есть это такой тип сущности, идентификатор сущностей (первичный ключ) которого содержит идентификаторы (первичные ключи) других сущностей. В данном случае можно сказать, что идентификационно-зависимый тип сущности зависит от других типов сущностей как логически, так и физически (поскольку понятие ключа можно относить к этим двум уровням). Заметим, что некоторые проектировщики баз данных ограничивают слабые типы сущностей именно идентификационно-зависимыми типами сущностей. Понятно, что последняя интерпретация слабых типов сущностей есть самой узкой из трех вышеприведенных.

Рассмотрим идентификационно-зависимый тип сущности на примере двух типов сущностей Квартира и Дом [Крёнке Д., 2003, гл. 3, с. 89]. Пусть первичным ключом (идентификатором) типа сущности Дом будет атрибут НазваниеДома. Для того, чтобы идентифицировать сущности типа сущности Квартира, необходимо знать не только номер квартиры и ее внутренние характеристики, но и дом, в котором она находится; значит, идентификатором типа сущности Квартира будут атрибуты НазваниеДома и НомерКвартиры. Поскольку идентификатор типа сущности Квартира включает в себя идентификатор типа сущности Дом (а именно НазваниеДома), то тип сущности Квартира есть идентификационно-зависимым от типа сущности Дом.

Возможна ситуация, когда некоторые типы сущностей, от которых зависит существование слабого типа сущности, входят в слабый тип связи несколько раз в разных ролях.

Например, слабый тип связи Контракт, который вводит слабый тип сущности Контракты, включает (соединяет) сильные типы сущностей Актеры, Кинофильмы, Студии [Гарсиа-Молина Г., Ульман Дж., Уидом Дж., 2004, гл. 2, с. 82]. Связь отражает, в частности, факт составления контракта между киностудией, которая владеет правами на съемку кинофильма, и другой киностудией, которая разрешает актеру принимать участие в работе над данным фильмом (так как актер является штатным сотрудником этой киностудии). Как видим, тип сущности Студии входит в слабый тип связи Контракт в двух ролях – “студия-актер” (студия, в штате которой находится актер) и “студия-фильм” (студия, которая снимает фильм). Таким образом, слабый тип сущности Контракты зависит от существования типов сущностей Актеры, Кинофильмы и Студии, причем тип сущности Студии детализирует слабый тип сущности два раза.

Слабый тип связи можно изобразить с помощью ациклического ориентированного графа (точнее дерева единичной глубины), где множество вершин – множество типов сущностей, которые являются участниками слабого типа связи, множество дуг – множество пар вершин (началом дуги будет вершина, которая отвечает слабому типу сущности, концом дуги – вершина, которая отвечает сильному типу

сущности). При данной интерпретации случай, когда сильный тип сущности входит в данный слабый тип связи несколько раз в разных ролях, мы упрощаем и изображаем в виде одной дуги.

На рис. 1 показан граф, изображающий слабый тип сущности E , существование которого зависит от сильных типов сущностей F_1, F_2, \dots, F_n , и который (слабый тип) вводится слабым типом связи R .

Следует отметить, что понятия сильного и слабого типа сущностей относительные, если их рассматривать для отдельных слабых типов связи (локальное рассмотрение). Так, один тип сущности в одном слабом типе связи может быть слабым типом сущности, а в другом – сильным типом сущности.

Рассмотрим введенные понятия для модели (глобальное рассмотрение). Как известно, модель – это множество типов сущностей с заданными на них типами связей. Все типы связи модели разделяются на слабые типы связи и сильные типы связи. Как отмечалось выше, слабые типы связи предназначены для введения слабых типов сущностей.

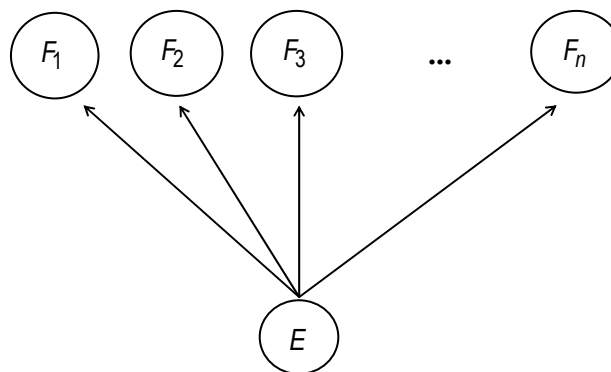


Рис. 1. Граф слабого типа связи

Слабый тип сущности в модели – это такой

тип сущности, который вводится с помощью некоторого слабого типа связи модели. Заметим, что слабый тип сущности может входить в единственный слабый тип связи, который его вводит (и принадлежит модели).

Сильный тип сущности в модели – это такой тип сущности, который не является слабым типом сущности в модели.

Рассмотрим модель по слабым типам связей; то есть модель, которая состоит из всех типов сущностей модели и слабых типов связей, которые вводят соответствующие слабые типы сущностей. Выше была показана локальная графовая интерпретация слабого типа связи (то есть для отдельно взятого слабого типа сущности, соответствующих сильных типов сущностей и слабого типа связи, который вводит слабый тип сущности). Аналогично можно изобразить и модель по слабым типам связей (рис. 2), которая получается путем объединения графов слабых типов связи модели и изолированных вершин, интерпретирующие сильные типы сущностей в модели, которые не являются участниками слабых типов связи данной модели.

Анализируя данный ориентированный граф модели, можно выделить такие его особенности:

- множество вершин графа совпадает с множеством всех типов сущностей модели;
- множество дуг состоит из множества пар вершин, где началом дуги будет вершина, которая отвечает слабому типу сущности, а концом дуги – вершина, которая отвечает сильному типу сущности в слабом типе связи, который вводит слабый тип сущности;
- вершина, которая отвечает сильному типу сущности в модели, или изолированная (то есть степень вершины равняется 0), или тупиковая (то есть ей инцидентны только входящие дуги);
- вершина, которая отвечает слабому типу сущности в модели, инцидентна исходящим дугам, которые соединяют ее с соответствующими вершинами типов сущностей, от которых зависит существование данного слабого типа сущности.

При построении модели базы данных проектировщики сначала определяют типы сущностей и типы связей. Для проверки правильности построения слабых, сильных типов сущностей и соответствующих слабых типов связей, рассматривают соответствующую модель по слабым типам связей.

Данная модель будет корректно построенной (назовем ее *W-корректной* от англ. weak – слабый), если выполняются следующие условия:

- для любого слабого типа сущности в модели существует единственный слабый тип связи, который его вводит;
- модель по слабым сущностям представляется ациклическим ориентированным графом.

Уточним требования к корректности модели в терминах частичного порядка. Для этого рассмотрим множество M всех типов

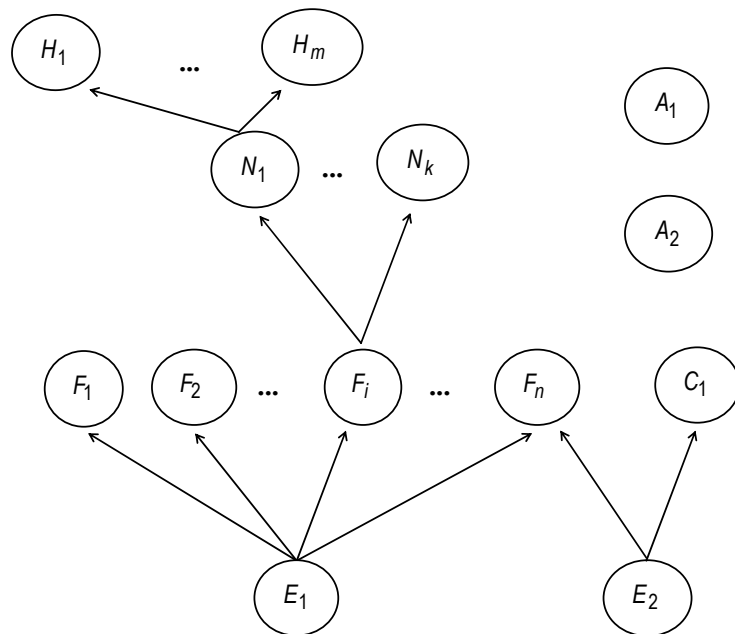


Рис. 2. Граф модели по слабым типам связей

сущностей этой модели. На этом множестве зададим бинарное отношение $<$: $E_1 < E_2 \stackrel{def}{\Leftrightarrow}$ существует слабый тип связи модели, который вводит слабый тип сущности E_1 , причем тип сущности E_2 принимает участие в этом типе связи и отличен от E_1 .

Рефлексивно-транзитивное замыкание отношения $<$ на множестве M обозначим через \leq^* (то есть $\leq^* \stackrel{def}{=} \bigcup_{i=0,1,2,\dots} <^i$, где $<^0 \stackrel{def}{=} \Delta_M$ – диагональ на множестве M , $<^{i+1} \stackrel{def}{=} < \circ <^i$, \circ – композиция

бинарных отношений).

В графе модели по слабым типам связей (рис. 2) каждая дуга отвечает парам бинарного отношения $<$.

Утверждение. Модель будет *W-корректной* тогда и только тогда, когда отношение \leq^* является частичным порядком. Тип сущности является сильным типом сущности в модели тогда и только тогда, когда этот тип является максимальным элементом частично упорядоченного множества M . Тип сущности является слабым типом сущности в модели тогда и только тогда, когда этот тип не является максимальным элементом частично упорядоченного множества M . □

Доказательства первого и второго утверждений непосредственно вытекают из определений, третье утверждение вытекает из второго. □

Диаграмма сущностей и связей

ER-проект состоит из документации к проекту, в которой отображены основные объекты и их свойства, и схемы базы данных (графического изображения).

Схема базы данных в ER-модели – это графическое представление совокупности типов сущностей с определенным для каждого из них конечным непустым множеством атрибутов и совокупности типов связей, для каждого из которых определено конечное (возможно, пустое) множество атрибутов.

Данную схему еще называют диаграммой сущностей и связей (entity-relationship diagram) или ER-диаграммой (ER-diagram). Такие диаграммы стандартизированы, но не весьма жестко; существует несколько систем обозначений, например, обозначения Чена [Коннолли Т., Бегг К., Страчан А., 2000, часть VIII, приложение Д], система обозначений с использованием значка "воронья лапка" [Коннолли Т., Бегг К., Страчан А., 2000, часть VIII, приложение Д] и система обозначений UML [Крэнке Д., 2003, гл. 3].

Рассмотрим систему обозначений Чена. Для задания принадлежности элемента к определенному виду используется специальная геометрическая фигура:

- прямоугольник – для типа сущности;
- эллипс – для атрибута;
- ромб – для типа связи.

Имена атрибутов, типов сущностей и типов связей указывают внутри соответствующих геометрических фигур.

Если тип сущности является слабым типом сущности, его принято изображать в виде двойного прямоугольника.

Сильный тип связи обозначается ромбом, а слабый тип связи – ромбом с двойным контуром. Для представления типов-участников типов связей используются линии, которые соединяют соответствующие типы сущностей с типом связи. Вид типа связи обозначается над линиями в виде соответствующих надписей возле типов сущностей. Например, если это вид бинарной связи "один ко многим", то делают надписи 1, М, соответственно, возле соответствующих типов сущностей.

Участники типа связи с полным участием соединяются со знаком типа связи двойной линией, а участники типа связи с частичным участием – одинарной линией. Можно обозначать структурные ограничения, которые накладываются на тип связи, в виде надписей (max, min) над линиями соединения возле соответствующих типов сущностей.

Линии, которые соединяют тип связи с типами сущностей, можно обозначать текстовыми метками, которые описывают определенные роли соответствующих типов сущностей. Также любые объяснения к элементам модели можно делать в виде текстовых надписей.

Выводы

Разработка моделей данных в рамках модели "сущность-связь" значительно упростилась в последние годы, поскольку теперь инструменты для построения ER-диаграмм входят в состав многих популярных CASE-средств. К таким продуктам относятся ERwin [Маклаков С.В., 1999, Маклаков С.В., 2005], Rational Rose [www.interface.ru/rational/rasdiag.htm, www.interface.ru/rational/oocase.htm] и CASE Studio [www.casestudio.com]. ER-диаграммы, созданные с помощью CASE-средств, как правило имеют более "красивый" вид, эти диаграммы проще изменять и адаптировать. Данные средства используют разные нотации для создания ER-диаграмм, некоторые из нотаций появились именно благодаря CASE-средствам.

Основное задание последующей работы – рассмотрение и формализация расширенной модели "сущность-связь".

Литература

[Буй Д.Б., Сильвейструк Л.М., 2006 а] Модель "сутність-зв'язок": формалізація сутностей та зв'язків // Вісник Київського університету. Серія: фіз.-мат. науки. – 2006. – Вип. 3. – С. 143-152.

[Буй Д., Сильвейструк Л., 2006 b] Формализация структурных ограничений в модели "сущность-связь" // Proceedings of the XII-th International Conference "Knowledge-Dialogue-Solution" (June 20-25, 2006, Varna, Bulgaria). – Sofia. – 2006. – P. 223-229.

