Information technology for determining structure of social group based on fuzzy c-means

Oksana Mulesa, Fedir Geche, Anatoliy Batyuk

Abstract – **Task of determination structure of social groups was considered as the task of fuzzy clustering. The adaptation of fuzzy c-means was done to solve the problem. Analysis of features design information-analytical system for solving the problem was done. Functional scheme IAS were developed.**

Keywords – **Information technology, Fuzzy c-means, Fuzzy clustering, Linguistic variable, Structure of social group, Information-analytical system.**

I. Introduction

The social structure of society is a set of interrelated and interacting social groups. Studying its properties can solve important social, political, economic, medical and other problems. Results of studying the structure of social groups is the basis for the prediction of the emergence and development of various social processes, planning and implementation of preventive and precautionary actions to prevent negative phenomena in society in general and in the community including forecasting the development of markets for goods and receiving labor market characteristics and so on. Quantitative and qualitative composition of the social group are among the parameters that characterize its structure. The problem of the division of society into social groups is an important task among of problems that arise in the study of social structure. It can be mathematically represented as a clustering problem.

II. Clustering problem

The task of clustering is to identify groups of objects that are closest to each other on some criteria.

The mathematical formulation of the clustering problem is such [1]:

Let X is the set of objects, Y is the set with numbers (names, labels) of clusters; $d(x,x')$ is the given function of the distance between objects; $X^m = \{x_1, x_2, ..., x_m\} \subset X$ is the final selection of objects. We must split the sample into subsets which do not intersect, called clusters, so that each cluster consisting of objects similar in metric d and objects of different clusters were significantly different. Also, each object $x_i \in X^m$ is given a number.

Oksana Mulesa – Uzhgorod National University, Pidgirna Str., 46, Uzhgorod, 88000, UKRAINE,
E-mail: mulesa.oksana@gmail.com
Fedir Geche – Uzhgorod National University, Pidgirna Str., 46, Uzhgorod, 88000, UKRAINE,
E-mail: fgeche@hotmail.com
Anatoliy Batyuk — ACS Department, Lviv Polytechnic National University, S. Bandery Str., 12, Lviv, 79013, UKRAINE,
E-mail: abatyuk@gmail.com

Known methods among of the clustering methods are statistical cluster analysis methods [1], [2], tree-like methods [3], neural network [4], [5] and genetic algorithms [6], [7].

III. Problem of determining structure of social groups as the task of fuzzy clustering

Verbal formulation of the problem of determining the structure of social groups can be done so [8]: to execute grouping of members on socio-demographic characteristics in a given social group; determine the quantity composition and representatives of each component of a given group. Representatives of the group are characterized by the values of attributes included in their socio-demographic profile. These characteristics usually include gender, age, educational level, professional activity, marital status, social level, religious affiliation and more.

Criteria for clustering persons depends on the problem that is solved. This can be a clustering persons for their consumer preferences, by purchasing power, the nature of social behavior and so on. The difficulty of making such a clustering is the fact that there are cases where some persons can simultaneously belonging to different groups. Therefore, we present the challenge for the problem of fuzzy clustering, which will build a mathematical model as follows:

Let $X = \{X_i = (x_{i1}, x_{i2}, ..., x_{iM}), i = \overline{1,N}\}$ is the set of objects, which are characterized by multiple attributes $K = \{K_1, K_2, ..., K_M\}$, that is x_{ij} is the value of j th attribute for i th object number. We must to divide this set X on G fuzzy clusters for a given criterion. That is, we have to build an algorithm to determine the degree of membership of object to each of the clusters:

$$X_i \xrightarrow{\Theta} (\mu_1(X_i), \mu_2(X_i), ..., \mu_G(X_i)),$$

where $\mu_g(X_i)$ is the degree of membership of X_i in the cluster g, $g = \overline{1,G}$, and $\sum_{g=1}^{G} \mu_g(X_i) = 1$.

The method of fuzzy c-means based on the theory of fuzzy sets. The basis of this method is the position that

each object may belong to different clusters with different values of membership function [9].

IV. ADAPTATION OF THE FUZZY C-MEANS

At the beginning of the procedure clustering we should solve the following problems [8]:

1. *The task of ranking attributes*. Verbal formulation of the problem can be done so: we must for each attribute from set K to indicate the impact degree to the possibility of membership to clusters.

Mathematically, the task consists in identify the functional dependencies Ω, for the implementation mapping

$$\Omega: K \to \Re^+,$$

and determination the impact of attributes on the formation of clusters.

