The Method of the Scientific Directions Potential Forecasting in Infocommunication Systems of an Assessment of the Research Activity Results

Andrii Biloshchytskyi
Department of technology management
Taras Shevchenko National University of Kyiv, KNU
Kyiv, Ukraine
bao1978@gmail.com

Alexander Kuchansky,
Department of cybersecurity and computer engineering
Kyiv Nat. Univ. of Construction and Archit., KNUCA
Kyiv, Ukraine
kuczanski@gmail.com

Yurii Andrashko
Department of System Analysis and Optimization Theory
Uzhhorod National University, UzhNU
Uzhhorod, Ukraine
yurii.andrashko@uzhnu.edu.ua

Svitlana Biloshchytska
Department of informat. technol. designing and applied math Kyiv Nat. Univ. of Construction and Archit., KNUCA
Kyiv, Ukraine
bsvetlana2007@ukr.net

Anastasiia Dubnytska, Vladimir Vatskel
Information technologies center
Kyiv Nat. Univ. of Construction and Archit., KNUCA
Kyiv, Ukraine
anastasiakuziomko@gmail.com, v.vatskel@it-lynx.com

Abstract—The method of the scientific directions potential forecasting in Infocommunication systems of an assessment of the research activity results is described. Formulas for calculation of forecasting for the partition of the set of scientific directions are given. Also assessments of scientific activity of scientists for previous periods are taken into account in the calculation. Linearly-weighed moving average model is applied to create the forecast of potential of the scientific directions. The method can be used for identification of the perspective directions of researches, which are formed in the scientific environment.

Keywords—potential forecasting; directions of scientific research; assessment of the research activity results.

I. INTRODUCTION

According to the Solow model of economic growth in stationary form, growth rate of labor productivity is equal to growth rate of scientific and technical progress, and rate of economic growth is equal to the sum of rate of the scientific and technical progress and growth rate of the population [1]. That is ensuring economic growth of the state depends on scientific and technical progress, in turn is directly connected with quality and an orientation of scientific research. Therefore, an important task for the state, the private companies, which use scientific research, is the creation of effective criteria for evaluation of the research activity results of the scientific and scientific directions in whom these scientists are united.

One of an assessment of quality components of research work of the scientific direction is the assessment of the main results of this work: scientific publications and citings on them. For this purpose it is necessary to estimate the level of citing specific authors, and then, considering scope of their researches, to carry out a procedure of a classifying for these scientific directions. Each scientific direction will include in that case a certain number of authors. Potential of the scientific direction to be defined by an assessment of research activity results of scientists, which belong to this direction. As in the scientific directions authors who actively are engaged in scientific activity are integrated and publish their researches, the scientific directions can be considered as dynamic objects, which have the history of development, and potential, that is those opportunities and means, which can be used for further development. History and current state of development of the scientific directions can be defined on the basis of indicators which are freely provided on the Internet. These are, first of all, indicators of citing publications of the authors belonging to certain directions: Hirsh's index, i10-index, g-index, PR-q etc.

Research problem is the creation of a method for calculation of the short-term forecast of potential of the scientific directions. It is supposed that the method can be used for identification of the perspective directions of researches, which are formed in the scientific environment. It will allow exercising operational management of various aspects of the concrete scientific directions at the level of the state, to attract the international grants, to create consortia for carrying out a scientific and educational research and so forth. As the scientific direction, as an object, dynamically changes and these changes are reflected in one form or another on the Internet,

978-1-5386-0983-5/17/$31.00

functioning of this object can happen in some Infocommunication system.

The Infocommunication system of an assessment of the scientific activities results represents a combination of:

- the web application which realizes search, storage and processing of large volumes of information on publishing activity of scientists from the Internet;
- information processing methods (classification methods for partition of scientists in the appropriate scientific directions, methods of creation of scalar and vectorial estimates of research and development activities results, forecasting methods of potential of the directions, etc.);
- information sources (information on citing scientists from scientometric databases);
- customers of information (the higher educational institutions, scientific research institutes, the private companies which are responsible for a scientific policy).