- [Гарсиа-Молина Г., Ульман Дж., Уидом Дж., 2004]. Системы баз данных. Полный курс.: пер. с англ. – Москва: Издательский дом “Вильямс”, 2004. – 1088 с.
- [Коннолли Т., Бегг К., Страчан А., 2000]. Базы данных: проектирование, реализация и сопровождение. Теория и практика, 2-е изд.: пер. с англ. – Москва: Издательский дом “Вильямс”, 2003. – 1440 с.
- [Крёнке Д., 2003]. Теория и практика построения баз данных. 8-е изд. – Санкт-Петербург: “Питер”, 2003. – 800 с.
- [Маклаков С.В., 1999] VPwin и ERwin. CASE-средства разработки информационных систем. – Москва: “ДИАЛОГ-МИФИ”, 1999. – 256 с.
- [Маклаков С.В., 2005] Создание информационных систем с AllFusion Modeling Suite. – Москва: “ДИАЛОГ-МИФИ”, 2005. – 432 с.
- [Редько В.Н., Брона Ю.Й., Буй Д.Б., Поляков С.А., 2001] Реляційні бази даних: табличні алгебри та SQL-подібні мови. – Київ: Видавничий дім “Академперіодика”, 2001. – 198 с.
- [Чен П., 1995] Модель “сущность-связь” – шаг к единому представлению о данных // Системы управления базами данных. – 1995. – № 3. <http://www.osp.ru/dbms/1995/03/271.htm>.
- www.interface.ru/rational/rasdiag.htm
- www.interface.ru/rational/oocase.htm
- www.casestudio.com

Информация об авторах

Буй Дмитрий – заведующий лабораторией проблем программирования, Киевский национальный университет имени Тараса Шевченко, факультет кибернетики: Украина, Киев, 03680, пр. Глушкова 2, корп.6; e-mail: buy@unicyb.kiev.ua

Сильвейструк Людмила – аспирантка, Киевский национальный университет имени Тараса Шевченко, факультет кибернетики: Украина, Киев, 03680, пр. Глушкова 2, корп.6; e-mail: slm-klm@rambler.ru

MULTIALGEBRAIC STRUCTURES EXISTENCE FOR GRANULAR COMPUTING

Alexander Kagramanyan, Vladimir Mashtalir, Vladislav Shlyakhov

Abstract: *In different fields a conception of granules is applied both as a group of elements defined by internal properties and as something inseparable whole reflecting external properties. Granular computing may be interpreted in terms of abstraction, generalization, clustering, levels of abstraction, levels of detail, and so on. We have proposed to use multialgebraic systems as a mathematical tool for synthesis and analysis of granules and granule structures. The theorem of necessary and sufficient conditions for multialgebraic systems existence has been proved.*

Keywords: *granular computing, multirelations, multioperations.*

ACM Classification Keywords: *I.2.4 Knowledge representation formalisms and methods: relation systems.*

Introduction

Granular computing explores knowledge from different standpoints to reveal various types of structures and information embedded in the data [Zadeh, 1997, Bargiela, Pedrycz, 2002]. A paradigm of granular computing consists in grouping elements together (in a granule) by indistinguishability, similarity, proximity or functionality in arbitrary feature or signal spaces. Taking into account a semantic interpretation of why two objects are put into the same granule and how two objects are related with each other it provides one of a general methodology for intelligent data analysis on different levels of roughening or detailing [Pal et al., 2005, Yao, Yao, 2002].

Internal, external and contextual properties of granules, collective structure of a family of granules and hierarchical structure of granules represent a possible foundation for qualitative/quantitative characterization of levels of abstraction, detail, control, explanation, difficulty, organization and so on. Focusing on high conceptual level issues by ignoring much irrelevant details, granular computing are actively used in computational intelligence [Doherty et al., 2003], information granulation based on rough sets [Yao, 2001, Pal et al., 2005], data mining [Yao, 2006], interval analysis, cluster analysis, machine learning and many others [Yager, 2002, Lin, 2003, etc.]. The integration of multiple views on different types of granulation and granular structure may provide more useful data analysis tools [Lin, 2003, Yao, 2005]. One of a number of possible approaches is to use multialgebraic systems [Mashtalir, Shlyakhov, 2003] as mathematical apparatus for synthesis and analysis of granules and granule structures.

Thus, we need tools providing a granular linkage, i.e. formal operations and relations determined on granules. Furthermore, this linkage has to be induced either by information embedded in the data or by given close coupling with field of application. These questions are at present far from being solved. But the important point to note here is the search of necessary and sufficient conditions for existence of multialgebraic systems as enough general tool of granular computing.

Motivation of granular computing modeling by multialgebraic systems

If we choose any natural number $p \in \mathbb{N}$ then we can consider a ternary relation

$$E(n_1, n_2, n_3) = \begin{cases} 1, & (n_1 + n_2) \bmod p = n_3 \bmod p; \\ 0, & (n_1 + n_2) \bmod p \neq n_3 \bmod p \end{cases} \quad (1)$$

where $a \bmod b$ defines a as modulo b residue, i.e. $a \bmod b \triangleq a - \lfloor a/b \rfloor \times b$ and $\lfloor \cdot \rfloor$ is a floor function. It is easily seen, if we hold fixed $k \in \{1, 2, 3\}$ then we get an equivalence relation P_k^E , e.g. for $k = 1$

$$P_1^E(n_1, n'_1) = 1 \Leftrightarrow E(n_1, n_2, n_3) \equiv E(n'_1, n_2, n_3).$$

This equivalence partitions set of natural numbers into residue classes modulo p . Indeed, if remainders in division n_1 and n'_1 by p are the same then for arbitrary n_2 and n_3 continued equality is

$$\begin{aligned} (n_1 + n_2) \bmod p &= (ps_1 + r_1 + ps_2 + r_2) \bmod p = (r_1 + r_2) \bmod p = \\ &= (ps'_1 + r_1 + ps_2 + r_2) \bmod p = (n'_1 + n_2) \bmod p. \end{aligned} \quad (2)$$

Here it is implied that $n_1 = ps_1 + r_1$, $n'_1 = ps'_1 + r_1$, $n_2 = ps_2 + r_2$, and $s_1, s'_1, s_2 \in \mathbb{N}$. From (2) it follows that $E(n_1, n_2, n_3) \equiv E(n'_1, n_2, n_3)$. The converse proposition is the valid one also. If $E(n_1, n_2, n_3) \equiv E(n'_1, n_2, n_3)$ then remainders in division n_1 and n'_1 by p have to be equal, if not when $n_1 = ps_1 + r_1$, $n'_1 = ps_1 + r'_1$ and $r_1 \neq r'_1$ under $n_2 = 0$, $n_3 = r_1$ we obtain, on the one hand,

$$(n_1 + n_2) \bmod p = (ps_1 + r_1) \bmod p = r_1 \bmod p = n_3 \bmod p,$$

i.e. $E(n_1, 0, r_1) = 1$. On the other hand,

$$(n'_1 + n_2) \bmod p = (ps'_1 + r_1) \bmod p = r'_1 \bmod p \neq n_3 \bmod p.$$

Since $r'_1, r_1 \leq p$ и $r'_1 \neq r_1$ then $E(n'_1, 0, r_1) = 0$, which contradicts the original assumption. Notice, from (1) we get $P_1^E = P_2^E = P_3^E$.

Let us sum up. The carrier of original relation is the set of natural numbers \mathbb{N} but the induced equivalence demonstrates the significant carrier change: we have got a finite set $\Pi = \{0, 1, \dots, p-1\}$.

New relation, which will be named a multirelation and denoted by E^M in the sequel, is generated on new carrier. As before it is ternary relation but the domain is Π^3 instead of \mathbb{N}^3 and multirelation E^M acquires the new

property that can be expressed as an operation

$$r_1 \oplus r_2 = r_3 \Leftrightarrow E_m(r_1, r_2, r_3) = 1 \Leftrightarrow E(n_1, n_2, n_3) = 1$$

where sign “ \oplus ” denotes p congruence addition and $n_i = ps_i + r_i, i = 1, 2, 3, r_i \in \{0, 1, \dots, p - 1\}$. Operations on equivalence classes here and subsequently will be denoted by F^M .

If they follow terminology of algebraic system a triplet $\langle A, R, Q \rangle$ (here A is arbitrary set (carrier), R is a relation suite, Q is a set of operations) is called a model on conditions that $Q = \emptyset$ and it is said to be an algebra if $R = \emptyset$. Consequently, from the model $\langle \mathbb{N}, E, \emptyset \rangle$ we pass on to the algebra $\langle \Pi, \emptyset, \oplus \rangle$ whose carrier is well-known algebraic structure viz a cyclic Abelian group of p -th order.

It is necessary to understand that original carrier can represent a set and carrier of induced relation on equivalence classes can be Cartesian product of different sets. Let us consider one more example

$$E(n_1, n_2, n_3) = \begin{cases} 1, & (n_1 + n_2) \bmod p_1 = n_3 \bmod p_2; \\ 0, & (n_1 + n_2) \bmod p_1 \neq n_3 \bmod p_2 \end{cases} \quad (3)$$

where $p_1 \neq p_2$. It should be noted that $P_1^E = P_2^E \neq P_3^E$, i.e. E_m is Cartesian product $A \times B$ where $A = \{0, 1, \dots, p_1 - 1\}$, $B = \{0, 1, \dots, p_2 - 1\}$. As may be seen from (3) the multirelation E^M as a ternary relation is defined on $A^2 \times B$ and represents an operation from A^2 into B . There is no difficulty in understanding that under certain p_1 and p_2 not only an equivalence inequality appears but a level of partition detail and equivalence nesting are changed. For instance, if $p_1 = 4, p_2 = 2$ then $P_1^E = P_2^E \subseteq P_3^E$ as \mathbb{N} is partitioned into 4 classes corresponding to residues of division $A = \{0, 1, 2, 3\}$ at the expense of $P_1^E = P_2^E$. Equivalence P_3^E partitions original set \mathbb{N} into 2 classes from even and odd numbers, i.e. $B = \{0, 1\}$. In this connection classes $\{0, 2\}$ belong to the set of even numbers, classes $\{1, 3\}$ form part of odd numbers set respectively.

In analyzed examples the original relation E induces the multirelation (more precisely the multioperation), i.e. an operation with ranges of definition as equivalence classes. It may seem that a similar situation is observed all the time, however this is by no means always the case. Consider the binary relation E (tab. 1) which is defined on the Cartesian product $\{1, 2, 3, 4, 5\} \times \{a_1, a_2, b_1, c_1, c_2, c_3, c_4, c_5, c_6\}$.

Table 1

		A		B	C					
		a_1	a_2	b_1	c_1	c_2	c_3	c_4	c_5	c_6
Π_I	1	0	0	1	1	1	1	1	1	1
	2	0	0	1	1	1	1	1	1	1
Π_{II}	3	1	1	1	0	0	0	0	0	0
	4	1	1	1	0	0	0	0	0	0
	5	1	1	1	0	0	0	0	0	0

It should be clear that the induced equivalences P_1^E and P_2^E dissect the first set $A_1 = \{1, 2, 3, 4, 5\}$ into 2 classes: $\Pi_I = \{1, 2\}$, $\Pi_{II} = \{3, 4, 5\}$ and the second one $A_2 = \{a_1, a_2, b_1, c_1, c_2, c_3, c_4, c_5, c_6\}$ into 3 classes: $A = \{a_1, a_2\}$, $B = \{b_1\}$, $C = \{c_1, c_2, c_3, c_4, c_5, c_6\}$. Thus, the multirelation E_m is defined on the Cartesian product of induced equivalence classes, i.e. $\{\Pi_I, \Pi_{II}\} \times \{A, B, C\}$ (tab. 2).

Table 2

	A	B	C
Π_I	0	1	1
Π_{II}	1	1	0

This multirelation can be represented as two explicit mappings associating $\{\Pi_I, \Pi_{II}\}$ with $\{A, B, C\}$. Denote induced mappings as F^{EM} and $(F^{EM})^{-1}$ in both directions then $F^{EM}(\Pi_I) = \{B, C\}$, $F^{EM}(\Pi_{II}) = \{A, B\}$ and $(F^{EM})^{-1}(A) = \{\Pi_{II}\}$, $(F^{EM})^{-1}(B) = \{\Pi_I, \Pi_{II}\}$, $(F^{EM})^{-1}(C) = \{\Pi_I\}$. Single-valuedness is lacking in both cases therefore we are not able to indicate multioperation.

Thus, algebraic model can lead either to multimodels or to multialgebra and there arises an important question: when do two relations with different arities generate one carrier?

Necessary and sufficient conditions for multirelations carriers equality

Let A_1, A_2, \dots, A_n be any given sets. Consider n -arity relation $E(x_1, \dots, x_n)$ on Cartesian product of arbitrary carriers $A_1 \times \dots \times A_n$. A trivial verification shows that

$$P_k^E(x_k, x'_k) = 1 \Leftrightarrow E(x_1, \dots, x_k, \dots, x_n) \equiv E(x_1, \dots, x'_k, \dots, x_n) \tag{4}$$

constitutes an equivalence relation and partitions may be regarded on each A_k . The understanding of the appearance mechanism of P_k^E awaits further investigation.

If $A_k = A_l$ then relations P_k^E and P_l^E can be compared. For instance,

$$[P_k^E(x_k, x'_k) = 1 \Rightarrow P_l^E(x_k, x'_k) = 1] \Leftrightarrow P_k^E \subseteq P_l^E,$$

i.e. P_k^E fulfills more detail partition than P_l^E and information can be analyzed with greater exactness. Using terminology of relation $E(x_1, \dots, x_n)$ we get in that case

$$\begin{aligned} E(x_1, \dots, x_{k-1}, x_k, x_{k+1}, \dots, x_n) \equiv E(x_1, \dots, x_{k-1}, x'_k, x_{k+1}, \dots, x_n) &\Rightarrow \\ \Rightarrow E(x_1, \dots, x_{l-1}, x_k, x_{l+1}, \dots, x_n) \equiv E(x_1, \dots, x_{l-1}, x'_k, x_{l+1}, \dots, x_n). \end{aligned}$$

Generally, on $\{1, 2, \dots, n\}$ the relation $E(x_1, \dots, x_n)$ produces the four-valued indicator function

$$f(k, l) = \begin{cases} -1, & A_k = A_l, P_k^E \subseteq P_l^E; \\ 0, & A_k \neq A_l, P_k^E \not\parallel P_l^E; \\ 1, & A_k = A_l, P_k^E = P_l^E; \\ 2, & A_k = A_l, P_l^E \subseteq P_k^E \end{cases}$$

where symbol " $\not\parallel$ " denotes relation incomparability. Let us introduce notations

$$\begin{aligned} X &= E(x_1, \dots, x_{k-1}, x_k, x_{k+1}, \dots, x_n) \equiv E(x_1, \dots, x_{k-1}, x'_k, x_{k+1}, \dots, x_n), \\ Y &= E(x_1, \dots, x_{l-1}, x_k, x_{l+1}, \dots, x_n) \equiv E(x_1, \dots, x_{l-1}, x'_k, x_{l+1}, \dots, x_n) \end{aligned}$$

then it leads to the following sufficiently clear statement.

Proposition 1. For arbitrary n -arity relation $E(x_1, \dots, x_n)$ values of the indicator function $f(k, l)$ are specified by conditions

- if $X \Leftrightarrow Y$ then $f(k, l) = 1$,
- if $X \Rightarrow Y$ then $f(k, l) = -1$,
- if $X \Leftarrow Y$ then $f(k, l) = 2$,
- if $X \not\parallel Y$ then $f(k, l) = 0$.

Definition 1. Arbitrary n -arity relation $E(x_1, \dots, x_n)$ is said to be internally (k, l) -coherent if and only if $A_k = A_l$

and $P_k^E = P_l^E$.

