International Journal of Pure and Applied Mathematics
Volume 116 No. 3 2017, 777-788
ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
doi: 10.12732/ijpam.v116i3.23
ijpam.eu

MATHEMATICAL APPARATUS IN THE METHODOLOGY OF RECOGNITION OF IMAGES ON INFORMATIONAL AND SOCIOLOGICAL POLYGONS

Oleg M. Barbakov[1 §], Marina L. Belonozhko[2]

[1]Department of Business Informatics and Mathematics
Tyumen Industrial University
625000, 38 Volodarsky Str., Tyumen, RUSSIAN FEDERATION

[2]Department of Marketing and Government Administration
Tyumen Industrial University
625000, 38 Volodarsky Str., Tyumen, RUSSIAN FEDERATION

Abstract: The article presents the methodology of pattern recognition in the formulation and solution of management tasks at informational and sociological ranges. An original mathematical apparatus is proposed, which is based on connectivity components, similarity measures and a rational complex of indirect features characterizing the direct target property of images and sub-images designated at the polygon. The authors substantiated the necessity of using non-standard methods of mathematics, in particular, the technology of pattern recognition when conducting experiments in virtual space to improve the efficiency of social objects management.

AMS Subject Classification: 03F03, 97E30
Key Words: pattern recognition technology, informational and sociological polygon, measure of similarity, connectivity component, direct target property

1. Introduction

Social phenomena and processes have their own specifics. They are multifactorial, historical, some elements interact with each other and the external

Received: July 24, 2017
Revised: September 4, 2017
Published: October 27, 2017

§Correspondence author

environment, are interconnected by complex connections, susceptible to the influence of subjective factors, difficult for qualification, and with this essentially differ from models that are used in natural science and engineering studies [1].

In the overwhelming majority of cases, when studying social objects, there are no clear ideas about their substantive and structural part, their goals, tasks, the laws of their functioning are not clear, and therefore it is virtually impossible to propose a theory that can explain and implement the mechanism for managing such phenomena and processes. In such cases it is possible and easier to propose and justify the methodology of setting and making managerial decisions, using non-standard methods of applied mathematics (new technology of pattern recognition) [2, 3].

2. Materials and Methods

Such a pattern recognition methodology was used in geological exploration when it was impossible to create a coherent theory in the absence of clear ideas about the content of objects, tasks, models.

Geologists – mathematicians and theorists of the Novosibirsk school proposed to use the methodology of pattern recognition when solving geological exploration problems, including the following elements:

– geological-prospecting experiment;

– interpretation models on the computer;

– calculation of economic efficiency of the proposed solution [4].

Investigating regional management and forming a predictive model of its effectiveness, you can apply the technology of pattern recognition, correcting the stages of conducting experiments in the virtual space. It is supposed to outline the main criteria for the degree of regional management effectiveness – levels, these levels are described by a certain set of properties characterizing them. Then the content part of the informational and sociological range will be determined by the sets of these criteria (properties) [5].

The statement of the management task at the informational and sociological range is carried out as follows: the goal is formulated. For example, in our study it is necessary for people with a fairly low standard of living to determine those criteria and indicators that need to be changed by them in order to improve their level and quality of life. Within the framework of the training that takes place at the information and sociological testing range, objects are distinguished – groups of the population with different living standards, they are described, and all together form the predictive model. Thus, on a computer within a specific programming environment, based on the use of certain algorithms, a computational base is created.

3. Results and Discussion

When the administrative task is set up on the informational and sociological range, the so-called direct property is defined, which is the objects of population groups with different living standards, each linking the living standard (direct or target property) with a set of indirect properties, which determines this level to some extent.

Indirect properties – these are the signs that quantitatively or qualitatively characterize the living standards of the population groups. Sociological surveys are carried out for their measurement, including questionnaires, analysis of secondary documents, and arrays of these signs are also included in the content part of the informational and sociological polygon. The most characteristic features that affect the direct property are: the standard of living, the material component of the vital activity of the region's population, the comfort of living and work, the costs of social security.

The informational and sociological polygon represents images of the initial approximation according to the direct property: the living standard of the population of the Tyumen region in the space of indirect properties – characteristics.

You can design the dependence of the direct property on both a single feature and their combinations on the ECM. At the same time, different combinations affect the direct property in different ways.

The subspace of indicators is divided into a number of areas, each of which has a certain number of points.

Naturally, in the subspace of two indicators it is very difficult to select objects by the direct property, it would be easier to do this in the space of three, four properties and so on. Although this is also difficult, because the parts formed from the set of points intersect in the space of all indicators.