One way of calculating the function Ω values is based on the results of expert surveys. It is reduced to consecutive solutions of tasks numerical evaluation of facility. Among the methods for determining the numerical evaluation of the object are important such as the statistical method, utilitarian and egalitarian methods, heuristic methods etc [10]. Let $\alpha_j = \Omega(K_j)$, $j = \overline{1,M}$.

2. *The task of constructing linguistic variables*. We build linguistic variables for each attributes from the set K [11]. Construction can be made based on the findings of experts. Denote by $\langle I_j, T_j, Q_j, S_j, P_j \rangle$ parameters that correspond to linguistic variable K_j, $j = \overline{1,M}$; R_j is the the power of the set T_j; $\mu_{rT_j}(x_j)$ is the membership function for r th therm from set T_j, $r = \overline{1,R_j}$.

After solving these tasks, we turn to run iterative procedure of the adapted method of fuzzy c-means. The algorithm of method is as follows:

Step 1. We set the parameter for stop the algorithm $\varepsilon \in (0;1)$ and the degree of fuzziness m, $1 \le m < \infty$. When $m = 1$ we get a partition in which each object belongs to only one cluster. When we increase the parameter m, degrees of fuzziness for selected clusters are increased.

We define randomly the degree of membership to each cluster for each object from the set $X = \{X_i = (x_{i1}, x_{i2}, \ldots, x_{iM}), i = \overline{1,N}\}$ and form the matrix $U^0 = (u_{ig})$. $u_{ig} \in [0;1]$, where u_{ig} is starting degree of membership for i th object on g th cluster $(i = \overline{1,N}, g = \overline{1,G})$, and $\sum_{g=1}^{G} u_{ig} = 1$, $\forall i \in \{1,\ldots,N\}$.

We define the function of the distance between two objects of $X_{i_1}, X_{i_2} \in X$ so:

$$d(X_{i_1}, X_{i_2}) = \frac{\sum_{j=1}^{M} \left(\alpha_j \cdot \frac{1}{R_j} \sum_{r=1}^{R_j} \left| \mu_{rT_j}(x_{i_1 j}) - \mu_{rT_j}(x_{i_2 j}) \right| \right)}{\sum_{j=1}^{M} \alpha_j} \quad (1)$$

where $x_{i_1 j}, x_{i_2}$ are relevant components of vectors X_{i_1}, X_{i_2}.

Step 2. We determine the value of membership functions for the components of vectors that are current cluster centers as follows:

$$\mu_{rT_j}(c_j^{(g)}) := \frac{\sum_{i=1}^{N} (u_{ig})^m \cdot \mu_{rT_j}(x_{ij})}{\sum_{i=1}^{N} (u_{ig})^m},$$

where $C^{(g)} = (c_1^{(g)}, c_2^{(g)}, \ldots, c_M^{(g)})$ is current center of g th cluster ($g = \overline{1,G}$).

Step 3. We find the distance from the object to the cluster centers by Eq. (1). Let $d_{ig} = d(X_i, C^{(g)})$, $i = \overline{1,N}$, $g = \overline{1,G}$.

Step 4. We calculate the matrix coefficients as follows: if $\exists g_1 \in \{1,2,\ldots,G\}$: $d_{ig_1} = 0$, then $u_{ig_1} = 1$, $u_{ig} = 0$, $\forall g = g_1$, $g = \overline{1,G}$;

if $\forall g = \overline{1,G}$ $d_{ig} > 0$, then $u_{ig} = \dfrac{1}{\sum_{l=1}^{G} \left(\frac{d_{ig}}{d_{il}} \right)^{\frac{2}{m-1}}}$.

Step 5. We test conditions for stop the algorithm. Viz, if $\|U - U^*\|^2 < \varepsilon$, where U^* is the matrix that was calculated on the previous step, then the algorithm is ended, else we will go to the Step 2.

V. FEATURES OF DESIGNING INFORMATION-ANALYTICAL SYSTEM

Models and methods, which were proposed in this research, together with developed information-analytical system, will be united in the corresponding information technology.

Development of information-analytical system for solving the problem of determining the structure of the social group, according to the algorithm involves implementing the following tasks:

– forming and ranking set of signs describing the person;

– fuzzification of input data;

– partition initial set on clusters.