This research is continuation of researches on the analysis of research activity, which have been published in [2 - 5]. In particular, in work [2] it is described a selective method of short-term forecasting of temporary ranks on the basis of comparison of templates. In work [3] methods of finding of scalar estimates, in particular the PR-q index, and also methods of creation of an integrated vector assessment for determination of research activity results of scientists are described. These methods can be used as an alternative to the known indexes of citing, such as h-index, g-index etc. The conceptual model of finding of similarity between fragments of text information [4] and tabular data of various types [5] can be used for detection of similarity between the direction of scientific research of authors and their distributions in the corresponding scientific directions. Also for forecasting of results of research activity indistinct expert methods [6] can be used. In work [7] is considered comparisons of a Hirsh's index from a g-index taking into account various selections of scientific and scientific collections. In work [8] ways of establishment of functional dependence between the number of publications and the number of quotes are offered. For carrying out the effective analysis of printing activity, it is necessary to understand organizational aspects of functioning of the scientific environment and educational institutions where this environment is formed. In particular, in work [9] it is described model of an assessment of quality of functioning of educational institutions. This model demands tracking of scientific activity results changing , and also forecasting of this activity results. In work [10] the analysis is carried out and the structure of interaction of the main subjects as carriers of knowledge is synthesized.

II. THE PRIMARY RESEARCH MATHERIALS

Assume that $A=\{a_1,a_2,\ldots,a_n\}$ is a set of scientists, n is the number of scientists, and $P=\{p_1,p_2,\ldots,p_m\}$ is the set of articles published by the given scientists, m is the number of publications. Mark the set of all publications by a scientist a_i, $i=\overline{1,n}$ trough

$$P(a_i)=\left\{p_j\in P\,\middle|\,(a_i,p_j)\in U\right\},\ i=\overline{1,n},\ j=\overline{1,m},\quad (1)$$

where the set $U\subset A\times P$ reflects the authorship of the scientist a_i for the publications p_j.

Mark the set of publications in which the publications are cited scientist a_i for $i=\overline{1,n}$ та $j=\overline{1,m}$ through

$$C(a_i)=\left\{p_j\in P\,\middle|\,(p_i,p_y)\in C, p_y\in P(a_i), y=\overline{1,m}\right\},\quad (2)$$

where the set $C\subset P\times P$ reflects the set of publication in which the publications are cited scientist a_i.

Assessment of the research activities results of scientists a_i presented as function

$$q_i=Q(P(a_i),C(a_i)),\ i=\overline{1,n}\quad (3)$$

where the values of Q correspond to the best functional results of research relevant scientist a_i..

We will designate a set of publications in which the publication p_j is quoted through

$$C(p_j)=\left\{p_j\in P\,\middle|\,(p_y,p_j)\in C, y=\overline{1,m}\right\}.\quad (4)$$

The q_i estimates for these reasons can be determined on the basis of h-, g-, e-, i10-indexes, PR-q index, etc. Or may be an integral assessment that combines the results of evaluating scientific activity by different methods.

We will designate through $H=\{\eta_1, \eta_2, \ldots, \eta_h\}$ a set of the directions of scientific research, where h is the quantity of the considered directions of scientific research. Classification of scientists by the directions of researches is some display Λ from a set of A in a set of H, $\Lambda:A\to H$. Each direction of scientific research η_d there corresponds the class of scientists K_d,, who conduct researches in the specified scientific direction.

Assume that for each scientific a_i it is already set estimates of results of their scientific activities and the procedure of a classifying for partition of scientists in the directions is carried out. Assume that $K_1,K_2,\ldots,K_h$ are the classes consisting of results of research and development activities of scientists related a certain scientific direction, h is the quantity of classes. It is necessary to mark that the scientist can research at the same time in several scientific directions, so it can consist in several classes at the same time. For leaving of it is offered to estimate results of research and development activities for each direction separately as results of the independent scientist. As q_i measure values for each scientist dynamically change over time, we will consider that if fix these values with a certain periodicity (quarter, half-year, year, etc.), we can receive values which will reflect dynamics of change of results of research activities of each scientist in time points $t_1,t_2,\ldots,t_s$:

$$\left(q_{i1}^d,q_{i2}^d,\ldots,q_{is}^d\right)=\left(q_i^d(t_1),q_i^d(t_2),\ldots,q_i^d(t_s)\right),\quad (5)$$

where q_{ir}^d is an assessment of the research activities results of scientists a_i, which belongs to the class K_d, $d=\overline{1,h}$, and is recorded at discrete time points t_r, $r=\overline{1,s}$ $t_r=t_1+(r-1)\cdot\Delta t$, $r=\overline{2,s}$, $t_1<t_2<\ldots<t_s$, t_1 is the initial time point, Δt is a time interval of fixing research activity results.