Images of the initial approximation are such areas in the space of indirect features, each image corresponds to the value of the direct property. Images intersect in different values of indirect properties, within the images can be their sub-images: grouping objects – areas within larger areas.

Images of the initial approximation form an informational and sociological ground, on which occurs the formulation and solution of the management task.

The mathematical formulation of the problem is as follows. Informational and sociological polygon g has certain objects, that is, images of the initial approximation A. Long-term observations, examinations, surveys, questionnaires, analysis of statistical literature, periodicals provide us with intervals of direct property $\phi : A_1(\phi_1), A_2(\phi_2), \ldots, A_n(\phi_n)$, where N is the number of intervals of direct property. This is the number of images of the initial approximation.

The real space F^n will include a set of indirect properties, certain collections of which are realized in different real subspaces $F^n = (f_1, f_2, \ldots, f_n)$ where n – is the number of indirect properties or subspaces.

Then the so-called formal point, which is characterized by a certain combination of values of indirect properties: $a_i = (\phi_i, f_{1i}, f_{2i}, \ldots, f_{ni})$ is allocated in the real space F^n as a specific element of a specific object.

A metric of the real space F^n to compare different points will be a measure of the similarity between them. Such a metric is varied in different recognition algorithms. In the current study, a measure of the similarity of Yu.A. Voronin [6], the so-called pair measure of the similarity of the points a_i and a_j by the property f_k, $k = 1, n$:

$$\lambda_{fk}(a_i, a_j) = 1 - \left\{ \frac{f_{ki} - fj}{f_{\max} - f_{\min}} \right\}^2 \tag{1}$$

where f_{max} and f_{min} – are the maximum and minimum limits of values of indirect characteristics in real space.

The paired measure of similarity in a set of indirect properties is defined as follows:

$$\lambda_i(a_i, a_2) = \sum_{k=1}^{n} a_k \lambda_{fk}(a_i, a_j) \tag{2}$$

α_k – is the weight or informative value of indirect characteristics,

$$0 \leq a_k \leq 1, \sum_{k=1}^{n} a_k = 1. \tag{3}$$

The first stage of the new pattern recognition technology is the formulation of a management task at an informational and sociological polygon.

$A_j = \{a_k\}$, $k = 1, N$ is a mathematical description of the informational and sociological polygon g in real space, which includes images of the initial approximation.

A measure of the similarity between two points by an indirect property (formula 1), by a set of indirect properties (formula 2) will be a metric.

The formulation of the management task includes the following stages: the calculation of the recognition constant, the formation of images using connectivity components, the designation by the informativeness of an individual property with its quantitative evaluation of a rational complex of indirect properties, and the evaluation of the priority coefficients of indirect properties' combinations.

The threshold measure of similarity will be the recognition constant, and the similarity measures between each two points will be compared with it, and if this measure is above the threshold, these two points in the real space will be considered points of the same image.

A CC or connected component of points in a real space (it is also called elementary in the F space) will be a measure of the pair of points' similarity greater or equal to the threshold for a set of indirect properties:

$$\lambda_F(a_i, a_j) \geq \lambda_0 \tag{4}$$

Sometimes the measure of similarity in a set of indirect properties between two points is less than a threshold measure of similarity, but if there are a number of intermediate points that connect the two source points according to similarity measures, they can be combined into one connected component, that is, the following conditions are fulfilled:

$$\lambda_F(a_i, a_{i+1}) \geq \lambda_0, \lambda_F(a_{i+1}, a_{i+2}) \geq \lambda_0, ..., \lambda_F(a_{j-1}, a_j) \geq \lambda_0 \tag{5}$$

To designate the connected component, a threshold measure of similarity is calculated in the interval from 0 to 1, and the optimum is selected. If its value 0 one connected component takes up all real space, if 1 – every component of the space of indirect properties is a connected component. In the scientific literature there are several ways to calculate the threshold measure of similarity [7].

For example, its evaluation can be carried out with equal weights of indirect properties:

$$a_i = \frac{1}{n}, i = 1, n \tag{6}$$

The first method is based on the calculation of the lower threshold measure of similarity λ_0^n, which can be used for rough inaccurate primary approximation within the framework of setting the management task at the informational and sociological range.