It is reasonable to mention that equivalence relation

$$V_E(k,l) = 1 \Leftrightarrow A_k = A_l, P_k^E = P_l^E \quad (f(k,l) = 1)$$

is induced on $\{1,2,\dots,n\}$. This relation can be expressed as matrix of internal coherence $\Phi(E) = (V_E(k,l))$.

Proposition 2. Under corresponding renumbering of n -arity relation $E(x_1,\dots,x_n)$ arguments, the matrix of internal coherence $\Phi(E)$ can be represented as block-diagonal matrix

$$\Phi(E) = \left(\begin{array}{cccccccc} 1 & \dots & 1 & 0 & \dots & \dots & \dots & 0 \\ \vdots & \ddots & \vdots & & & & & \vdots \\ 1 & \dots & 1 & & & & & \vdots \\ 0 & & & \ddots & & & & \vdots \\ \vdots & & & & \ddots & & & \vdots \\ \vdots & & & & & 1 & \dots & 1 \\ \vdots & & & & & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & \dots & 1 & \dots & 1 \end{array} \right) \quad (5)$$

where $r_1 + r_2 + \dots + r_s = n$,

$$\left\{ \begin{array}{l} A_1 = \dots = A_{r_1} = B_1, \\ \dots \\ A_{n-r_s+1} = \dots = A_n = B_n, \end{array} \right. \quad \left\{ \begin{array}{l} P_1^E = \dots = P_{r_1}^E = L_1^E, \\ \dots \\ P_{n-r_s+1}^E = \dots = P_n^E = L_s^E. \end{array} \right.$$

Proposition 2 yields information that $E(x_1,\dots,x_n)$ has the carrier $B_1^{r_1} \times \dots \times B_s^{r_s}$ and establishes s different equivalences L_i^E on $B_i \times B_i$. From now on, B_i/L_i^E stands for cosets and $[B_i/L_i^E]^{r_i}$ denotes the direct product of equal cosets, i.e. we can conclude that desired relation forms on $[B_1/L_1^E]^{r_1} \times \dots \times [B_s/L_s^E]^{r_s}$ or actually on $B_i/L_i^E \times \dots \times B_i/L_i^E$.

Definition 2. Arbitrary n -arity relation $E(x_1,\dots,x_n)$ induces on $B_i/L_i^E \times \dots \times B_i/L_i^E$ a relation E^M which will be referred to as a multirelation.

As it has already been stated above, it is important to understand that significance should be assigned to the simultaneous application of relations.

Definition 3. Two arbitrary relations n -arity $E_1(x_1,\dots,x_n)$ on $A_1 \times \dots \times A_n$ and m -arity $E_2(x_1,\dots,x_m)$ on $C_1 \times \dots \times C_m$ are externally (i,j) -coherent if and only if $A_i = A_j$ and $P_i^{E_1} = P_j^{E_2}$.

Obviously, on $\{1,2,\dots,n\} \times \{1,2,\dots,m\}$ an equivalence relation V_{E_1,E_2} and $(n \times m)$ matrix of external coherence $\Phi(E_1,E_2)$ are introduced similarly to the one relation case. More precisely, elements $l_{ij}, i = \overline{1,n}, j = \overline{1,m}$ of matrix $\Phi(E_1, E_2)$ are specified by expression

$$l_{ij} = \begin{cases} 1, A_i = A_j, P_i^{E_1} = P_j^{E_2}, \\ 0, otherwise. \end{cases}$$

Proposition 3. Two arbitrary relations $E_1(x_1,\dots,x_n)$ on $A_1 \times \dots \times A_n$ and $E_2(x_1,\dots,x_m)$ on $C_1 \times \dots \times C_m$ induce two multirelations E_1^M, E_2^M with the same carrier if and only if by rows (column) transpositions the matrix of external coherence $\Phi(E_1,E_2)$ is reduced to the block-diagonal form

$$S^M([x_{\alpha(n)}]_{P(n)}) = 1 \Leftrightarrow S(x_{\alpha(n)}) = 1. \quad (6)$$

Under condition (6) a multirelation S^M is congruent dependent on $P(n)$ and S what we denote by $S^M = \text{con}(P(n), S)$ for brevity.

Remark 1. It is easy enough to understand that the definition of congruent dependence is correct if and only if $P(n) = h - \text{fac} S$.

Remark 2. It is a simple matter to show that $P(n) = h - \text{fac} S$ if and only if $P(n)$ is induced by n -arity relation S satisfying (4).

Definition 5. Partial factor $P(n)$ will be named factor (full factor) with notation $P(n) = \text{fac} S$ if and only if

$$\forall x_{\alpha(n-1)} \in A(n-1), \forall x_{\alpha_n}^n, x_{\alpha'_n}^n \in A(n): S(x_{\alpha(n)}) = S(x_{\alpha'(n)}) = 1 \Rightarrow P_n[x_{\alpha_n}^n, x_{\alpha'_n}^n] = 1$$

where $x_{\alpha(n)} = (x_{\alpha(n-1)}, x_{\alpha_n}^n)$, $x_{\alpha'(n)} = (x_{\alpha(n-1)}, x_{\alpha'_n}^n)$.

Consequently, we get a multioperation F^M

$$F^M([x_{\alpha(n)}]_{P(n)}) = [x_{\alpha_n}^n]_{P_n} \Leftrightarrow S(x_{\alpha(n)}) = 1$$

where $[x_{\alpha_n}^n]_{P_n}$ is the coset of the set A_n in regard to the equivalence P_n and the element $x_{\alpha_n}^n$ belongs to this coset.

Remark 3. It is easy enough to see that $[x_{\alpha_n}^n]_{P_n}$ is unique coset. In this connection $F^M = \text{con}(P(n), S)$ and $P(n) = \text{fac} S$ if and only if $P(n)$ is induced by n -arity relation S satisfying (4).

The theorem of a multialgebraic system existence under given external equivalence and the same carrier had been proved [Mashtalir et al., 2003] and with mentioned notations it can be represented as follows.

Theorem 1. Suppose that A is arbitrary carrier, P is given equivalence on A^2 and $\Sigma_S = \{P, S_1, \dots, S_\beta, \dots\}$ is a family of n -arity relations then a model $\langle A, \Sigma_S \rangle$ generates multialgebraic system $\langle A/P, \{F_\xi^M\}, \{S_\eta^M\} \rangle$ where

$$F_\xi^M = \text{con}(P^n, S_\xi), \quad S_\eta^M = \text{con}(P^n, S_\eta) \text{ and}$$

$$\exists \Sigma_{1S}, \Sigma_{2S} \subset \Sigma_S : \Sigma_{1S} \cap \Sigma_{2S} = \emptyset, \quad \Sigma_{1S} \cup \Sigma_{2S} = \Sigma_S \setminus P, \quad S_\xi \in \Sigma_{1S}, \quad S_\eta \in \Sigma_{2S}$$

if and only if

$$\forall S_\beta \in \Sigma_S \setminus P \Rightarrow P^n = \begin{cases} h - \text{fac} S_\beta, & S_\beta \in \Sigma_{1S} \setminus P, \\ \text{fac} S_\beta, & S_\beta \in \Sigma_{2S} \setminus P. \end{cases}$$

It should be emphasized that any n -arity relation E forms its equivalence $L^E = \prod_{i=1}^s L_i^E$ on the carrier $B_1 \times \dots \times B_s$ (see the explication of expression (5)) which is determined by the matrix of internal coherence. Further, the carrier structure is direct product of matrix blocks. Hence, the consideration of relations and matrices of external coherence by pairs gives possibilities to establish conditions that due to proposition 3* all pairs $(S_{\beta'}, S_{\beta''})$ from this collection represent difunctional relations $V_{S_{\beta'}, S_{\beta''}}$. Granting remarks 1–3, we can restate theorem 1 and give more strong assertion of necessary and sufficient conditions for multialgebraic systems existence.

Theorem 2. Let $\{S_1, \dots, S_\beta, \dots\}$ be a family of arbitrary arity relations whose carriers may be different then multialgebraic system is induced if and only if

i) $\Sigma_S = \{L^E, S_1, \dots, S_\beta, \dots\}$, L^E is an equivalence induced by $S_1, \dots, S_\beta, \dots$,

ii) $\exists \Sigma_{1S}, \Sigma_{2S} \subset \Sigma_S : \Sigma_{1S} \cap \Sigma_{2S} = \emptyset, \Sigma_{1S} \cup \Sigma_{2S} = \Sigma_S \setminus L^E$,

iii) $\forall S_\beta \in \Sigma_S \setminus L^E \Rightarrow \underbrace{L^E \times \dots \times L^E}_n = \begin{cases} h-fac S_\beta, & S_\beta \in \Sigma_{1S} \setminus L^E, \\ fac S_\beta, & S_\beta \in \Sigma_{2S} \setminus L^E, \end{cases}$

iv) $\forall S_{\beta'}, S_{\beta''} \in \Sigma_S \setminus L^E$ and $\forall S_{\beta'}, S_{\beta''}$ is difunctional relation.

Thus, a factorization of information in any feature space conceptually is one of the basic methods providing an interpretation of data. On the one hand, identification can be required up to given or explored equivalence relations set. With another, construction of equivalence classes often represents an essence and a purpose of data processing. We have introduced and proved conditions describing interdependence of different levels information representations.

Conclusion

Different types of granulation represent different aspects of data and provide different types of knowledge embedded in data. An intelligent data analysis based on granular computing deals with theories, methodologies, techniques and tools that provide consideration what is relevant and permit to ignore irrelevant details. Granular computing involves two-way communications upward and downward in a hierarchy of different abstraction levels that represent different granulated views of problems understanding. It is reasonable to assume that granules relations satisfying various axiomatics and operations with operands corresponding to granules offer advantages for formalization of transformations and interpretations in multilevel processing of arbitrary nature data.

There exist two distinct varieties of relations concerning data to be analyzed. First of all, we should emphasized internal (embedded) interrelationships of original data. Thus, latent information that induces relationships between granules has to be explored. In the second place, a relation concerned with applications can be introduced on the original data. In both cases joint analysis has to be carried into effect.

Usually there are possibilities of empirical verification of the properties only at the lower level of abstractions, i.e. with the use of original data. Partitions and coverings can be normally valid models of granulation, and properties of relations along with operations in conformity with equivalence or tolerance classes generate a basic interest. In other words, the problem consists in an examination of original data to know properties of granule families. In our opinion multialgebraic systems can be sufficiently adequate tools in order to formalize elements of detailing or roughening such as granule, granulated view, granularity and hierarchy in the framework at least formal mathematical structures. We have established necessary and sufficient conditions of producing relations (with the same carrier) on granules induced by relations associated with original data. Furthermore, we have found conditions of multialgebraic systems existence. As development of these results it should be indicated the investigation of specific algebraic structures on original data such as semigroup, group, ring, different vector spaces etc. There arise several problems (it seems that peculiar but, vice versa, very important). Among them it should be noted comparisons of granule families for which there are no two ways about an introduction of an admissible metric on granule structures, e.g. on set partitions [Bobrowsky et al., 2006, Mashtalir et al., 2006] since it is often necessary to have dealings with a whole family of partitions and we have to be able to compare these partitions.

Bibliography

- [Bargiela, Pedrycz, 2002] A. Bargiela, W. Pedrycz Granular computing: an introduction. Boston, Kluwer Academic Publishers, The Kluwer International Series in Engineering and Computer Science. Vol. 717. 2002. 478 p.
- [Bobrowsky et al., 2006] L. Bobrowsky, V. Mashtalir, V. Shlyakhov Metrics on arbitrary partitions of measurable sets. Global Information Systems. Problems and Tendencies of Progress. Kharkov, KnNURE. 2006. pp. 70–71.
- [Doherty et al., 2003] P. Doherty, W. Lukaszewicz, A. Szalas Information granules for intelligent knowledge structures. Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing / G. Wang, Q. Liu, Y. Yao, A. Skowron (Eds.) Lecture Notes in Artificial Intelligence. Springer-Verlag Berlin Heidelberg. 2003. Vol. 2639. pp. 405-412.
- [Lin, 2003] T.Y. Lin Granular computing (Structures, Representations, and Applications). Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing. G. Wang, Q. Liu, Y. Yao, A. Skowron (Eds.) Lecture Notes in Artificial Intelligence. Springer-Verlag Berlin Heidelberg. 2003. Vol. 2639. pp. 16-24.
- [Mashtalir et al., 2006] Mashtalir V., Mikhnova E., Shlyakhov V., Yegorova E. A novel metric on partitions for image segmentation. IEEE International Conference on Video and Signal Based Surveillance, avss, p. 18, 2006.
- [Mashtalir, Shlyakhov, 2003] Mashtalir V.P., Shlyakhov V.V. Properties of multialgebraic systems in problems of comparative recognition. Cybernetics and Systems Analysis. Vol.39, No.6. 2003. pp. 790–804.
- [Pal et al., 2005] S.K. Pal, B.U. Shankar, P. Mitra Granular computing, rough entropy and object extraction. Pattern Recognition Letters. 2005. Vol. 26. pp. 2509-2517.
- [Yager, 2002] R.R. Yager Using granular objects in multi-source data fusion // Rough Sets and Current Trends in Computing. J.J. Alpigini, J.F. Peters, A. Skowron, N. Zhong (Eds.). Lecture Notes in Artificial Intelligence. Springer-Verlag Berlin Heidelberg. 2002. Vol. 2475. pp. 324-330.
- [Yao, 2001] Y.Y. Yao Information granulation and rough set approximation. International Journal of Intelligent Systems. 2001. Vol. 16, No. 1, pp. 87-104.
- [Yao, Yao, 2002] J.T. Yao, Y.Y. Yao Induction of classification rules by granular computing. Rough Sets and Current Trends in Computing / J.J. Alpigini, J.F. Peters, A. Skowron, N. Zhong (Eds.). Lecture Notes in Artificial Intelligence. Springer-Verlag Berlin Heidelberg. 2002. Vol. 2475. pp. 331-338.
- [Yao, 2005] Y.Y. Yao Perspectives of granular computing. Proceedings of 2005 IEEE International Conference on Granular Computing. 2005. Vol. 1. pp. 85-90.
- [Yao, 2006] Y.Y. Yao Granular computing for data mining // Proceedings of SPIE Conference on Data Mining, Intrusion Detection, Information Assurance and Data Networks Security / B.V. Dasarathy (Ed.), Kissimmee, Florida, USA. 2006. pp. 1-12 (624105).
- [Zadeh, 1997] L.A. Zadeh Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic // Fuzzy Sets Systems. 1997. Vol. 19. pp. 111-127.