Then (3) can be written as:

$$q_{ip}^{d} = Q\left(P\left(a_i, t_p\right), C\left(a_i, t_p\right)\right),\ a_i \in K_d,\ d = \overline{1,h} \qquad (6)$$

where $P\left(a_i, t_p\right)$ is the set of scientists publications a_i, which were published before the time point t_p, that are all publications of the scientist which have been published in time point $t_1,t_2,\ldots,t_p$, $p \le s$, so in time points t_c, $t_c \le t_p$ where $c = \overline{1, p-1}$; $C\left(a_i, t_p\right)$ is the set of publications, in which the publications are cited scientist a_i, in time points $t_c \le t_p$ where $c = \overline{1, p-1}$; q_{ip}^{d} is the assessment of the research activities results of scientists a_i, which belongs to the class K_d for a period of time preceding the time point t_p.

We will calculate a total score of the research activities results for all scientists who belongs to the appropriate scientific direction or a class K_d for the period preceding time point t_r, $r = \overline{1,s}$:

$$\overline{q}_r^{d} = \sum_{i-1}^{card(K_d)} q_{ir}^{d}, \qquad (7)$$

where $\overline{q}_r^{d}$ is the total score of the research activities results for all scientists who belongs to the class K_d for the period preceding time point t_r; *card(K_d)* is the number of scientists who belongs to the scientific direction or a class K_d.

We will consider other approach of assessment of the results of scientific activity for the direction of scientific research. According to this approach of the publication of all scientists conducting scientific research in the specified direction are considered as the general result of this direction. The corresponding estimates of results of research activity, in particular indexes of citing, are calculated for the scientific direction in general, but not for each scientist separately.

We will designate through P_r^{d} a set of all publications of the scientists, who belong to the class K_d. Moreover, the publications have been published in previous of the t_r timepoints. We will consider further

$$card\left(P_r^1\right) \ge card\left(P_r^2\right) \ge \ldots \ge card\left(P_r^h\right),\ r = \overline{1,s},$$

so the classes of scientists are ordered as nondecrease of number of publications in each of timepoints of $t_1,t_2,\ldots,t_s$. Then the main indexes of citing can be calculated by formulas:

$$h_r^{d} = \max_{y=\overline{1,m}} \min\left\{y, card\left(P_r^{d}\right)\right\}, \qquad (8)$$

where h_r^{d} is the Hirsh's index of the class of scientists K_d, $d = \overline{1,h}$ at the timepoint t_r, $r = \overline{1,s}$,

$$g_r^{d} = \max_{y=\overline{1,m}} \min\left\{ y, \left\lfloor \sqrt{\sum_{x=1}^{y} card\left(P_r^{d}\right)} \right\rfloor \right\}, \qquad (9)$$

where g_r^{d} is the g-index of the class of scientists K_d, $d = \overline{1,h}$ at the timepoint t_r, $r = \overline{1,s}$

$$e_r^{d} = \sqrt{\sum_{y=1}^{h_r^{d}} card\left(P_r^{d}\right) - \left(h_r^{d}\right)^2}, \qquad (10)$$

where e_r^{d} is the e--index of the class of scientists K_d, $d = \overline{1,h}$ at the timepoint t_r, $r = \overline{1,s}$, and h_r^{d} is the Hirsh's index,

$$i10_r^{d} = card\left(C10_r^{d}\right), \qquad (11)$$

where $i10_r^{d}$ is the i10-index of the class of scientists K_d, $d = \overline{1,h}$ at the timepoint t_r, $r = \overline{1,s}$, and $C10_r^{d}$ is the a subset of publications P_r^{d}, quoted in other publications 10 or more times, so:

$$C10_r^{d} = \left\{p_y \in P_r^{d} \,\middle|\, card\left(C\left(p_y\right)\right) \ge 10\right\}, \qquad (12)$$

where *card($C(p_y)$)* is the number of pubications in which the publication p_y is quoted.