The first calculation of the lower threshold measure of similarity includes the following procedures. A certain point is compared with all other points of the real space by calculating paired measures of this point's similarity with others on a set of indirect properties. The closest point to a given point in space is at the maximum measure of similarity and is excluded from consideration. Next, the nearest point is taken and the same calculations are made. All the points of the real space are viewed in this way, and then, changing the starting point in space, we find a set of maximal similarity measures. For each iteration (its own starting point) there is the arithmetic average of the maximum paired similarity measures, the arithmetic mean of all the averages of the maximum similarity measures for each iteration is calculated, and it is taken as the lower threshold measure of similarity [8]:

$$\lambda_0{}^H = \frac{1}{N}\sum_{k=1}^{N}\frac{1}{N-1}\sum_{l=1}^{N-1}\lambda_F(b_1, b_K), \tag{7}$$

where N is the polygon's thickness; b_k – is the nearest point, or $\lambda_F(b_1, b_k) = max\lambda_F(b_1, b_j)$, the index $l = 1$ is assigned to the starting point b_k, $l = 2, N$ when j not equal to l – points of the shortest path.

The threshold measure of similarity can also be determined by calculating the upper threshold measure λ_0^b, which is used when it is necessary to more accurately divide the real space into connected components.

The first calculation of the upper threshold measure of similarity is carried out as follows. Each point of real space is taken and for it there is a maximum measure of similarity, and their arithmetic average is considered to be λ_0^b (formula 8):

$$\lambda_0{}^B = \frac{1}{N}\sum_{i=1}^{N}\max_j \lambda_F(b_i, b_j). \tag{8}$$

There is also a way to calculate the threshold measure of similarity, within which the optimal value of λ_0^{opt} is determined. It is calculated in the interval $[\lambda_0^n, \lambda_0^b]$, and on this interval changes of λ_0 must be a constant number of generated connected components. A function $\Psi(\lambda_0)$ of the number of generated CCs from λ_0 is constructed, there is a subinterval that generates the longest step of the function (Figure 1 shows the interval of similarity measures [0.25, 0.31]), its center (0.28) is located, which is the optimal threshold measure of similarity.

When there is a threshold measure of similarity, it is possible to select the preliminary connected components in each image of the initial approximation, for which the pair measures of similarity of each point of the real space with all other points of the image are calculated, but the influences of the indirect properties must be equal: $\alpha_1 = \ldots = \alpha_n = 1/n$. Only after this, any two points b for which the following condition is realized: $\lambda_F(b_i, b_j) \geq \lambda_0$, form the set B_1.

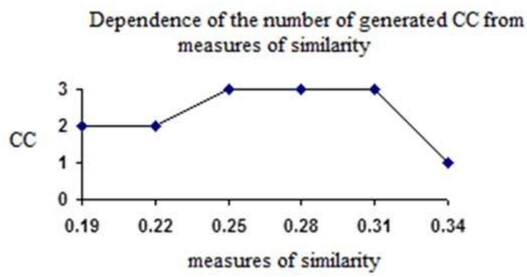

Figure 1: Measures of similarity in the connected component (CC)

There may be points not included in the created connected components, for which similarity measures are also calculated later, and new connected components are found, but for the new B_2 and so on. This process lasts until the entire set of points of the real space is exhausted, and points that are not included in any connected component are not taken into account when solving the problem.

Within the sequence of the stages fulfillment of setting the management task at the informational and sociological range, will be performed a rational complex of indirect properties selection. That is, the selection of the most informative attributes for the direct property. The mathematical formulation of this stage is as follows. We denote the rational complex of indirect properties by F^m, where m is the number of properties that are used in the calculation of the connected components.

Let us define some categories needed in the study for the mathematical formulation of the management task. The first concept is the range or amplitude of the indirect property or the maximum value of the indirect property minus the minimum value or on the image or on the sub-image or connected component, which is an image, or sub-image in certain cases. It is also necessary to introduce the category of the center of gravity's value of the connected component, which can be calculated in the subspace of the indirect property as the maximum value of such a property, minus the minimum value, divided by two.

The individual information coefficient for each indirect property f_i, $i = 1$, n in each sub-image (connected component) can be calculated by the formula 9:

$$a_{ij} = (1 - \frac{\sigma f_{ij}}{\Delta f_i})\frac{\Delta r_{ij}}{\Delta f_i}\frac{N_j}{N} \tag{9}$$

where σf_{ij} – is the range (amplitude) of the values of the f_i property on the sub-image $_j$ of a certain image; Δf_i – is the global span of the f_i property on the entire polygon; Δr_{ij} – is the average distance between the centers of gravity of the sub-image of the given image and all sub-images, only of another image; N_j , N – is the number of points in the j sub-image and on the entire polygon, respectively.

In the calculation it is possible to determine which indirect property contributes the greatest to the calculation of the connected components (its informativity is high), and the influences are necessarily normalized and their sum equals 1.

$$a_{ij}{}^{norm} = \frac{a_{ij}}{\sum\limits_{k=1}^{n} a_{ij}}, \tag{10}$$

where n is the number of properties.

