Algorithm of informative indicators system formation for development of business management tools

N V Shalanov¹, Z A Kapelyuk¹, A A Aletdinova²

¹Siberian University of the Consumer Cooperation, 26, Karla Marksa Ave., Novosibirsk, 630087, Russia
²Novosibirsk State Technical University, 20, Karla Marksa Ave., Novosibirsk, 630073, Russia

E-mail: ec_progn@sibupk.nsk.su, aletdinova@corp.nstu.ru

Abstract. The aim of the work is to improve the management decision – making tools based on informative indicators for organizations. To form a system of informative indicators the authors propose using the following methods. These are the assessment of the uniformity degree of empirical data for each indicator, the analysis and elimination of multicollinearity of indicators based on the matrix of paired correlation coefficients, the integrated assessment of the potential of the object of study using a set of indicators. A distinctive feature of this approach was the construction of a potential function that allows you to design a multidimensional object on a geometric plane. In that case, while forming a system of indicators, some indicators will be excluded taking into account their multicollinearity and weights in the potential function. The advantage of the proposed algorithm is using less expert assessments in scientific research.

1. Introduction
The basis of the modern theory of decision-making is a comprehensive concept of decision-making. It requires evaluation of all essential aspects of the problem situation and the use of logical thinking, human intuition, mathematical methods and technical means. Decision-making is a conscious choice from a set of alternatives. Decision-making tasks correspond to a wide range of practical situations for which there are many solutions and you need to choose the best. Their implementation in modern conditions requires the development of modern tools. Organizations are actively implementing information systems to support decision-making for operational business management. Decision-making tasks correspond to a wide range of practical situations for which there are many solutions and you need to choose the best. Their implementation in modern conditions requires the development of modern tools. Organizations are actively implementing information systems to support decision-making for operational business management. One of the main problems in the creation of such systems is a set of external and internal environment factors, for which it is necessary to choose informative indicators.

Scientists solve this problem in different ways. In their work, P. Delias, M. Doumpos, E. Grigoroudis, P. Manolitzas, N. Matsatsinis describe the development of mathematical models for the analysis of business processes based on clustering to support management decisions [1]. G. Vessia, R. Falconer, B. Zimmermann, A. M. Tarquis developed computational tools to obtain soil management decisions [2]. For the manufacturing industry B. Joung, J. Carrel, P. Sarkar, S. C. Feng made a study of sustainable development indicators, based on mutual similarity, for the development of a database [3].
Thus, one of the most important sections of statistical analysis is data analysis, which involves the study of the initial statistical information presented in the form of a data set describing the object of study. In this case, as a rule, the object of study is described in dynamics by a set of indicators. Consequently, it is represented as a multidimensional dynamic object. Statistical analysis of this data set implies the implementation of activities required to assess the informative indicators that describe the object.

Assessment of the informative indicators of the system requires a number of methods and techniques, which help to solve this problem. There are many approaches and methods of assessment of informative indicators related to both heuristic and analytical methods. However, almost all of them are reduced to the use of expert assessments at a particular stage of solving the problem.

The use of a wide range of different methods of multivariate quantitative analysis to determine if the indicators are informative is discussed in the works of scientists: L. Yu, H. Liu, L. Wang, J. Wu, S. L. Huang, L. Zheng, X. Xu, L. Zhang, J. Huang, M. García-Torres, F. Gómez-Vela, B. Melián-Batista, J. M. Moreno-Vega, G. E. Box, G. M. Jenkins, G. C. Reinsel, G. M. Ljung [4-7].

In accordance with the objectives of the study, they propose the following requirements for the formation of a data set for mathematical research: checking the homogeneity of the selected study groups by methods of expert assessments or multivariate statistics (for example, cluster analysis); normalization of indicators; reducing the dimensionality of the indicator space; standardized description of indicators; construction of classification scales of indicators.

2. Methods

We propose an algorithm for evaluating the informative indicators of the system based on the analysis of the initial data set. It includes the following methods:

• assessment of the uniformity degree of empirical data for each indicator;
• evaluation of multicollinearity of indicators based on the matrix of pair correlation coefficients;
• integrated assessment of the facility potential by a set of indicators;
• formation of a system of informative indicators to describe the object.