Authors' Information

Kagramanyan Alexander – Kharkov National University, Svobody sq., 4, Kharkov, Ukraine, 61077

Mashtalir Vladimir – Kharkov National University of Radio Electronics, Lenina ave., 14, Kharkov, Ukraine, 61166, mashtalir@kture.kharkov.ua

Shlyakhov Vladislav – Kharkov National University of Radio Electronics, Lenina ave., 14, Kharkov, Ukraine 61166

PREDICTION OF PROPERTIES AND STATES OF DYNAMIC OBJECTS THROUGH ANALOGICAL INFERENCE

Yurij Kuk, Helen Lavrikova

Abstract: *A technique of analogical inference for prediction of properties and states of dynamic objects the attributes which vary randomly in time is suggested. The technique is based on linear transformation of attributes of objects with the purpose to ensure the maximum probability of correct prediction of their properties and states.*

Keywords: *artificial intelligence, analogical inference, dynamic objects, prediction of properties.*

ACM Classification Keywords: *1.2.4 Artificial Intelligence: knowledge representation formalisms and methods.*

Introduction

The report analyses complex multiple parameter dynamic objects, whose operation is determined by the set of attributes, which by its nature can take random values. Many such objects exist in the nature and in the human society. Functioning of economical and the environmental systems, of safety systems, complex objects of new technical equipment, etc., as a rule, depends on the behavior of a number of various attributes. Random processes, which describe the values of these attributes, determine the properties and states of the objects in study. For prediction of properties and states of dynamic objects we will apply the structure predicate model of knowledge (SPMK) [1], which represents a special type of network structure - an object - predicate net. The properties of the objects and their relations in it are described with predicates. This net represents a graph the nodes of which correspond to the objects and the predicates. Two types of objects are examined: the complex and the primary objects. Primary objects are understood as the objects that form a part of complex objects. This work examines a special case when some sets of the objects possessing a certain property act as complex objects. Let us designate such complex objects as groups. A characteristic property of each object of the set of objects of the given group is also a property of this complex object. The objects included into the into group act as the primary objects. The primary predicates are understood as the predicates indicating properties and relations of primary objects, and the predicates of complex objects are understood as the predicates indicating properties and relations of complex objects. We will divide the predicates of primary and complex objects expressing numerical characteristics of the objects into initial and derivative, that is those received as a result of mathematical transformations of the initial ones. We will divide all derivative predicates into the levels of transformation as follows. Initial predicates we will consider as the predicates of zero level of transformation. The predicates received by transformation of initial predicates, we will designate as predicates of the 1-st level of transformation. The predicates received by transformation of predicates of the 0 and 1-st levels, we will name predicates of the 2-nd level of transformation, etc. Semantic object predicate net is a four-layer graph of a pyramidal net, separate strata of which form its nodes. The first stratum P_0 corresponds to the initial predicates, indicating the properties and relations of the primary objects, and also various levels of transformation obtained basing on their derivative predicates. Elements P_0 are primary predicates. The second stratum A corresponds to the names of the primary objects. When interpreted they form subject domains of the primary predicates. The third stratum S corresponds to the names of the complex objects. In our case these are the names of the groups. The fourth stratum V are the predicates indicating the properties and relations of the complex objects, and also the derivative predicates of various levels of transformation received on their basis. In our case these are predicates which characterize various groups of objects. Elements V are the predicates of the complex objects. Their subject domains are complex objects. Curves of the lower and upper layers connect the nodes representing the objects, to the nodes, representing the predicates, and are directed from the primary and the complex objects to the predicates. They are used in interpretation of the predicates. Let ω mean the repetition

factor of some predicate. Then the presence of the ω curves originating from the ω objects and converging in the given predicate, corresponds to the logical value of the predicate "true" when these objects are substituted in the predicate, and to the value "false" – in case of substitution of the object in the predicate in the absence of the curve connecting the given object to the predicate. Curves of the middle layer connect the nodes corresponding to the primary objects, to the nodes representing the complex objects. The primary elements at which the curves originate are a part of those complex objects at which these curves end. On the basis of the analysis of such object predicate net in operation, a procedure of obtaining of new knowledge about the properties and states of unknown dynamic objects through processing of experimental data is suggested. The procedure is based on measurement of distances between the predicates of the objects in study, the properties and states of which are unknown, and the predicates of groups.

Prediction of the properties of dynamic objects through analogical inference

Let us perform prediction of properties and states of dynamic objects through analogical inference, using SPMK. The inference of reasoning by analogy is the inference based on transference of reasoning from the examined area into the homomorphic area, i.e. the area somewhat similar to that examined. J groups of the objects G_1, \dots, G_J of a certain subject domain will act as an examined area. Each of these groups consists of the objects possessing the characteristic specific for this group. These sets of objects form a training sample for SPMK. Available is also a set of objects - the so-called examination sample G_e . The task is to determine for each object from the set G_e one group G_j or several groups G_j, \dots, G_l to which it is homomorphic. After that it is possible to draw the following inference by analogy: the given object of set G_e either possesses a property of one group G_j or totally possesses the properties of several groups G_j, \dots, G_l . Such inference is nothing else but some new knowledge of the object in study. It is obvious, that reliability of such knowledge requires the further testing in practice. In the work ion we suppose, that the degree of similarity of the objects from the examination sample to the objects of the area in study is determined by the distance which is measured by means of the measure constructed in a certain way, between the attributes of the objects from the set G_e and the predicates of the objects of the groups G_1, \dots, G_J . The rule of inference by analogy can be formulated as follows. Let P_1, P_2, \dots, P_J be the sets of predicates of groups G_1, \dots, G_J , and V_1, V_2, \dots, V_J correspondingly be the properties describing the objects of these groups. Let P_e be a set of predicates of some object from the set G_e . Then, if the distance between the sets of the predicates P_j and P_e : $d(P_j, P_e) < r_1$, where r_1 is a certain threshold, than $P_e \rightarrow (V_j) \wedge (\neg V_i), i \neq j$ with a certain reliability q_1 , i.e. the inference is made that the object possesses property V_j and does not possess properties $V_i, i \neq j$.

Statement of a problem

Let there be a certain dynamic object which can be in one of L states. The information on the state is read with M sensors at discrete instants $t = 1, 2, 3, \dots$. The output of each of the sensors represents samplings of the signal amplitudes describing the states of the corresponding sector of the object in study and being the values of the time series. The task is to predict by observations of the sensors the state of a dynamic object, preliminary experimental data of observations of the sensors in different states of the dynamic object being available. To solve the problem let us act as follows. The experimental data for the time series for the i -th sensor and the l -th state of a dynamic object we will designate as $X^{(i,l)}(t), i = 1, \dots, M, l = 1, \dots, L, t = 1, 2, 3, \dots$. Each time series $X^{(i,l)}(t)$, containing S measurements we will divide in Z the equal cuts, containing $n = S/Z$ measurements. We will consider these informational cuts as some primary objects, of descriptions of which we will form training and examination samples. Let us select initial predicates - attributes by means of which we will

describe the obtained objects. Let us designate as $s_k^{(i,j)}$ the k -th object generated from the k -th cut of time series $X^{(i,l)}(t)$. Let us examine the general case when observable cuts of the time series represent non-stationary processes. For determination of attributes of the primary objects we will act as follows. We will decompose each cut of the time series in a Fourier series: $X^{(i,l)}(t) = a_0 + \sum_j [a_j \cos(\lambda_j t) + b_j \sin(\lambda_j t)]$, for $j=1, \dots, n, l=1, \dots, L$. Numbers λ_j are referred to as the frequencies, and the magnitudes $p_j = (a_j^2 + b_j^2) * (n/2)$ - as the periodograms. Let us take as the attributes of the object $s_k^{(i,l)}$ N specific features of the function which describes the dependence of p_j on λ_j , and is referred to as the spectrogram. For example, the quantity of maxima of this function, frequency λ_j to which the maxima of the periodogram correspond, and the values of the periodograms corresponding to them. Let us designate selected attributes as $x_{1k}^{(i,l)}, x_{2k}^{(i,l)}, \dots, x_{Nk}^{(i,l)}$. Depending on the state of examined dynamic object, statistical characteristics of initial attributes of the primary objects will vary. To predict the states of dynamic objects let us divide Z objects of each time series $X^{(i,l)}(t), i=1, \dots, M, l=1, \dots, L$, into K and W objects where K objects belong to the training sample, and W objects - to the examination sample. $K+W=Z$. As a result we will receive $J=M \times L$ groups of objects of the training sample:

$$G_1(l) = \{s_1^{(1,l)}, s_2^{(1,l)}, \dots, s_K^{(1,l)}\}, \dots, G_M(l) = \{s_1^{(M,l)}, s_2^{(M,l)}, \dots, s_K^{(M,l)}\}, l=1, \dots, L.$$

Let us generate SPMK in which the role of the primary objects is played by the considered above cuts of the time series. Initial primary predicates of these objects are features of spectrograms of corresponding time series. As complex objects $G_1(l), \dots, G_M(l), l=1, \dots, L$, let us take $J=M \times L$ groups of objects of the training sample: $G_1(l), \dots, G_M(l), l=1, \dots, L$. To each node of the complex object $G_i(l)$ enters K curves from corresponding objects $s_k^{(i,l)}$, to each of which enters N curves from corresponding initial primary predicates $x_{1k}^{(i,l)}, x_{2k}^{(i,l)}, \dots, x_{Nk}^{(i,l)}$. The problem consists in the following. Basing on the available experimental data obtained from all sensors to determine the state of a dynamic object, i.e. it is required to determine unknown parameter l by the sample $x_{1k}^{(i,l)}, x_{2k}^{(i,l)}, \dots, x_{Nk}^{(i,l)}, i=1, \dots, M$. Or, in the other words, it is required to select from L sets of groups: $G_1(l), \dots, G_M(l), l=1, \dots, L$, a certain single set to which the sample $x_{1k}^{(i,l)}, x_{2k}^{(i,l)}, \dots, x_{Nk}^{(i,l)}, i=1, \dots, M$, corresponds in the best way.

Representation of knowledge about a training sample by means of SPMK

To solve a problem of prediction of the state of a dynamic object, for all objects of the SPMK by the initial primary predicates, derivative predicates of various levels of transformation are constructed. As derivative predicates of the 2-nd level of transformation let us take average values of attributes for each group of training objects. For convenience of let us put the attributes of an object $s_k^{(i,l)}$ in form of vector Euclid space R_N : $x_{1k}^{(i,l)}, x_{2k}^{(i,l)}, \dots, x_{Nk}^{(i,l)}$, besides that when considering any l -th state of dynamic object, we will drop an index l in the formulas. Then for the group of objects $G_i = \{s_1^{(i)}, s_2^{(i)}, \dots, s_K^{(i)}\}$ derivative predicates of the 2-nd level of transformation coincide with components of vector $h^{(i)} = (\bar{x}_1^{(i)}, \bar{x}_2^{(i)}, \dots, \bar{x}_N^{(i)})$ the coordinates of which are equal to the component-wise average values of vectors $x_k^{(i)} = (x_{1k}^{(i)}, x_{2k}^{(i)}, \dots, x_{Nk}^{(i)})$ of all training objects which are included into the given group:

$$\bar{x}_1^{(i)} = \frac{1}{K} \sum_{\nu=1}^K x_{1\nu}^{(i)}, \bar{x}_2^{(i)} = \frac{1}{K} \sum_{\nu=1}^K x_{2\nu}^{(i)}, \dots, \bar{x}_N^{(i)} = \frac{1}{K} \sum_{\nu=1}^K x_{N\nu}^{(i)}.$$

Let us take as derivative predicates of the 3-rd level of transformation deviations of the values of the attributes of objects from the group averages. For the object $s_k^{(i)}$ belonging to the group of training objects $G_i = \{s_1^{(i)}, s_2^{(i)}, \dots, s_k^{(i)}\}$, derivative predicates of the 3-rd level of transformation coincide with components of vector $\tilde{x}_k^{(i)} = (x_{1k}^{(i)} - \bar{x}_1^{(i)}, x_{2k}^{(i)} - \bar{x}_2^{(i)}, \dots, x_{Nk}^{(i)} - \bar{x}_N^{(i)})$. As far as the states and properties of the objects in study are essentially influenced by the existence of relations between the attributes of the object, we will take as derivative predicates of the 4-th level of transformation we will take values of relations between all pairs of the attributes of objects from every group. We will measure the value of relations by means of sample factors of a covariance. To determine the values of these derivative predicates we will introduce matrixes $A^{(1,l)}, \dots, A^{(M,l)}$, $l = 1, \dots, L$ the columns of which consist of deviations of the values of the attributes of objects from their group averages for corresponding groups of objects. They contain N lines and accordingly K columns and look like following:

$$A^{(1,l)} = \begin{pmatrix} x_{11}^{(1,l)} - \bar{x}_1^{(1,l)} & \dots & x_{1K}^{(1,l)} - \bar{x}_K^{(1,l)} \\ \dots & \dots & \dots \\ x_{N1}^{(1,l)} - \bar{x}_1^{(1,l)} & \dots & x_{NK}^{(1,l)} - \bar{x}_K^{(1,l)} \end{pmatrix}, \dots, A^{(M,l)} = \begin{pmatrix} x_{11}^{(M,l)} - \bar{x}_1^{(M,l)} & \dots & x_{1K}^{(M,l)} - \bar{x}_K^{(M,l)} \\ \dots & \dots & \dots \\ x_{N1}^{(M,l)} - \bar{x}_1^{(M,l)} & \dots & x_{NK}^{(M,l)} - \bar{x}_K^{(M,l)} \end{pmatrix}$$