The weight of the image is the arithmetic average of the sub-patterns' weights (formula 11):

$$a_{ij}{}^{sr} = \frac{\sum\limits_{j=1}^{K} a_{ij}{}^{norm}}{K}, \tag{11}$$

where K is the number of sub-images in the image.

Normalization of the images' weights is performed using a rational complex of properties after calculating the weights of the indirect properties as the arithmetic average weights of the images.

The rational complex includes only those weights for which the condition is realized:

$$a_0 = \frac{\sum\limits_{i=1}^{n} a_i}{n}. \tag{12}$$

The rational complex of indirect properties and the calculated coefficients of individual informativeness of each indirect property underlies the repetition of a new process iteration of setting the management task at the polygon. The new iteration includes the calculation of the recognition constant (threshold measure of similarity, lower, upper or optimal), the allocation of images based on the definition of connected components, the refinement of sub-images, that is, the formation of an informational and sociological polygon.

The formulation of the management task consists in the formation of the so-called informational and sociological polygon with well-studied objects (images and sub-images) for which a direct target property and a set of indirect signs of these objects' qualitative aspects (that are characteristic for each) are defined. The images and rational complexes of indirect properties are correlated with each other.

The content of the management task formulation is as follows. The real space of the informational and sociological polygon includes a set of points from which images and sub-images with their values of the direct target property and the corresponding sets of indirect properties that are used to solve the management task, are formed.

The stage of solving the problem involves bringing the groups' objects with a low standard of living to groups with a high standard of living by changing the ratio of indirect properties using the computational base of the informational and sociological polygon. Such objects are compared with the points of images and sub-images of the polygon in the space of indirect properties with the help of their rational complex, individual informativities and taking into account the priority of indirect properties' combinations. These objects, using a geometric approach, are compared with the nearest points in the images and sub-images on the polygon in the subspaces of the indirect properties and by the maximum multiple projective measure of similarity with the points of a certain image for the sought object, the direct target property of this image [9] (formula 13) is adopted:

$$\Lambda(a_j, A_0^1) = \sum_{i=1}^{n} \beta_i \sum_{m=1}^{C_n^i} \max \sum_{k=1}^{i(m)} a_k (1 - \Delta_k^2), \tag{13}$$

where A^1 – is the image from the polygon;

a_j – is the object under analysis;

α_k – is the coefficient of individual informativeness of the k indirect property (its weight);

$i(m)$ – is a rational complex of indirect properties;

β_i – the priority coefficient of indirect properties' combinations;

$$\Delta_k = \frac{f_k(a_i) - f_k(a_j)}{f_k^{\max} - f_k^{\min}}, \tag{14}$$

where a_i – is the object image of A_0^1.

$f_k(a_i), f_k(a_j)$ – are the values of the k property of the sought object and of the A_0^1 images' object, respectively.

In this case, the coefficients k and β_i are necessarily normalized.

The combination of indirect properties of two, three properties and so on is taken into account by the coefficient β_i of the priority of combinations of indirect properties' subspaces, while it [7] normalizes all elements of the formula, after which the multiple projective measure of similarity is measured in the interval from 0 to 1:

$$\sum_{i=1}^{m}\sum_{k=1}^{C_m^i}\beta_{ik}=1. \qquad (15)$$

The normalization is carried out taking into account each k subspace.

The calculation of the priority coefficients of indirect properties' combinations in the current study is based on the relationship between two properties of Yu.A. Voronin [8], in which each combination of indirect properties is weighed, that is, various combinations of different indirect properties are taken into account. The measure of the relationship between direct target and indirect properties is calculated using the formula 16:

$$\sigma(f_t,\phi)=\frac{\sum\limits_{ij}\lambda(f_{ti},f_{tj})\lambda(\phi_i,\phi_j)}{\sqrt{\sum\limits_{ij}\lambda^2(f_{ti},f_{ij})\sum\limits_{ij}\lambda^2(\phi_i,\phi_j)}}, \qquad (16)$$

where $t=\overline{1,m}$; $i,j=1,2,\ldots,N$; $i\neq j$; ϕ – is a direct property; f – indirect property; m – rational complex of indirect properties; N – number of points in the set.

$$\lambda(f_{ti},f_{tj})=1-\{\,\frac{f_{ti}-f_{tj}}{f_{\max}-f_{\min}}\,\}, \qquad (17)$$

where $f_{max}-f_{min}$ – the range of values of the property on the set.

By analogy with the above formula $\lambda(\phi_i,\phi_j)$ is defined.