For finding the PR-q indexes of classes of scientists it is necessary to solve the system of the linear algebraic equations:

$$\tilde{q}_r^{d} = \sum_{z=1}^{h} \beta_{dz} \xi_z q_r^{z}, \qquad (13)$$

where $\tilde{q}_r^{d}$ is the PR-q index of the class of scientists K_d, $d = \overline{1,h}$ at the timepoint t_r, $r = \overline{1,s}$, $\tilde{q}_r^{z}$ is the PR-q index of the class of scientists K_z at the timepoint t_r, β_{dz} is the the coefficient determined by the number of citings publications of scientists of the class K_d in publications of scientists of the class K_z. ξ_z is the coefficient which provides the existence of the uncommon solution of system of the linear algebraic equations (13).

On the coefficients β_{dz} and ξ_z the same restrictions are imposed as well as when finding PR-q of indexes of scientists [3].

Also it is possible to apply one of methods of finding of integrated assessment to assessment of results of scientific activity of the direction. In more detail methods of definition of these estimates it is considered in work [3].

Let q_r^{d} is the total score of the research activities results for all scientists who belongs to the class K_d for the period preceding time point t_r. The value q_r^{d} is calculated according to the one of formulas (7-13). If we calculate such estimates for various periods, then we will receive a discrete time series $\left(q_1^{d}, q_2^{d}, \ldots, q_s^{d}\right)$. To analyze the dynamics of the research activities results in a given scientific direction, we find the relation $\frac{q_2^{d}}{q_1^{d}}, \frac{q_3^{d}}{q_2^{d}}, \ldots, \frac{q_s^{d}}{q_{s-1}^{d}}$, so, we will construct values:

$$V_a^{d} = \frac{q_{a+1}^{d}}{q_a^{d}},\ a = 1,2,\ldots,s-1, \qquad (14)$$

where V_a^{d} is the potentials of the scientific direction which are defined by a class K_d.

Then the time series of potentials of the scientific directions or a class of K_d has an appearance $\left(V_1^d, V_2^d, \ldots, V_{s-1}^d\right)$, $d = \overline{1,h}$. The forecast of the potential of a scientific direction with horizon 1 is defined as a function of F from retrospective data:

$$\hat{V}_s^d = F\left(V_{s-u}^d, V_{s-u+1}^d, \ldots, V_{s-1}^d\right), \quad (15)$$

where u is a number of retrospective data, elected to calculate the forecast.

The forecast of the potential of the scientific direction with the horizon 2 is determined by (15) and (16):

$$\hat{V}_{s+1}^d = F\left(V_{s-u+1}^d, V_{s-u+2}^d, \ldots, V_{s-1}^d, \hat{V}_s^d\right), \quad (16)$$

$\hat{V}_{s+1}^d$ is the forecast of the potential of the scientific direction 2 points ahead.

F is a function that defines a forecast of potential. For example, for F you can choose a linearly-weighted moving average (LWMA):

$$\hat{V}_s^d = \left(\sum_{k=1}^{u} k\right)^{-1} \sum_{k=1}^{u} k \cdot V_{s-u+k-1}^d, \; d = \overline{1,h}, \quad (17)$$

where $\hat{V}_s^d$ is the forecast of the potential of the scientific direction for 1 point ahead on the basis of a linearly-weighted moving average.

III. RESULTS OF THE RESEARCHING

Formalization of a method of the scientific directions potential forecasting is the result of a research. The method assumes that splitting into the scientific directions and history of estimates changes of the results of research activity of scientists are set. For application of a method the procedure of a classifying of scientists for the directions has to be carried correctly out, and also enough estimates of the scientific activity results of scientists, calculated for the previous periods (sufficient volume of retrospective data) has to be set. The forecast on the described method allows estimating direction potential. If $\hat{V}_s^d < 1$, the direction which represents the K_d class, has the negative dynamics and potential, if $\hat{V}_s^d > 1$ has the positive potential, if $\hat{V}_s^d = 1$, there is no growth of potential, therefore any changes there are no both the positive, and the negative in the results of activities of the scientific direction.