To solve the tasks within the framework of the proposed algorithm, it is necessary to formalize the listed methods included in it. To solve the problems in the framework of the proposed algorithm, it is necessary to formalize these techniques. Let the multidimensional dynamic object of the study be described by a set of indicators \( x_1, x_2, \ldots, x_n \). In this case, the dynamics of the object is presented for the period \([t_1, t_N]\). Description of the dynamics of the object of study for this period in the framework of these indicators is presented in the form of a matrix "time-sign", the elements of which are the values of the \( j \)-th sign at the time \( t_i \).

Let us consider an algorithm for evaluating the informative indicators of a system describing a multidimensional dynamic object in the context of each of the indicated techniques.

2.1. Assessment of uniformity of empirical data

One of the tasks of data analysis is the task of assessing the homogeneity of the initial empirical data, reflecting the dynamics of the multidimensional dynamic object, presented in the form of a matrix "time-sign".

As is known, the generalized indicators reflecting the homogeneity of the initial empirical data are the mean square deviation and the coefficient of variation. These characteristics provide estimates of the measure of values dispersion of a indicator relative to its average value. They show that the smaller their values, the more homogeneous the empirical data.

In addition, the assessment of the population homogeneity of empirical data is carried out according to the degree of uniformity. Thus:

\[
\beta_j = \frac{x_j - \bar{x}_j}{x_j} \times 100\%,
\]

where \( \beta_j \) is the degree of uniformity of the \( j \)-th indicator;
\( \bar{x}_j \) – average value of the \( j \)-th indicator;
\[ \sigma_j \] – mean square deviation of the \( j \)-th indicator.

The extent of closeness of the degree of uniformity to 100% shows the degree of homogeneity of empirical data for each indicator.

### 2.2. Assessment of multicollinearity of indicators

System indicators which describe a multidimensional dynamic object often depend on each other and this dependence is not obvious. One of the methods used to evaluate this dependence is the correlation analysis method. According to this method, the correlation matrix (a matrix of the pairs of correlation coefficients between system indicators) is calculated from the initial data set. The high correlation that exists between two indicators is called collinearity, and between several indicators – multicollinearity.

To avoid multicollinearity there is a rule to omit from the analysis one or several related indicators when the absolute value of correlation between them and the remaining indicator is more than 0.7. However, this is done only when the researcher believes that it will not affect the core of the problem under consideration.

### 2.3. Integrated assessment of the facility potential

Homogeneous objects with different levels of development have different potential in the process of changing its internal structure and the environment. In our opinion, the potential of a multidimensional dynamic object is a quantitative measure of its level of development by a set of indicators describing it.

The potential of the object is the level of its development. To determine its level, you should build an axis with the beginning and direction. The value of the object state projection at the time of observation will be a quantitative measure of the object development level.

As a potential function, we propose the following:

\[
y_i = \sum_{j=1}^{n} \alpha_j \frac{x_{ij}}{\sigma_j}.
\]

If the values of the indicators are equal to zero values, the development level will be equal to zero too.

There is a problem of determining the highest level of development. Let the object tends to this state, and it is called the reference, denote it as \( X^* = (x_1^*, x_2^*, \ldots, x_n^*) \). This reference state of the object will correspond to the level of development equal to 100, that is:

\[
y^* = \sum_{j=1}^{n} \alpha_j \frac{x_j^*}{\sigma_j} = 100.
\]

For the proposed axis, which marks the level of development of the object, we propose a scale. Let the value on this scale be calculated as follows:

\[
y^* - 100
\]

Then, from this proportion:

\[
C(t_i) = \frac{y_i - y^*}{100}.
\]

This formula allows you to find the value of the potential on the constructed scale.

Let the object of study is described by a system of indicators \( x_1, x_2, \ldots, x_n \).

In dynamics for the period \( t \) it is represented by the matrix "time-sign", the elements of which are the values of the \( j \)-th sign at the time \( t_i \). To calculate the potential of a multidimensional dynamic object at the time of observation, it is necessary to perform the following calculations [8]:

1. Calculate mean values of indicators:

\[
\bar{x}_j = \frac{\sum_{i=1}^{n} x_{ij}}{N}.
\]

2. Calculate the mean square deviations of indicators:
\[ \sigma_j = \sqrt{\frac{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2}{n}}. \]