We form matrixes $B^{(1,l)} = A^{(1,l)} A^{(1,l)T}$, \dots , $B^{(M,l)} = A^{(M,l)} A^{(M,l)T}$, $l = 1, \dots, L$, where the index T indicates the operation of transposition of the matrix. They look like following:

$$B^{(1,l)} = \begin{pmatrix} \sum_{\nu=1}^K (x_{1\nu}^{(1,l)} - \bar{x}_1^{(1,l)})^2 & \sum_{\nu=1}^K (x_{1\nu}^{(1,l)} - \bar{x}_1^{(1,l)})(x_{2\nu}^{(1,l)} - \bar{x}_2^{(1,l)}) & \dots & \sum_{\nu=1}^K (x_{1\nu}^{(1,l)} - \bar{x}_1^{(1,l)})(x_{N\nu}^{(1,l)} - \bar{x}_N^{(1,l)}) \\ \sum_{\nu=1}^K (x_{2\nu}^{(1,l)} - \bar{x}_2^{(1,l)})(x_{1\nu}^{(1,l)} - \bar{x}_1^{(1,l)}) & \sum_{\nu=1}^K (x_{2\nu}^{(1,l)} - \bar{x}_2^{(1,l)})^2 & \dots & \sum_{\nu=1}^K (x_{2\nu}^{(1,l)} - \bar{x}_2^{(1,l)})(x_{N\nu}^{(1,l)} - \bar{x}_N^{(1,l)}) \\ \dots & \dots & \dots & \dots \\ \sum_{\nu=1}^K (x_{N\nu}^{(1,l)} - \bar{x}_N^{(1,l)})(x_{1\nu}^{(1,l)} - \bar{x}_1^{(1,l)}) & \sum_{\nu=1}^K (x_{N\nu}^{(1,l)} - \bar{x}_N^{(1,l)})(x_{2\nu}^{(1,l)} - \bar{x}_2^{(1,l)}) & \dots & \sum_{\nu=1}^K (x_{N\nu}^{(1,l)} - \bar{x}_N^{(1,l)})^2 \end{pmatrix}$$

.....

$$B^{(M,l)} = \begin{pmatrix} \sum_{\nu=1}^K (x_{1\nu}^{(M,l)} - \bar{x}_1^{(M,l)})^2 & \sum_{\nu=1}^K (x_{1\nu}^{(M,l)} - \bar{x}_1^{(M,l)})(x_{2\nu}^{(M,l)} - \bar{x}_2^{(M,l)}) & \dots & \sum_{\nu=1}^K (x_{1\nu}^{(M,l)} - \bar{x}_1^{(M,l)})(x_{N\nu}^{(M,l)} - \bar{x}_N^{(M,l)}) \\ \sum_{\nu=1}^K (x_{2\nu}^{(M,l)} - \bar{x}_2^{(M,l)})(x_{1\nu}^{(M,l)} - \bar{x}_1^{(M,l)}) & \sum_{\nu=1}^K (x_{2\nu}^{(M,l)} - \bar{x}_2^{(M,l)})^2 & \dots & \sum_{\nu=1}^K (x_{2\nu}^{(M,l)} - \bar{x}_2^{(M,l)})(x_{N\nu}^{(M,l)} - \bar{x}_N^{(M,l)}) \\ \dots & \dots & \dots & \dots \\ \sum_{\nu=1}^K (x_{N\nu}^{(M,l)} - \bar{x}_N^{(M,l)})(x_{1\nu}^{(M,l)} - \bar{x}_1^{(M,l)}) & \sum_{\nu=1}^K (x_{N\nu}^{(M,l)} - \bar{x}_N^{(M,l)})(x_{2\nu}^{(M,l)} - \bar{x}_2^{(M,l)}) & \dots & \sum_{\nu=1}^K (x_{N\nu}^{(M,l)} - \bar{x}_N^{(M,l)})^2 \end{pmatrix}$$

To determine the value of relation between the i -th and the j -th attributes for some group of objects it is necessary to take value of the element on the intersection of the i -th line and the j -th column in the matrix of dispersion B with the upper index of the corresponding group and to divide it into the number of objects in this group, i.e. in K , then the diagonal elements will take values of sample variances of the attributes. The values of these matrixes and consequently the values of the derivative predicates of the 4-th level of transformation, without division into K , determine a mutual scatter of vectors of the attributes of the given group of objects in relation to the vector of their average values in the vector space R_N . Therefore we will call such matrixes matrixes of dispersion. Let us take values of the elements of the total matrix $B = B^{(1,1)} + \dots + B^{(M,L)}$ as

derivative predicates of the 5-th level of transformation. As the derivative predicates of the 6-th level of transformation we will take average values for each attribute of objects from the incorporated training sample, i.e. without consideration of belonging of an object to a certain group. Values of derivative predicates of the 6-th level of transformation coincide with components of the vector $h = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N)$, where

$$\bar{x}_1 = \frac{1}{M * L} \sum_{l=1}^L \sum_{i=1}^M \bar{x}_1^{(i,l)}, \bar{x}_2 = \frac{1}{M * L} \sum_{l=1}^L \sum_{i=1}^M \bar{x}_2^{(i,l)}, \dots, \bar{x}_N = \frac{1}{M * L} \sum_{l=1}^L \sum_{i=1}^M \bar{x}_N^{(i,l)}.$$

As derivative predicates of the 7-th level of transformation we will take elements of matrix of dispersion for the attributes of objects from the incorporated training sample, without consideration of their belonging to a certain group. Let us designate this matrix as - W . The quantity of its elements, and consequently the quantity of such derivative predicates is N^2 .

Optimization of the net

Optimization of the net will be understood as its transformation in order to receive the optimal solution of the problem of prediction of properties and states of dynamic objects. As it was already noted, the values of the attributes of objects can be examined in form of vectors or points of the Euclidean space R_N . Vectors of attributes of objects of different groups will be mapped in this space with some sets of points which will be placed in it in a certain way. Let us put an optimization problem to find some linear transformation of vectors of attributes of objects so that for the objects with new values of attributes following two conditions will be met. Condition 1) - the points corresponding to the average values of transformed attributes for different groups of objects - the so-called centers of groups, - would be located as far as possible from each other; and condition 2) - the points corresponding to the same group of objects would be concentrated around their centers as close as possible. For solution of this problem we used the values of all derivative predicates represented in the net, besides that, the methods of linear algebra and optimization were used. The extremum was found by means of Lagrange factors. Let us take without proof the following basic theorem which allows to construct the necessary transformation.

Theorem 1. Let $B = B^{(1,1)} + \dots + B^{(M,L)}$ be a total matrix of dispersion for separate groups, and W - be a matrix of dispersion for attributes of the incorporated training sample. Let $V = W - B$.

Then $N-1$ eigenvectors $C_1 = (c_{11}, \dots, c_{N1})$, $C_2 = (c_{12}, \dots, c_{N2})$, ..., $C_{N-1} = (c_{1,N-1}, \dots, c_{N,N-1})$ of the matrixes $VB^{(-1)}$ determine hyperplane Q , projections of vectors of attributes of objects on which met the above formulated conditions 1) and 2) of the problem of optimization.

Let's construct a hyperplane Q with the help of the vectors $C_1 = (c_{11}, \dots, c_{N1})$, $C_2 = (c_{12}, \dots, c_{N2})$, ..., $C_{N-1} = (c_{1,N-1}, \dots, c_{N,N-1})$ found in theorem 1 Let us project N - dimensional vectors

$$x_{1k}^{(i,l)}, x_{2k}^{(i,l)}, \dots, x_{Nk}^{(i,l)}, k = 1, \dots, K, l = 1, \dots, L,$$

of the attributes of objects $s_k^{(i,l)}$, $k = 1, \dots, K, l = 1, \dots, L$ on it, and designate projections of these vectors as $z_k^{(i,l)} = (z_{1k}^{(i,l)}, \dots, z_{(N-1)k}^{(i,l)})$, $k = 1, \dots, K, l = 1, \dots, L$. Their dimension is equal $N-1$, and their components are equal to:

$$z_{pk}^{(i,l)} = \sum_{j=1}^N c_{jp} x_{jk}^{(i,l)}, p = 1, \dots, N-1, k = 1, \dots, K.$$

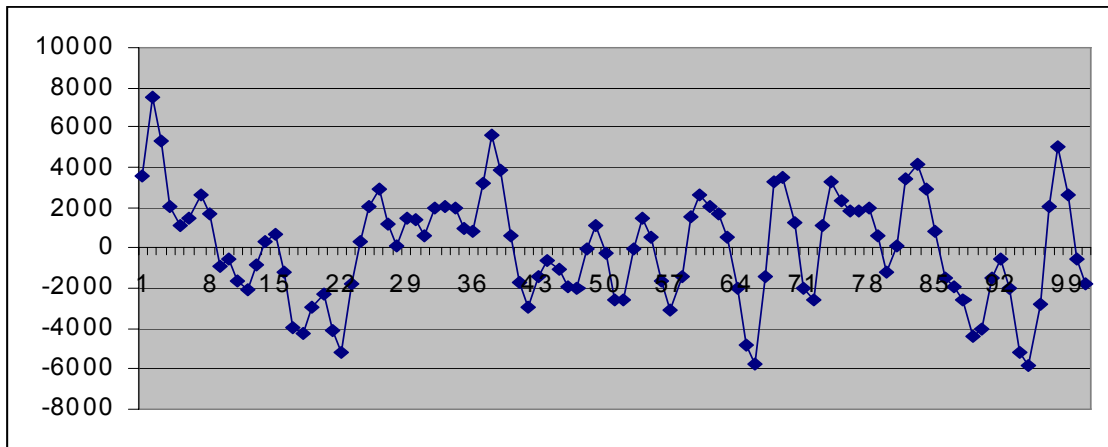


Fig.1. Time series from 6-th vibration sensor. Load of 8 tons.

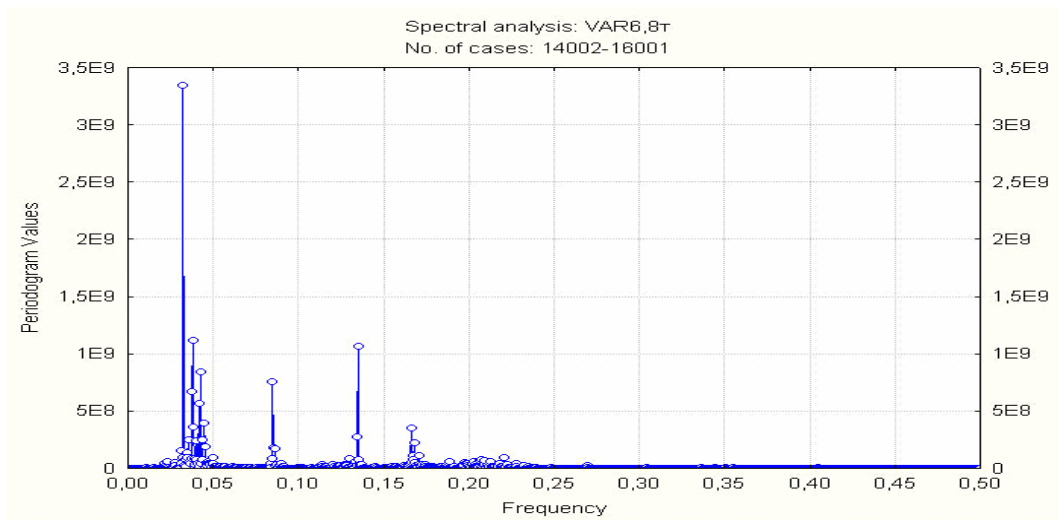


Fig.2. The typical spectrogram of time series from the 6-th vibration sensor. Load of 8 tons.

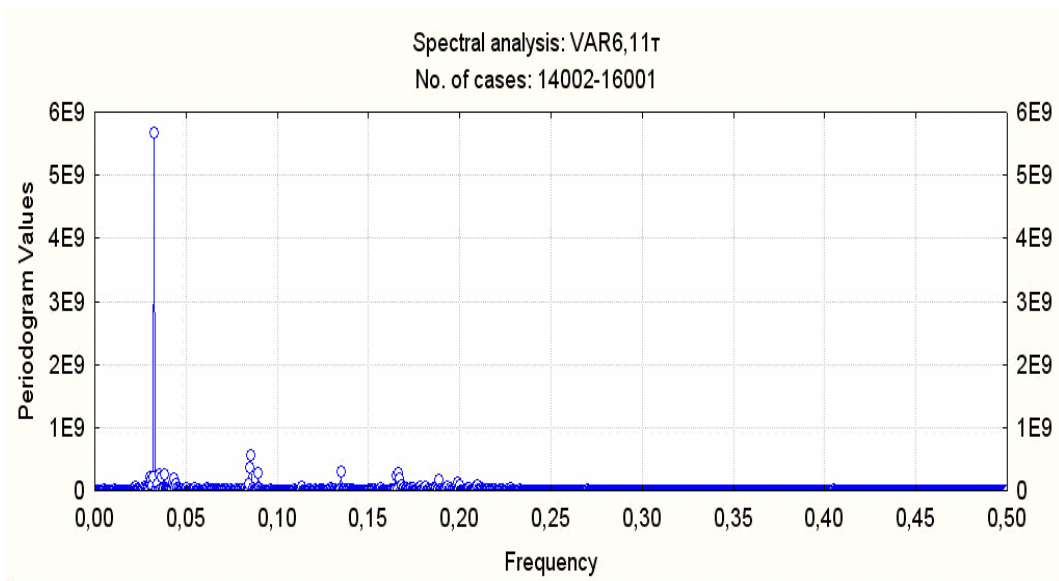


Fig. 3. The typical spectrogram of time series from the 6-th vibration sensor. Load of 11 tons.