The measure of the relationship between the direct objective and indirect property in the management task is defined as follows. The path of the nearest points with respect to the direct property is determined as the link between the direct and t indirect property, first in the subspace of the direct and given t indirect property, and different j current points are used. The same is done for all indirect properties with subsequent normalization by the sum of all these properties. The priority coefficient of indirect properties' combinations is calculated by the formula:

$$\beta_{ik}=\sum_{t=1}^{i}a_{kt}\sigma_{kt}(f_t,\phi). \qquad (18)$$

Each specific combination will have its own priority coefficient, and then it will be used in calculations by formula 13.

And so the multiple projective measure of the sought object's similarity with the image points on the informational and sociological polygon is calculated, and then the analyzed object is evaluated in accordance with the following decisive rule:

$$\max \Lambda_F(a_j, {A_0}^1) \geq \lambda_o; |\Lambda_F(a_j, A_0^1) - \Lambda_F(a_j, {A_0}^S)| \geq \varepsilon. \tag{19}$$

Or there is a refusal of the object's evaluation.

When the decisive rule is executed, the evaluated object corresponds to the direct target property of the image with which its multiple projective measure of similarity is maximal. If we take an example of a direct target property – the standard of living, then when the ratio of the indirect properties of the sought object changes, we can achieve the maximum projective measure of similarity with the image of the direct target property's group – a high standard of living. At the stage of setting the problem, images of the direct target property will be formed on the informational and sociological polygon, and new images will be formed during the new iterations of the task formulation. An original computational base of various images used at the stage of solving the management task is created, in particular, the evaluation of new objects in the sense of their belonging to the images of the informational and sociological polygon [10].

4. Conclusion

Thus, the formulation and adoption of managerial decisions on the informational and sociological polygon includes the formation of an informational and predictive model, and the result of the implementation of the decision is entered into the database at the stage of decision's execution. The object is changed, information about its altered state is analyzed by the management entity, and in case of inadequate change of the object within the goals and defined tasks, it is possible to correct the predictive model at the stage of control over the decisions' execution. The feedback of the subject and the control object is realized, interaction and mutual influence contribute to achieving the most effective results [11].

Computer technologies make it possible to implement the constructed predictive models for the formulation and adoption of a management solution in the virtual space, which will allow the selection of various alternative solutions for the computer based on the criteria developed, and then a full management cycle, including social planning, design, organization, social experiments, control of the managerial decisions' results.

References

[1] G.V. Osipova, Sociological research: methods, methodology, mathematics and statistics, In: *Dictionary – reference book: Sociology*, Science, Moscow, **4**, (1991), 178.

[2] V. Zhuravel, Automated Information Systems as Means of Improving the Investigation of Murders, *Journal of the National Academy of Legal Sciences of Ukraine*, Vol. 1, No 88, (2017), 161-172.

[3] .V. Melnichuk, Information and reference work on documents: the foundations of the organization, *Materials of the Afanasiev Readings*, Vol. 1, No 10, (2012): 192-197.

[4] Yu.A. Budyansky, I.V. Barausova, Matrix interpretation models in field geophysics, *Proceedings of ZapSibNIGNI*, (1987), 22-28.

[5] O.M. Barbakov, Regional management: realities and perspectives, Lan, St. Petersburg, (1999), 317.

[6] Yu.A. Voronin, and others, Introduction of measures of similarity and connection for solving geological and geophysical problems, *Mathematical problems of geophysics*, Siberian Branch of AN SSSR, Novosibirsk, Vol. 2, (1971), 305-310.

[7] O.M. Barbakov, Ye.M. Shaposhnikova, Algorithms of mathematical support. Recognition complex "Eureka", *Proceedings of ZapSibNIIGNI: Methods of geological interpretation of complex geophysical data*, (1989), 38-51.

[8] Yu.A. Budyansky, A.N. Yermakova, Construction of a mathematical model of the socio-economic ensemble, *Proceedings of ZapSibNIIGNI: Economic problems of development of the West Siberian oil and gas complex*, (1990), 104-114.

[9] Yu.A. Budyansky, Geological interpretation of complex geophysical data, Nedra, Moscow, (1992), 265.

[10] M. Belonozko, O. Barbakov, S. Siteva, Higher Education Quality Management in Virtual Space of a Regional Higher Educational Institution, *Mediterranean Journal of Social Sciences*, Vol. 6, No. 3, (2015)

[11] O.M. Barbakov, V.G. Kiselyov, Methodology of carrying out social experiments in virtual space, *Tyumen State University Herald*, Vol. 8, (2012).