UDC 658.62.018

A.I. Mikhalyov, A.A.Stenin, I.G. Drozdovich, S.A.Stenin, Yu. A. Tymoshin

CLASSIFICATION AND ASSESSMENT OF THE QUALITY OF INDUSTRIAL PRODUCTS

Annotation. To solve the problem of classifying industrial products on the basis of a statistical method, a procedure is proposed for the formation of a ranked scale of characteristics. Based on this, an assessment of the quality of a specific industrial product within the class under consideration is proposed using the empirical criterion proposed in the article, which allows us to avoid the disadvantages of known methods (in particular, integral, differential, etc.) by using the combinatorial approach.

Keywords: industrial product, ranking of features, empirical criterion, combinatorial approach.

Introduction

The solution of the problem of improving the quality of products, which is very relevant for modern industry, is connected with the consideration of a multitude of tasks, among which an important role is played by the problem of deciding on the quality of a product [1-3]. Objectivity of the decision depends on sources of information about the quality of the product. Information about the state of the product is borne by various signs (factors), and given that the concept of quality is complex and multifaceted, the number of indicators describing all aspects of product quality can reach several tens or even several hundred. Obviously, not all of them are equivalent, not all contain the necessary amount of information about the quality of the product. Naturally, the problem arises: how many signs and what characteristics need to be controlled in order to make a decision about the quality of the product that satisfies the given criterion, that is, it is necessary to select the most important indicators for evaluation of quality or signs, which are called informative. In this case, the significance or informative character of the characteristics should be assessed quantitatively, since only in this case it is possible to give preference to one or another characteristic, which is difficult to do with a qualitative approach to the selection of characteristics. Formation of informative signs, as a rule, is carried out in two stages. At the first stage, physical analysis is used, on the basis of which the entire set of measured characteristics is selected (with excess according to a given criterion), which to some extent characterizes the quality of the products. At the second stage, a mathematical approach is applied, when the most informative (significant) are selected from the entire excess set.

The quality level is a relative characteristic based on a comparison of the values of the most informative features of the evaluated industrial products with the corresponding values taken as a basis for comparison.

Review of existing solutions

At present, in order to solve the problem of forming a ranked scale of characteristics, in most cases the procedure for analyzing a pre-defined ensemble of characteristics is formalized. Sometimes, however, the procedure for building up this ensemble is formalized, if necessary. However, in this case, as a rule, only the obtaining of pre-prescribed properties is realized. There are two approaches to the solution of the above problem [4].

The first approach is that from the very beginning, the setting for finding a small number of signs of great informative is taken. However, all the methods, used in this way, are still based on heuristics and empirics, i.e. the choice of attributes is determined by the intuition, experience and imagination of the developer. However successful the constructed system of attributes can be, it can not be proved that it is better than some other [5,6].

The second approach is that of a large number of initial characteristics, according to some criterion of informative character of the characteristics, the smallest number of the most useful traits is selected [7,9].

Both approaches to the construction of an informative system of characteristics require the definition of criteria for assessing the informative of the characteristics.

To date, various criteria for the informative of attributes have been developed, based on the methods of mathematical statistics and information theory. The most recognized of them were criteria reflecting the distances between class distributions [8,9].

From the above approaches to constructing an informative system of characteristics, the second approach is more constructive than the first . On its basis, it is possible to link functionally the criterion of informative of the characteristics with the probability of recognition error [8,9].

Evaluation of the quality of industrial products involves the compliance of values of informative features of the evaluated industrial products to the requirements of consumers and the choice, if necessary, of the direction of improving their quality.

A wide application in the recent past was the practice of combining (addition, multiplication, etc.) individual quality characteristics (complex, group, generalized, integral) [3]. The difference in this case in the dimension of features was overcome by using dimensionless relative (relative to objects taken for the base) of their values. Relative significance of the signs was taken into account by the importance factors. However, the signs carried their physical essence with them in the formulas and led to absurdities, when, for example, with the same complex quality indicators, one of the compared cars could not move, but had higher performance of other characteristics. At the same time, with the correct formation of groups of characteristics for a particular class of industrial products and the synthesis of a combinatorial quality criterion, an effective system for assessing the quality of industrial products can be built.