Solution of problem of prediction of properties and states of dynamic objects

The linear transformation found in solution of optimization problem will transform the vectors of attributes $x_{1k}^{(i,l)}, x_{2k}^{(i,l)}, \dots, x_{Nk}^{(i,l)}$, $l = 1, \dots, L$ into the vector $z_k^{(i,l)} = (z_{1k}^{(i,l)}, \dots, z_{(N-1)k}^{(i,l)})$, $l = 1, \dots, L$, lying on the hyperplane Q . For the group of objects $G_i(l) = \{s_1^{(i,l)}, s_2^{(i,l)}, \dots, s_K^{(i,l)}\}$, $l = 1, \dots, L$, with transformed attributes, the derivative predicates of the 2-nd level the transformations expressing the average group values of initial attributes will be transformed into the components of the vector of the center $z^{(i,l)} = (\bar{z}_1^{(i,l)}, \bar{z}_2^{(i,l)}, \dots, \bar{z}_N^{(i,l)})$, $l = 1, \dots, L$, for the given group of objects. These components look like following:

$$\bar{z}_1^{(i,l)} = \frac{1}{K} \sum_{v=1}^K z_{1v}^{(i,l)}, \bar{z}_2^{(i,l)} = \frac{1}{K} \sum_{v=1}^K z_{2v}^{(i,l)}, \dots, \bar{z}_N^{(i,l)} = \frac{1}{K} \sum_{v=1}^K z_{Nv}^{(i,l)}, l = 1, \dots, L.$$

We will solve the problem of prediction of properties and states of dynamic objects as follows. Let $x = (x_1, x_2, \dots, x_N)$ be a vector of attributes of some examination object. In the 1-st step we find the projection of this vector to the hyperplane Q . As a result we receive the vector $y = (y_1, y_2, \dots, y_{N-1})$. Its components are equal to $y_p = \sum_{j=1}^N c_{jp} x_j$, where the factors c_{jp} , $p = 1, \dots, N-1$, $j = 1, \dots, N$, are components of eigenvectors

$$C_1 = (c_{11}, \dots, c_{N,1}), C_2 = (c_{12}, \dots, c_{N,2}), \dots, C_{N-1} = (c_{1,N-1}, \dots, c_{N,N-1})$$

of the matrix $VB^{(-1)}$. In the 2-nd step the Euclidean distance between the vector y and all centers of groups $z^{(1,l)}, z^{(2,l)}, \dots, z^{(M,l)}$, $l = 1, \dots, L$, are found. Let us designate them as $d_{1,l}, \dots, d_{M,l}$.

$$d_{i,l} = \sqrt{(y_1 - \bar{z}_1^{(i,l)})^2 + (y_1 - \bar{z}_2^{(i,l)})^2 + \dots + (y_1 - \bar{z}_N^{(i,l)})^2}, i = 1, \dots, M, l = 1, \dots, L.$$

We find the group to which the vector y is most close. Then we conclude by analogy, that this examination object possesses a property or a state of objects of the found group.

Problem in discernment of operating modes of the aircraft

Let's consider a problem of recognition of operating modes of helicopter with the loads of 8 and 11 tons. In solution of the problem we used the data obtained as a result of vibration meter tests of the main reducer of helicopter Ka 32N04. The data were taken from 6 sensors of the helicopter, from each of which 48000 samplings were obtained. The example of the graph of the time series, describing the first 100 samplings from the 6-th sensor for the load of 8 tons is represented on fig.1.

It was found that for reliable recognition on the basis of load spectrograms (8 or 11 tons) readings from only one sensor are sufficient. For this work the 6-th sensor was selected.

By the attributes describing prominent characteristic features of the spectrogram of the objects from the training sample, on the basis of the spectrogram of an object from the examination sample the conclusion was made about the operating mode of the helicopter. The typical spectrogram for the time series from the 6-th vibration sensor and the load of 8 tons is represented on fig. 2., and for the load of 11 tons - on fig.3.

One of the differences of the spectrogram on fig. 2 (load of 8 tons) from the spectrogram on fig. 3 (load of 11 tons) is that in the frequency interval of 0,1-0,15 there is a maximum which is essentially higher than the corresponding maximum of the spectrogram on fig. 3 in the same frequency interval. On the basis of this attribute only it is possible to determine reliably an operating mode of the helicopter.

Conclusion

The work suggests a technique of prediction of properties and states of dynamic objects, which is based on measurement of Euclidean distances between the attributes of the objects.

References

1. V.N.Koval, Yu.V.Kuk. A technique of acquisition of new knowledge in object-predicate semantic nets. // the Artificial intelligence. - 2005. - №3. - p. 25 - 36.

Authors' Information

Yurij Kuk – The Institute of Cybernetics of National Academy of Science of the Ukraine, the senior scientist, address: 40 Glushkov ave., Kiev, Ukraine; 03680; e-mail: 1913@i.com.ua

Helen Lavrikova – The Institute of Cybernetics of National Academy of Science of the Ukraine, address: 40 Glushkov ave., Kiev, Ukraine; 03680; e-mail: icdepval@ln.ua

АВТОМАТИЧЕСКИЙ СИНТЕЗ ФУНКЦИОНАЛЬНЫХ СХЕМ

Наталья Филатова, Олег Ахремчик, Олег Куприянов

Аннотация: Доклад посвящается интеллектуализации компьютерной обработки информации в области автоматизированного проектирования систем управления. Основой автоматического синтеза служат модели структуры и описания составляющих ее основных элементов. Алгоритмы автоматического синтеза схем являются результатом применения технологий инженерии знаний и поиска данных при определении отношений порядка. Связи на структурной схеме детализируются посредством проверки условий совместимости и установления дополнительных элементов для их истинности. Представляется клиент-серверная архитектура программного обеспечения для автоматического синтеза функциональных схем. Основным ее компонентом является база данных. Задания, схемы и описания элементов составляют ее декларативную часть. Правила синтеза – процедурную часть. Количество генерируемых вариантов ограничивается. Апробация программного обеспечения доказывает возможность автоматического синтеза целей систем управления в ходе разработки функциональных схем.

Ключевые слова: представление знаний, САПР, автоматический синтез.

1. Задача автоматического синтеза функциональных схем

На современном этапе развития САПР систем управления остается не решенной проблема автоматического синтеза схем. Существующие программные средства автоматизированного проектирования поддерживают решение этих задач в интерактивном режиме. Все процедуры выбора, оценки варианта технического решения выполняет инженер. В этой ситуации существенно ограничивается число анализируемых альтернативных вариантов схем, что приводит к снижению качества проектного решения.

Все технологии разработки схем систем управления включают четыре взаимосвязанных этапа (табл.1). Каждый этап завершается созданием определенного типа схемы. Выделение отдельных этапов в

решении этих задач связано с блочно-иерархическим представлением объекта проектирования (ОП) в САПР [1, 2].

Таблица 1 позволяет обосновать требования к структуре знаний предметной области. Действительно, при создании структур, моделирующих схему рассуждения эксперта, разрабатывающего схему системы автоматизации (СА), можно ограничиться знаниями, охватывающими первые 4 вертикальных уровня проектирования. Знания, связанные с методами создания блоков, (охватывают вторые 4 вертикальных уровня проектирования) не будут востребованы при разработке систем СА, как и аналогичные структуры третьего уровня.

Таблица 1.

Объект проектирования	Уровни проектирования (типы схем)			
	1	2	3	4
Система автоматизации (СА)	Структурная схема (СС)	функциональная схема (ФС)	Принципиальная электрическая схема (ПЭС)	Монтажно-коммутационная схема (МКС)
Блок / устройство	- " -	- " -	- " -	- " -
Элемент блока	- " -		- " -	- " -

Каждому уровню проектирования соответствует, по меньшей мере, одна модель объекта проектирования (схемы СА). В САПР схем СА должна поддерживаться иерархия схемных моделей вида : м_СС => м_ФС; м_ФС => м_ПЭС; м_ПЭС => м_МКС. Процедурные знания САПР должны охватывать методы преобразования всех выделенных моделей объекта проектирования.

Вертикальные уровни декомпозиции объекта проектирования позволяют выделить семантические группы предметных отношений. Каждая такая группа включает отношения, связанные с выполнением одной или нескольких операций. Горизонтальные уровни декомпозиции объекта проектирования позволяют выделить его структурные особенности, которые должны найти отражение в концептуальной структуре предметной области. На основе таблицы можно сделать вывод, что объект проектирования "схема СА" приводит к иерархии трех классов сущностей:

- система автоматизации, заданная на множестве промышленных систем;
- технический блок (устройство), определенный на множестве общепромышленных технических средств автоматики (ТСА);
- элемент блока, определенный на множестве типовых элементов замены.

Как показывает опыт создания интеллектуальных САПР, любая модель j-ого уровня проектирования опирается только на соответствующую модель (j+1)-ого уровня [3]. Поддерживать связи моделей j-ого уровня с моделями (j+2)-ого уровня не целесообразно. Следовательно, в концептуальной структуре знаний рассматриваемой САПР можно ограничиться двумя первыми классами сущностей - схема СА и блок.

Полученные выводы были учтены при разработке системы Functional Design of a Control System (FDcS) для автоматизированной поддержки начальных этапов проектирования систем автоматизации технологических объектов. Главное назначение этой системы – генерация множества вариантов технических решений (в виде функциональных схем автоматизации) при слабых ограничениях на область поиска вариантов.

Постановка задачи синтеза схемы СА и методы ее решения зависит от уровня проектирования. Разработка структурной схемы выполняется с использованием методов теории автоматического управления, и поддерживается множеством известных программ моделирования систем. Построенную модель м_СС можно рассматривать как единственное решение, являющееся техническим заданием для всех последующих этапов.

При формировании функциональной схемы решается вопрос технической реализации системы автоматизации, в схему включают блоки технических средств автоматики (ТСА), выпускаемые промышленностью. На основе одной структурной схемы можно создать несколько различных вариантов функциональных схем (ФС), отличающихся по стоимости и некоторым эксплуатационным характеристикам. Учитывая большую размерность множества и постоянные изменения на рынке ТСА, инженер проектировщик обычно ограничивается одним вариантом функциональной схемы.

Переход от структурной схемы к функциональной означает конкретизацию теоретического решения на базе некоторого комплекса технических средств. Если ТСА не единственно, а представлено конечным множеством блоков, то теоретически можно получить все возможные варианты функциональных схем:

$$m_CC \Rightarrow m_FC_1, m_FC_2, \dots, m_FC_N,$$

причем количество схем (N) будет зависеть только от объема БД ТСА.

Модели m_CC и m_FC различаются между собой степенью детализации описания объекта (схемы автоматизации). При переходе от модели m_CC к модели m_FC увеличивается как число блоков и связей, так и число характеристик, используемых в описаниях блоков.

В рамках каждой модели схема рассматривается как совокупность (n) контуров. Каждый контур состоит из цепей, которые могут в общем случае иметь разную длину.

$$Sh = \bigcup_{i=1}^n Kon_i, \quad (\forall i) Kon_i = \bigcup_{j=1}^{m_i} Ze_j, \quad \bigcap_{j=1}^{m_i} Ze_j \neq \emptyset \quad (1)$$

m_i - число цепей в i -ом контуре

Любая цепь, в свою очередь, состоит из одной или более связей.

$$Ze_j = S_{1,2}^j \circ S_{2,3}^j \circ \dots \circ S_{g,k_j}^j, \quad S_{r,t}^j = S^j(B_r, B_t) \quad B_r, B_t \in B \quad (2)$$

где: k_j – число связей в j -ой цепи (длина цепи), $S(B_r, B_t)$ – связь блока r -ого с блоком t -ым, B – множество блоков ТСА.

Порядок следования связей между элементами цепи строго определен, поэтому цепь представляет последовательность связей, а не совокупность. Каждую связь будем рассматривать как пару элементов (блоков), которые находятся в отношении “соединять”:

$$R^S = \langle \text{соединять}, x_i, y_r \rangle \quad (3)$$

где x_i – выход блока B_i , y_r – вход блока B_r .

Тогда отношение $(x_i R^S y_r)$ соответствует высказыванию “выход i -ого блока соединен с входом r -ого блока” и формируется связь $S(B_i, B_r)$. Каждую связь можно рассматривать как элементарный орграф, имеющий вершину-исток (Is) и вершину-сток (St). Для связи $S(B_r, B_t)$ соответственно $Is = B_r$, $St = B_t$.

Для построения отношения порядка (Ze) на множестве связей необходимо, чтобы для каждой пары связей выполнялось требование:

$$S_i(Is, St) \cap S_j(Is, St) \rightarrow St_i = Is_j \quad (4)$$

В множестве блоков ТСА (B) можно выделить три подмножества: измерительных преобразователей (ИП= B^1), исполнительных устройств (ИУ= B^3) и подмножество B^2

$$B^2 = B \setminus (B^1 \cup B^3) \quad (5)$$

Блоки из B^1 , B^3 выбирают исходя из требований технологов к вектору входов (выходов) объекта управления. Другие блоки (из B^2) выбирают с учетом характеристик B^1 , B^3 .

Операции выбора блоков ТСА для заданной структурной схемы относят к этапу разработки ФС, однако, при создании САПР <<FDcS>> было введено допущение:

- считать, что для каждого элемента цепи структурной схемы, выполнена операция назначения, определившая наименование блока ТСА.

Тогда при создании ФС задача выбора блоков для построения Z_e будет решаться на множестве B^2 , что приведет как к уменьшению размерности задачи перебора, так и к точному заданию истоков и стоков всех цепей ФС.

При введенных допущениях модель ФС будет отображением m_{CC} на множестве концептов ТСА, включенных в цепи ФС (рис.1).

Рассмотрим на примере взаимосвязь между моделями схемы. На 1-ом уровне проектирования структурную схему можно представить в виде графа ($G:: IP \rightarrow P \rightarrow IU$). Гомоморфное отображение этой структуры на таксономию "ТСА" позволяет получить модель m_{CC} и ее граф $G1:: (IP=ТСП) \rightarrow (P=БК1/Р130) \rightarrow (IU=МЭО)$. Между моделями G и $G1$ существует гомоморфизм. При разработке ФС на основе графа $G1$ формируется граф $G2$.

Очевидно, отображение вида $G2$ (или m_{FC}) построено правильно, если объединение $G1$ и $G2$ даст транзитивное замыкание $G2$ на вершинах из $G1$.

Рассмотрим формальную постановку задачи построения множества моделей функциональных схем на основе единственного варианта структурной схемы.