The offered method of forecasting of potential of the scientific directions can be used for comparison of various directions among themselves. The most perspective directions of scientific research there correspond the maximum values of the forecast of potential and vice versa: the least perspective directions there correspond the minimum values.

Feature of the offered method is the need of presence of sufficient volume of retrospective data (history of results of scientific research). Therefore this method isn't applied to assessment of potentials of the new directions of scientific research which results are insufficiently published during the previous periods of time.

IV. CONCLUSION

The method for calculation of the short-term forecast of potential of the scientific directions is constructed. The method can be used in Infocommunication systems for the analysis of the scientific environment for the purpose of definition of those scientific directions, which are perspective or have sufficient potential from the point of view of investment into these directions, cooperation with scientists, which belong to these directions or make the significant contribution to their development, etc. Potential of the scientific directions in this method completely is defined by publishing activity of authors who belong to them, the number of citings their scientific works, which are reflected in the calculated estimates, or indexes of citing.

REFERENCES

[1] R. Solow. “A contribution to the theory of economic growth,” *The Quarterly Journal of Economics*, vol. 70, no. 1, pp. 65-94, 1956.

[2] S. Singh. “Pattern modeling in time-series forecasting,” *Cybernetics and Systems. An International Journal*, vol. 31, no. 1, pp. 49-65, 2000.

[3] A. Biloshchytskyi, A. Kuchansky, Yu. Andrashko, S. Biloshchytska, O. Kuzka and O. Terentyev. “Evaluation methods of the results of scientific research activity of scientists based on the analysis of publication citations,” *Eastern-European Journal of Enterprise Technologies*, vol. 3, no. 2 (84), pp. 4-10, 2017. DOI: 10.15587/1729-4061.2017.103651

[4] A. Biloshchytskyi, A. Kuchansky, S. Biloshchytska and A. Dubnytska. “Conceptual Model of Automatic System of Near Duplicates Detection in Electronic Documents,” in *Iformation Technologies and Computer Modelling: Inter. Conf. CADSM*-2017*:* Lviv Politechnic Nat. Univ., Polyana, 2017, pp. 381-384. DOI**:** 10.1109/CADSM.2017.7916155

[5] P. Lizunov, A. Biloshchytskyi, A. Kuchansky, S. Biloshchytska and L. Chala, “Detection of near duplicates in tables based on the locality-sensitive hashing method and the nearest neighbor method,” *Eastern-European Journal of Enterprise Technologies*, vol. 6, no. 4 (84), pp. 4-10, 2016. DOI: 10.15587/1729-4061.2016.86243

[6] O. Mulesa and F. Geche. “Designing fuzzy expert methods of numeric evaluation of an object for the problems of forecasting,” *Eastern-European Journal of Enterprise Technologies*, vol. 3, no. 4 (81), pp. 37-43, 2016. DOI: 10.15587/1729-4061.2016.70515

[7] L. Egghe. “The Hirsch index and related impact measures,” *TOC*, vol. 44, no. 1, pp. 65-114, 2010. DOI: 10.1002/aris.2010.1440440109

[8] M. Gagolewski and R. Mesiar. “Monotone measures and universal integrals in a uniform framework for the scientific impact assessment problem,” *Information Sciences*, vol. 263, pp. 166-174, 2014. DOI:10.1016/j.ins.2013.12.004

[9] T. Otradskaya and V. Gogunskii. “Development process models for evaluation of performance of the educational establishments,” *Eastern-European Journal of Enterprise Technologies*, vol. 3, no. 3 (81), pp. 12-22, 2016. DOI: 10.15587/1729-4061.2016.66562

[10] V. Gogunskii, O. Kolesnikov, K. Kolesnikova and D. Lukianov. “"Lifelong learning" is a new paradigm of personnel training in enterprises,” *Eastern-European Journal of Enterprise Technologies*, vol. 4, no. 2 (82), pp. 4-10, 2016. DOI: 10.15587/1729-4061.2016.74905