With this in mind, on the design phase of information-analytical system is necessary to solution of such problems:

– forming the list of attributes that influence the entry an object to a cluster; forming the set of possible values for each attribute;

– construction of expert procedures for ranking of attributes;

– construction of procedures for formation of linguistic variables for each of the selected attributes;

– software development for preprocessing of input data;

– development of analytical unit of information-analytical system.

Development of information-analytical system must be made according to levels of system models:

goals => problems (models) => methods (algorithms) => tools.

A set of mathematical models and methods are formed the basis of the information-analytical system. Among them are the following:

– models and methods of determining the competence of experts;

– models and methods of determining the numerical evaluation of the object based on expert interviews;

– indirect methods of construction of membership functions of fuzzy sets based on the opinions of experts;

– models and methods of fuzzy clustering.

Formally functional diagram of information-analytical system is shown in Fig. 1.

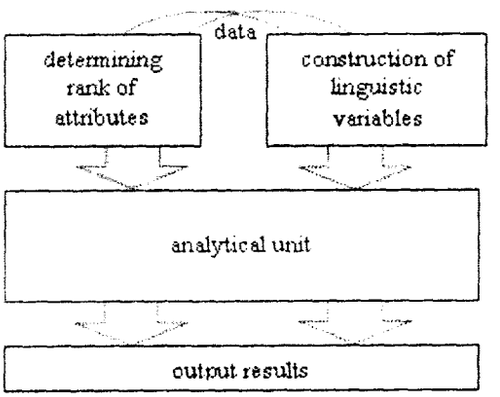

Fig.1. Functional diagram of information-analytical system

VI. Summary and Conclusion

The work is devoted to development of information technology for determine the structure of social groups on the basis of fuzzy c-means. The following steps were implemented for the design of effective technology:

– The problem of determining the structure of social groups had been studied. We have analyzed the problem and demonstrated its features associated with character of input data.

– Verbal and mathematical formulation of the problem of determining the structure of social groups as fuzzy clustering tasks were done. It provides the ability for the execution of the partition into clusters, in which each object can simultaneously belong to several clusters with different degrees of membership.

– The adaptation of fuzzy c-means was developed for solving formulated problem.

– The features of designing information-analytical system for solving the problem of determining the structure of social groups had been found; it functional scheme was developed. The information technology has been tested for a social group "Migrant workers" as a group with high risk of HIV/AIDS. The numerical experiments have confirmed its effectiveness.

References

[1] Dzh. Tu and R Honsales, "The principles of pattern recognition", *"Myr"*, Moscow, 412 p., 1978.

[2] K. Fukunaha, "Introduction to statistical pattern recognition theory", *Nauka. Hlavnaia redaktsyia fyzyko-matematycheskoi lyteratury*, Moscow, 368 p., 1979.

[3] V. Pliuta, "Comparative multivariate analysis in econometric modeling", *Finansy i statystyka*, Moscow, 175 p., 1989.

[4] T. Kohonen, "Self-organization and associative memory", *Springer Verlag, New-York, 2d. ed.*, 312 p. 1988.

[5] A. N. Gorban, A. Yu. Zinovyev, "Method of Elastic Maps and its Applications in Data Visualization and Data Modelling". *Int. Journal of Computing Anticipatory Systems, CHAOS, Vol. 12,* pp. 353-369. 2002.

[6] V. Ye. Snytiuk, "Forecasting. Models. Methods. Algorithms: Tutorial" *"Maklaut"*,Kyiv, 364 p., 2008.

[7] V. Snytiuk, "Evolutionary clustering of complex objects and processes", *XI-th International Conference «Knowledge-Dialogue-Solution», Varna, T. 1,* pp. 232-237., 2005.

[8] O. Yu. Mulesa, "Adaptation of fuzzy c-means method for determination the structure of social groups". *Technology audit and production reserves*, [S.l.], v. 2, n. 2(22), pp. 73-76, apr. 2015. ISSN 2312-8372. doi:http://dx.doi.org/10.15587/2312-8372.2015.41014.

[9] Yu. V. Smeshko, "On a criterion to select the exponential weight classification algorithm of fuzzy C-means", *Molodezh y nauk, Krasnoyarsk,* pp. 19-25. 2012.

[10] O. F. Voloshyn, S. O. Mashchenko, "Decision theory: tutorial", *Publishing and Printing Center "Kyivskyi universytet"*, Kiev, 366 p., 2010.

[11] S. A. Orlovskyi, "Decision making with fuzzy initial information", *Nauka. Hlavnaia redaktsyia fyzyko-matematycheskoi lyteratury*, Moscow, 208 p., 1981.