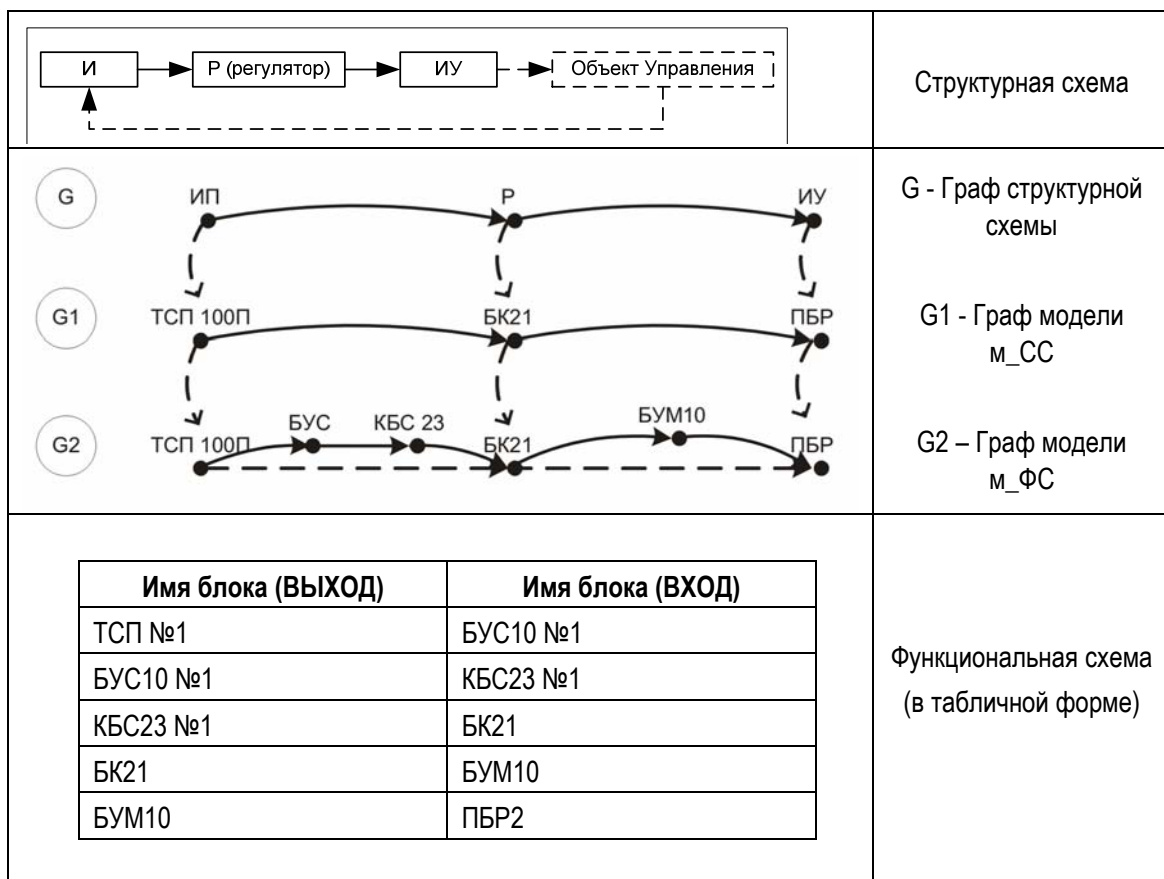


Рис.1. Связи между структурной и функциональной схемами и их моделями.

Пусть имеется модель m_{CC} , заданная n контурами, каждый контур состоит из m_i цепей (1). Введено ограничение на длину каждой цепи ($k_j=1$). Тогда вместо (2) каждая цепь вырождается в связь.

$$(\forall j) Z e_j = S^j(x_r, y_t), \quad B_r, B_t \in B \tag{6}$$

где: x_r – выходы блока B_r , y_t – входы блока B_t .

Каждый блок (B_i) может быть представлен множеством атрибутов, которое можно представить в виде нескольких подмножеств:

$$B_i :: A_i = (A_{i1}, A_{i2}, \dots, A_{ip}) = \bigcup_{p=1}^{pp} A_{ip} \quad (7)$$

В зависимости от размерности (pp) множества A_{ip} можно создавать различные по степени детализации описания технических блоков. Для иерархии схем автоматизации будет справедливо:

$$B_i^1 \subset B_i^2 \subset B_i^3, \text{ но тогда } A_i = \bigcup_{p=1}^3 A_{ip} \quad (\forall p) A_{ip} \subseteq A_i, \quad \bigcap_{p=1}^3 A_{ip} = \mathbf{0} \quad (8)$$

где: B_i^1, B_i^2, B_i^3 – (соответственно) описания i -ого блока для структурной, функциональной и принципиальной электрической схем.

При создании модели m_CC используют описания блоков (B_i^1) с минимальным числом атрибутов. В рассматриваемой постановке задачи множество A_{i1} , включает названия блока, модели и выпускающей фирмы.

Необходимо на основе модели m_CC вида (1,6,7) найти такое множество функциональных схем $\{m_FC_i\}$, в котором для каждой m_FC_i выполняются условия:

а) число и состав контуров не меняется при переходе от m_CC к любой m_FC_i ,

б) число цепей не меняется при переходе от m_CC к любой m_FC_i ,

в) для каждой связи (S_i), входящей в цепь из модели m_FC , должны выполняться ограничения на вид функциональных преобразований и уровень сигнала, которые обеспечивает указанная электрическая связь:

$$(Ogr_fun(B_r) = Ogr_fun(B_t)) \& (Ogr_DSign(B_r) \leq Ogr_DSign(B_t)) = A, \quad (9)$$

$$A \Rightarrow S(B_r, B_t) \in m_FC$$

г) каждая цепь функциональной схемы имеет одинаковые с соответствующей цепью в структурной схеме истоки и стоки:

$$Ze_i \in m_CC, \quad Ze_i = S_i(x_r, y_t), \quad Ze_i^* \in m_FC, \quad Ze_i^* = S_1 \circ \dots \circ S_k, \quad Ze, \quad S_1 \cap S_i = x_r, \quad S_k \cap S_i = y_t \quad (10)$$

Переход от m_CC к m_FC осуществляется путем последовательной проверки ограничений (9) для всех цепей CC . Возможны три варианта результатов проверки:

1) Если при анализе Ze вида (6) условие (9) выполняется, то $S^j(x_r, y_t)$ переносится из m_CC в m_FC без изменений.

2) Если для некоторой цепи Ze из m_CC условие (9) не выполняется. Тогда на множестве блоков B^2 необходимо найти такой B_l , что при его подстановке в (9) вместо B_r условия (9) будут выполняться. Тогда формируется новая связь $S^j(B_r, B_l)$. Модель m_FC принимает вид:

$$Ze_j = S^j(B_r, B_l) \circ SG(B_l, B_t) \quad (11)$$

Таким образом связь $S^j(x_r, y_t)$ из m_CC заменяется двумя новыми: скорректированной связью $S^j(B_r, B_l)$ и связью-гипотезой $SG(B_l, B_t)$.

Расщепление связи происходит по правилу: исток в скорректированной связи равен истоку в родительской связи, сток связи-гипотезы равен стоку в родительской связи:

$$Is_{rt} = Is^{kor} \quad St_{rt} = St^{gip} \quad (12)$$

Скорректированная связь включается в м_ФС, для гипотетической связи повторяются операции проверки условий (9) и при необходимости коррекции путем расщепления.

3) Если хотя бы для одной цепи после завершения цикла коррекции не удастся выполнить условия (9-10), то в необходимо вносить изменения в модель м_СС.

2. Система проектирования функциональных схем "FDcS"

Программное комплекс (ПК) системы проектирования функциональных схем сочетает в себе функции клиентской части системы управления реляционной базой данных, графического редактора структурных схем и клиентскую и серверную части реализации алгоритма синтеза функциональных схем.

Разработка программы выполнена на языке С++ с использованием методов объектно-ориентированного программирования, библиотеки стандартных шаблонов STL, мультиплатформенной С++ библиотеки для разработки графического интерфейса приложений (GUI) QT от Trolltech, динамически подключаемой библиотеки libpq.dll, поставляющей интерфейс с сервером базы данных PostgreSQL.

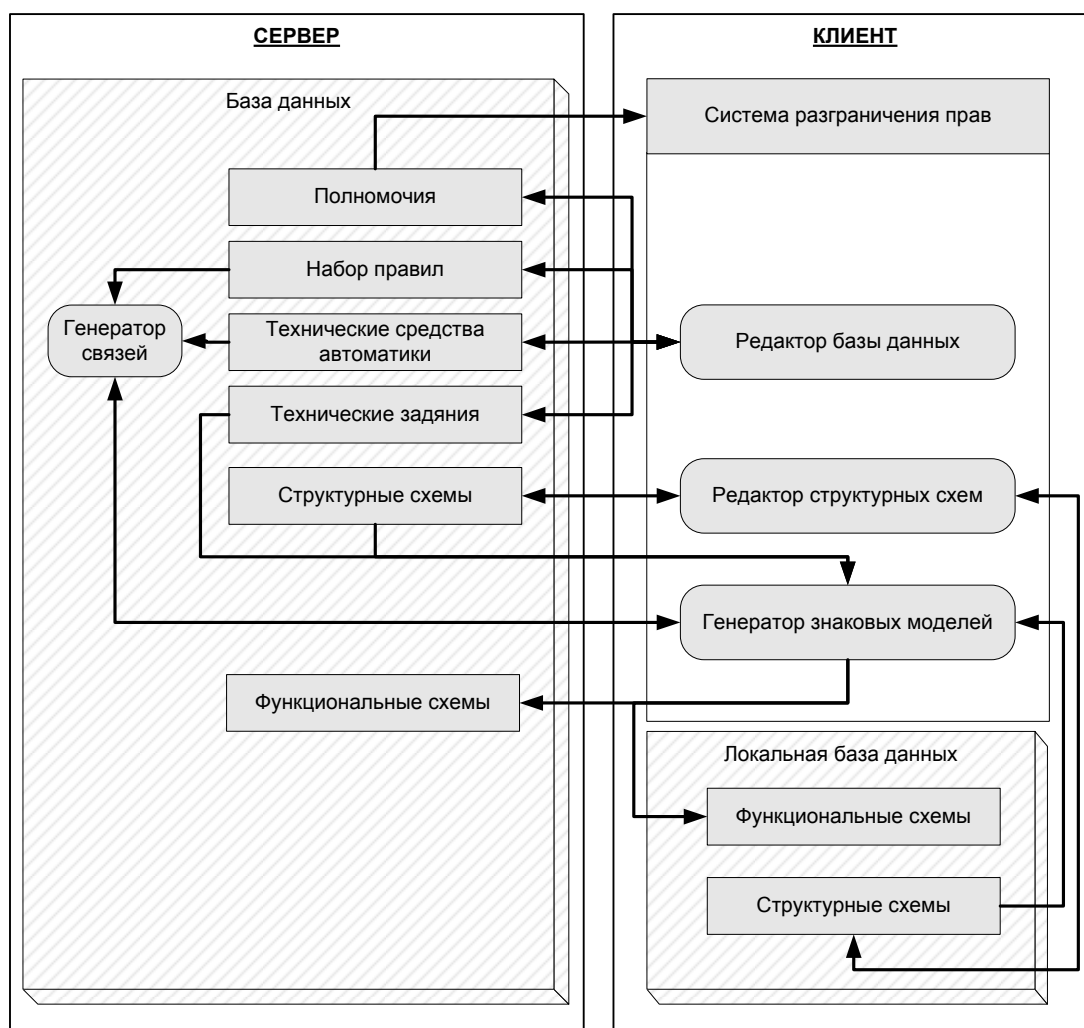


Рис.2. Архитектура системы проектирования функциональных схем "FDcS".

ПК имеет модульную, распределенную архитектуру, которая позволяет успешно функционировать в сетях различного масштаба, как в корпоративных (Intranet), так и в сети Internet. Допустима установка

программного обеспечения на одной рабочей станции, в этом случае на нее возлагаются дополнительные функции сервера баз данных.

Система распределения прав между группами пользователей позволяет защитить информацию в базе данных от действий неквалифицированных пользователей, а также ограничить доступ к административным функциям ПК. Архитектура системы представлена на рис.2.

Главным серверным компонентом ПК является база данных (БД). База данных включает декларативные и процедурные знания предметной области. Декларативную часть БД образуют данные по техническим средствам автоматизации, техническим заданиям, структурным схемам и сгенерированным функциональным схемам. К процедурной части БД относятся набор правил фокусировки поиска подстановок в продукции вида (9) и процедуры, образующие серверную часть модуля генерации знаковых моделей – генератор связей.

Клиентская часть ПК включает модули редактора БД, графического редактора структурных схем, генератора знаковых моделей и локальной БД структурных и функциональных схем. Все они работают под контролем модуля системы разграничения прав.

Модуль генератора знаковых моделей в качестве исходных данных получает макет технического задания с вербальным описанием требований к системе управления и ее структурной схемой. Синтез очередного варианта функциональной схемы осуществляется путем совместного взаимодействия модулей генератора знаковых моделей и генератора связей.

Ядром информационного обеспечения системы “FDcS” является база данных технических средств автоматизации - БД ТСА. Все множество ТСА по функциональному признаку разбиты на группы, составляющие иерархическую структуру (дерево) без множественного наследования [4]. В используемой реляционной базе данных (PostgreSQL) для хранения такого рода иерархии необходимо добавить в описание группы элементов дополнительное поле – ссылку на первичный ключ родителя.

Для каждого элемента указывается уникальный идентификатор, наименование, идентификатор группы, к которой принадлежит элемент (поле “tree_node”) и массив, хранящий полный путь (поле “path”) по иерархии групп начиная от корневой группы и заканчивая группой, к которой принадлежит элемент.

Дополнительное описание пути является избыточным, т.к. его всегда можно определить рекурсивно. Однако для быстрого выбора элементов, относящихся к определенной группе (не только указанной, но и всех ее наследников), рекурсивный способ не подходит из-за недостаточного быстродействия.

Скорость генерации цепей функциональной схемы определяется затратами времени на поиск допустимых подстановок в правило (9), выполняемых генератором связей. Для уменьшения времени генерации цепей при большом количестве элементов таксономии ТСА используются специальные правила фокусировки, сужающие область поиска элементов.

Введено два типа правил: ИСКАТЬ и НЕ ИСКАТЬ, которые задаются для пары групп элементов (G_1, G_2). Например, в соответствии с правилом <ИСКАТЬ(G_1, G_2)> при построении связи $S(B_r, B_s)$ если $B_r \in G_1$ или любой ее дочерней группы, поиск B_s для подстановки в (9) ограничивается на таксономии ТСА группой G_2 или любой ее дочерней группой. Правило <НЕ ИСКАТЬ((...), G_2)> позволяет вести поиск допустимых подстановок в (9) среди всего множества элементов, за исключением элементов указанной группы (G_2).

Приведем пример использования правил. Для указанных узлов иерархии групп элементов

- измерительные преобразователи (ИП)
- └─ нормирующие преобразователи (НП)
- └─ первичные преобразователи (ПП)

можно ввести правило RULE 2 :: <ИСКАТЬ(ПП, НП)>. Интерпретация RULE 2 :: Поиск дополнительных устройств для подключения к первичным преобразователям (датчикам) необходимо вести только в группе

нормирующих преобразователей, преобразующих сигналы датчиков в сигналы унифицированных диапазонов.

Другой важной проблемой генерации функциональных схем является резкое увеличение числа возможных вариантов схем при увеличении размерности множества блоков ТСА. Для решения этой проблемы в системе "FDcS" используются две стратегии: ограничения общего количества требуемых функциональных схем и исключение из процесса генерации технически не реализуемых схем.