3. Calculate the standardized values of indicators:
\[ z_{ij} = \frac{x_{ij}}{\sigma_j}. \]

4. Define expertise-based reference indicators:
\[ X^* = (x_1^*, x_2^*, \ldots, x_n^*). \]

5. Calculate the standardized values of references:
\[ z_j^* = \frac{x_j^*}{\sigma_j}. \]

6. Calculate the weights of the indicators in the potential function:
\[ \alpha_j = \frac{z_j^*}{\sqrt{\sum_{i=1}^{n}(z_j^*)^2}}. \]

7. Construct the potential function:
\[ y = \sum_{j=1}^{n} \alpha_j z_j. \]

8. Calculate the value of the potential function for the moment of observation \( t_i \):
\[ y_i = \sum_{j=1}^{n} \alpha_j z_{ij}. \]

9. Calculate the reference value of the potential function:
\[ y^* = \sum_{j=1}^{n} \alpha_j z_j^*. \]

10. Calculate the potentials for each object:
\[ \mathcal{C}(t_i) = \frac{y_i}{y^*} \cdot 100. \]

2.4. Formation of a system of informative indicators
The formation of a system of informative indicators according to the initial system of indicators is carried out depending on the purpose of the study. Typically, in any study the assessment of the potential for cumulative indicators is of interest. This explains the selection of informative indicators, that is, the exclusion from the system of those indicators that have the least weight in the potential function. Therefore, variables with large weights but multicollinear can also be excluded from the scorecard.

As a procedure for selecting informative indicators, we propose the following approach. First of all, the system of indicators should exclude signs among multicollinear and leave only one of them. However, we should not forget the weight of the signs in the potential function. Then, if the signs are collinear, then the one that has less weight in the potential function is excluded from these two signs, and the sign that has more weight remains. If the signs are multicollinear, then the one that has the most weight in the potential function is left.

Thus, excluding the indicators among collinear and multicollinear with regard to their weights in the potential function, we obtain a system of informative indicators, which is used for further research of a multidimensional dynamic object.

3. Conclusion
It should be noted that the formation methodology of an informative indicators system is extremely important in the theories of cluster analysis and taxonomy. The main tasks of these theories are the formation of multidimensional clusters, taxa, groups of objects from the whole population on a set of classification indicators. In this case, the classification indicators should be independent or orthogonal and at the same time informative. The proposed methodology solves this problem among others. The
advantage of this approach is that it is simpler and allows to reduce the use of expert assessments. Since most of the methods of evaluation of informative indicators are based on the heuristic method, which is based on expert evaluation of the significance of indicators. It should only be noted that to solve the problems of taxonomy, a data set of the "object-sign" matrix type is used, the elements of which are the values of the \( j \)-th sign inherent in the \( i \)-th object.

At the same time, the most difficult and main task of cluster analysis is the objects classes formation on a set of indicators. The authors propose to solve this problem on the basis of one classification and the most informative feature – the level of potential. And, thus, the problem of multidimensional classification is reduced to one-dimensional. While significantly simplifying the task, we do not lose information about the objects of research and achieve a high level of correctness in solving the problem of taxonomy.

Thus, the authors propose using methodical support for the development of information systems for business decision-making. The consistent application of the above four methods in the analysis of data is an algorithm for the formation of an informative system of indicators for the initial information set. Our algorithm will reduce the analyzed indicators amount of external and internal business environment, the cost of their collection and processing to get the optimum solution.

References
[1] Delias P, Doumpos M, Grigoroudis E, Manolitzas P, Matsatsinis N 2015 Knowledge-Based Systems 84 203-213
[2] Vessia G, Falconer R, Zimmermann B, Tarquis A M 2018 Biosystems Engineering 168 1–3
[3] Joung C B, Carrell J, Sarkar P, Feng S C 2013 Ecological indicators 24 148-157
[4] Yu L, Liu H 2004 Journal of machine learning research 5 (Oct) 1205-1224
[5] Wang L, Wu J, Huang S L, Zheng L, Xu X, Zhang L, Huang J. 2018 arXiv preprint arXiv 1811.08979
[6] García-Torres M, Gómez-Vela F, Melián-Batista B, Moreno-Vega J M 2016 Information Sciences 326 102-118
[7] Box G E, Jenkins G M, Reinsel G C, Ljung G M 2015 Time series analysis: forecasting and control (New York: John Wiley & Sons)
[8] Shalanov N V, Aletdinova A A 2018 Journal of Physics: Conference Series 1015 032004