При большом количестве заданных в БД блоков ТСА возможен случай, когда в процессе генерации цепи происходит многократная коррекция связи путем расщепления, для чего может использоваться один и тот же блок ТСА. Это может привести к заикливанию и образованию цепочек связей вида В1–В2–В3–В1–В2–В3–В1... В тоже время известно, что на практике цепи, состоящие из пяти и более связей, обычно не используются из-за низкой надежности, высокой сложности и стоимости. Для исключения заикливания введено ограничение на количество связей в цепи (m – настраиваемый параметр). При обнаружении в цепи более (m) связей процесс генерации цепи прекращается, а сама схема помечается как технически не реализуемая.

Заключение

Создание системы Functional Design of a Control System (FDcS) доказывает возможность перехода от параметрического синтеза структурных схем, отражающих динамику АСУТП, к автоматическому структурному синтезу функциональных схем в ряде измерительных и управляющих цепей. Модели, положенные в основу построения баз данных и знаний являются базой для промышленной технологии начальных стадий автоматизированного проектирования АСУТП.

Библиография

- 1.Норенков И.П., Введение в автоматизированное проектирование технических устройств. Москва, 2005.
- 2.Цветков В.Д. Системно-структурное моделирование и автоматизированное проектирование технологических процессов. Минск. 1990.
- 3.Филатова Н.Н. Автоматическое формирование знаний в САПР. Тверь. 1996.
- 4.О.Л.Ахремчик. Программная система функционального проектирования аппаратных комплексов АСУТП. / Программные продукты и системы, N2. 2007.

Информация об авторах

Филатова Наталья Николаевна - профессор, Тверской государственный технический университет, 170026, Ст. Никитина 22, Тверь, Россия; e-mail: nfilatova99@mail.ru

Ахремчик Олег Леонидович - доцент, Тверской государственный технический университет, 170026, Ст. Никитина 22, Тверь, Россия

Куприянов Олег Владимирович - магистр, Тверской государственный технический университет, 170026, Ст. Никитина 22, Тверь, Россия; e-mail: aven@tvcom.ru

60th Anniversary of



INSTITUTE OF MATHEMATICS AND INFORMATICS of Bulgarian Academy of Sciences

Acad. G. Bonchev Str., block 8, Sofia 1113, Bulgaria

Tel. (+359-2) 979-3824, Fax (+359-2) 971-3649

<http://www.math.bas.bg>

The Institute of Mathematics and Informatics (IMI) at BAS was founded in 1947 as Institute of Mathematics. At the beginning about ten research fellows were working at the Institute. In 1961 a computational centre was established as part of the Institute. Later specialist in Mechanics also worked at the Institute, hence and it was named Institute of Mathematics and Mechanics. Its present name dates from 1995

The Institute has considerable achievements in the field of Mathematics that are not discussed here.

The development of the Informatics in Bulgaria started at the Institute. Many researchers have built the career of Informatics specialists.

The Institute was the first in Bulgaria to buy an universal analog computing machine MH-7. The first Bulgarian computer was created at the Institute. Soon after that came into exploitation the first imported into Bulgaria computer "MINSK-2". An original software for this computer – auto code "MIKOD", operation systems "MID" and "MID-2", a system for symbol programming "MIKS" and a rich library of programs were created here as well.

The fellows of the Institute also carried out the first Informatics researches in Bulgaria. The Institute has a wide range of activities in Applied Informatics and it continues to produce original software for the solving important problems. Researchers from the Institute organized and taught the first courses in Informatics at the Sofia University "St. Kliment Ohridski" for students in Mathematics. In a short time a major in Informatics was launched with the help of the Institute and later on it became a speciality at the Sofia University. Researchers of the Institute prepared the first syllabus, textbooks, and manuals. The staff of the Institute is also involved in training teachers in Informatics for the secondary school.

In the course of the years the informaticians at focused upon the research activities and many of them are still lecturing Informatics at a number of Bulgarian universities.

Departments of IMI : *Algebra; Artificial Intelligence; Biomathematics; Complex Analysis; Differential Equations; Education in Mathematics and Informatics; Geometry and Topology; Information Research; Laboratory of Mathematical Linguistic; Logic; Mathematical Foundations of Informatics; Mathematical Linguistics; Mathematical Physics; Computational Mathematics; Operation Research; Probability and Statistics; Real and Functional Analysis; Software Engineering; Telecommunications Department.*

15th Anniversary of



ASSOCIATION OF DEVELOPERS AND USERS OF INTELLIGENT SYSTEMS

ADUIS consists of about one hundred members including ten collective members. The Association was founded in Ukraine in 1992. The main aim of **ADUIS** is to contribute to the development and application of the artificial intelligence methods and techniques. The efforts of scientists engaged in **ADUIS** are concentrated on the following problems: expert system design; knowledge engineering; knowledge discovery; planning and decision making systems; cognitive models designing; human-computer interaction; natural language processing; methodological and philosophical foundations of AI.

Association has long-term experience in collaboration with teams, working in different fields of **research and development**. Methods and programs created in Association were used for revealing regularities, which characterize chemical compounds and materials with desired properties. Some thousands of high precise prognoses have been done in collaboration with chemists and material scientists of Russia and USA.

Association can help **businessmen** to find out conditions for successful investment taking into account region or field peculiarities as well as to reveal user's requirements on technical characteristics of products being sold or manufactured.

Physicians can be equipped with systems, which help in diagnosing or choosing treatment methods, in forming multi-parametric models that characterize health state of population in different regions or social groups.

Sociologists, politicians, managers can obtain the Association's help in creating generalized multi-parametric "portraits" of social groups, regions, enterprise groups. Such "portraits" can be used for prognostication of voting results, progress trends, and different consequences of decision making as well.

Association provides a useful guide in technical diagnostics, ecology, geology, and genetics.

ADUIS has at hand a broad range of high-efficiency original methods and program tools for solving analytical problems, such as knowledge discovery, classification, diagnostics, prognostication.

ADUIS unites the creative potential of highly skilled scientists and engineers

Since 1992 **ADUIS** holds regular conferences and workshops with wide participation of specialists in AI and users of intelligent systems. The proceedings of the conferences and workshops are published in scientific journals.

ADUIS cooperates through its foreign members with organizations that work on AI problems in Russia, Byelarus, Moldova, Georgia, Bulgaria, Czechia, Germany, Great Britain, Hungary, Poland, etc. **ADUIS** is the collective member of the European Coordinating Committee for Artificial Intelligence (ECCAI).

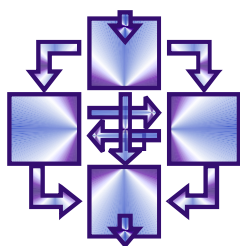
Products developed by ADUIS: **Confor**: Tools for Knowledge Discovery, Classification, Diagnostics and Prediction; **Analogy**: Tools for Solving Problems on the Basis of Analogy; **Manager**: Tools for Decision Support Systems Design; **Discret**: Tool for Discretization of Numerical Data; **Gobsec**: The System for Investment Scheduling.

For contacts: V.M. Glushkov Institute of Cybernetics; National Academy of Science of Ukraine;

Prospect Akademika Glushkova, 40, 03680 GSP Kiev-187, Ukraine;

Phone: (380+44) 5262260; Fax: (380+44) 5263348; E-mail: glad@aduis.kiev.ua

10th Anniversary of



Association for the Development of the Information Society

Acad. G. Bonchev St., block 8, Sofia 1113, Bulgaria
Tel. (+359-2) 979-3813, -3808, Fax (+359-2) 739-808
e-mail: ario@math.bas.bg, adis@einet.bg
<http://www.adis.org>

The Association for the Development of the Information Society (ADIS) was established in April 1997 and is an independent, non-government, non-profit organization with the non-commercial objective to support the development of the information society in Bulgaria. This objective is extensively defined in the Association's statute and includes:

- Interaction with individuals and organizations working for the development of the information society in Bulgaria and in the world.
- Support of the comprehensive utilization of the capacity of the information infrastructure and information technologies by all layers of society and all ages and professions, as well as by unemployed, ethnic minorities, people with disabilities, etc.
- Development and implementation of national and international projects whose goal is establishing, developing, and governing the information society.
- Participation in the elaboration and implementation of educational, promotional, and demonstration programs dedicated to information society issues.
- Participation in international activities on issues of the development of the information society, and maintenance of ties to and interaction with foreign and international organizations.
- Organization of conferences, forums, workshops dedicated to the information society.
- Publishing of a newsletter distributed among the individual and collective members of the Association.

Besides individual persons, the Association has as collective members from various regions of Bulgaria: Plovdiv University 'Paisii Hilendarski', Technical University—Gabrovo, the Police Academy, the Institute of Mathematics and Informatics, the Institute of Information Technologies, the Central Laboratory of Computer Security of the Bulgarian Academy of Sciences (Sofia), and other organizations. Societies in the cities of Plovdiv, Shoumen, and Bourgas have been formed as autonomous subsidiaries of the Association. Its membership and associated structures are growing quickly and already include foreign members. The Association has existed since recently but it unites people and organizations with several decades of experience in the field of computer science and information technologies. Since 1999, the Association has organized monthly national seminars in the framework of the Forum Global Information Society. The seminars are devoted to the development of the information society in all fields of the human activities and aspects. Other activities include implementing a project for training disabled (deaf) people to use computers and the Internet, a project for training secondary school teachers in a broad range of computer technologies, participation in the drafting of the Bulgarian national strategy for the Information Society, drafting of models and principals for creating, management and development of public centers for access to Internet, information and communication services and public e-information and e-services for the Bulgarian citizens as well as delivering of talks on Information Society issues at various national and regional events by members of the Association.

The Association gladly welcomes contacts with organizations from abroad whose activities are related to the development of the global information society.

15th Volume of



International Journal "Information Theories and Applications"

Verba volant, scripta manent !

International Journal "Information Theory and Applications" (IJ ITA) has been established in 1993 as independent scientific printed and electronic media. IJ ITA is edited by the Institute of Information Theories and Applications FOI ITHEA in collaboration with the Institute of Cybernetics "V.M.Glushkov", NASU (Ukraine) and Institute of Mathematics and Informatics, BAS (Bulgaria).

During the years, IJ ITA became as well-known international journal. Till now more than 600 papers have been published. IJ ITA authors are widespread in 39 countries all over the world: Armenia, Belarus, Brazil, Belgium, Bulgaria, Canada, Czech Republic, Denmark, Egypt, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Israel, Italy, Japan, Kirghizia, Latvia, Lithuania, Malta, Mexico, Moldavia, Netherlands, Poland, Portugal, Romania, Russia, Scotland, Senegal, Serbia and Montenegro, Spain, Sultanate of Oman, Turkey, UK, Ukraine, and USA.

IJ ITA major topics of interest include, but are not limited to:

INFORMATION THEORIES

<i>Artificial Intelligence</i>	<i>General Information Theory</i>
<i>Computer Intellectualisation</i>	<i>Hyper Technologies</i>
<i>Intelligent Networks and Agents</i>	<i>Information Models</i>
<i>Intelligent Technologies</i>	<i>Intellectualisation of Data Processing</i>
<i>Knowledge Discovery and Engineering</i>	<i>Knowledge-based Society</i>
<i>Knowledge Acquisition and Formation</i>	<i>Logical Inference</i>
<i>Distributed Artificial Intelligence</i>	<i>Natural language Processing</i>
<i>Models of Plausible Reasoning</i>	<i>Neuroinformatics</i>
<i>AI Planning and Scheduling</i>	<i>Philosophy and Methodology of Informatics</i>
<i>Bioinformatics</i>	<i>Quality of the Programs</i>
<i>Cognitive Science</i>	<i>Software Engineering</i>
<i>Decision Making</i>	<i>Theory of Computation</i>

APPLICATIONS

<i>Communication Systems</i>	<i>Multimedia Systems</i>
<i>Computer Art and Computer Music</i>	<i>Programming Technologies</i>
<i>Hyper Technologies</i>	<i>Program Systems with Artificial Intelligence</i>
<i>Intelligent Information Systems</i>	<i>Very Large Information Spaces</i>

More information about the IJ ITA rules for preparing and submitting the papers as well as how to take out a subscription to the Journal may be obtained from www.foibg.com/ijita.

Second Volume of



International Journal "Information Technologies and Knowledge"

Intelligo ut credam !

International Journal "Information Technologies and Knowledge" (IJ ITK) has been established in 2007 as independent scientific printed and electronic media. IJ ITK is edited by the Institute of Information Theories and Applications FOI ITHEA in collaboration with the Institute of Cybernetics "V.M.Glushkov", NASU (Ukraine); Institute of Mathematics and Informatics, BAS (Bulgaria); Institute of Information Technologies, BAS (Bulgaria); University of Hasselt (Belgium); Natural Computing Group (NCG) of the Technical University of Madrid (Spain); Astrakhan State Technical University (Russia); Taras Shevchenko National University of Kiev (Ukraine); University of Calgary (Canada); VLSI Systems Centre of Ben-Gurion University (Israel).

The main scope of the IJ ITK covers but is not limited to the theoretical research, applications and education in the area of the Information Technologies for:

- Knowledge Collecting and Accumulation
- Knowledge Discovery and Acquisition
- Knowledge Level Modeling
- Knowledge Management -Transfer and Distributing
- Knowledge Market
- Knowledge Representation and Processing
- Knowledge Utilization
- Knowledge-based Society
- Knowledge-based Systems

Many scientific and practical areas are connected to the topics of interest of IJ ITK:

- Business Informatics: e-Management, e-Finance, e-Commerce, e-Banking,
- Business Intelligence: Methodology, Tools and Technologies, Analytics and Statistics;
- Cognitive science
- Competitive Intelligence;
- Data Mining
- Decision Making
- e-Management in Governmental and Municipal Structures: Models, Systems, e-Government, etc.
- Information Technologies in Biomedicine
- Intelligent Communication Technologies and Mobile Systems
- Intelligent Robots
- Intelligent Systems
- Intelligent Technologies in Control and Design
- Modern (e-) Learning Information Technologies
- Multimedia Semantic Systems
- P2P e-Learning Applications
- Planning and Scheduling
- Socio-cognitive engineering
- Technology and Human Resource Issues
- Technology-based Blended, Distance and Open Learning
- Web-based Technologies and Systems, AI/Semantic Web
- etc.

More information about the IJ ITK rules for preparing and submitting the papers as well as how to take out a subscription to the Journal may be obtained from <http://www.foibg.com/iitk>.