Formulation of the problem

For assess the informative (significance) of the characteristics introduce a certain set of numbers, each of which will characterize the "usefulness" of a particular feature. Such numbers are called "weights", because they describe, as it were, the weights of the characteristics in the overall evaluation of the quality of the product. These numbers form a field of positive real numbers, and the largest weight should be assigned to the characteristic with the greatest informative. Suppose there is a set of N objects. And, for simplicity of explanation, this set consists of only two classes of objects K_1 and K_2. Each object of the set K is described by the same set **n** features: $x_1, x_2, \ldots, x_n$, whose values in the aggregate determine the belonging of the object to its class. If the feature space is considered as a linear metric, then the set of attributes generates two sums characterizing the closeness of objects to the classes K_1 and K_2:

$$S_1 = \sum_{s=1}^{n} a_s(\bar{x}_{1s} - x_s), \; S_2 = \sum_{s=1}^{n} a_s(\bar{x}_{2s} - x_s) \quad (1)$$

where $\bar{x}_{1s}$ and $\bar{x}_{2s}$ - mean values of the *s*-th attribute, determined by the collection of instances of the first and second classes, respectively; a_s - the weight coefficient of the *s*-th characteristic.

The task is reduced to finding a set of weight coefficients that would reduce the distance between objects within one class and increase the distance between objects of different classes [10]. Then the belonging of the L-th object to one of the classes can be estimated using the difference y = S_1-S_2, i.e $x_L \in K_1$ if S_1 <W, S_2> W; $x_L \in K_2$, if S1> W, S2 <W, where W is a certain threshold, with respect to which the

belonging of the L-th object to the corresponding class is estimated. The magnitude of threshold depends on the degree of overlapping of the classes K_1 and K_2, that is, on the set $G=\{\Omega_L\}$ of objects. Thus, it is necessary to find the set of weight coefficients that enter into the relation y = s1-s2 for minimize the intersection of classes.

Ranging of informative characteristics

Let the set G be the set of objects belonging to the intersection of the classes K_1 and K_2. Each element of a set can be characterize by the sum

$$S_L = \sum_{s=1}^{n} E_{sL} a_s \ . \tag{2}$$

The average value of such sums for objects $x_L \in G$ is

$$S = \frac{1}{m}\sum_{L=1}^{m} s_L = \frac{1}{m}\sum_{L=1}^{m}\sum_{s=1}^{n} E_{sL} a_s = \sum_{s=1}^{n} \overline{E}_s a_s \ , \tag{3}$$

where $\overline{E}_s$ - the mean value of the s-th characteristic over all objects of the set G.

In order to be able to accurately divide the objects of the set G belonging to different classes, it is necessary to find the direction of the maximum dispersion of objects in the feature space. Obviously, in this direction there must be a vector, the components of which are the weights a_s. Then these weight coefficients will generate such values of s_L that the set G will be represented as two groups, with the exception of objects for which s_L has a value close or equal $\overline{S}$.

Let us form an expression of the form

$$y = \frac{\sum_{L=1}^{m} \left(s_L - \overline{S}\right)^2}{\sum_{s=1}^{n} a_s^2} . \tag{4}$$

The sum $\sum_{L=1}^{m} \left(s_L - \overline{s}\right)^2$ in this expression will be more greater, the smaller the sum of squares, and the best separation of objects. Further, according to [10], the expression for **y** can be written in the matrix form: **y = aAa** '/ **aBa**', where **a** is a row - vector, **a**[1] is a column -vector, **B** is a unit matrix, and **A** is a matrix proportional to the sample covariance matrix whose elements are numbers

$$a_{sj} = \sum_{L=1}^{m} \left(E_{sL} - \overline{E}_s\right)\left(E_{jL} - \overline{E}_s\right) \tag{5}$$

Since the matrix A is given in the region of real numbers and is symmetric, all its eigenvalues are real numbers. Differentiating the expression for **y** for each a_s, we obtain a system of equations:

$|yB - A|a' = 0.$ (6)

For nonzero a_s, the system of equations is solvable only if its determinant $|yB - A| = 0$. For the relation (6), we can define the characteristic polynomial of the matrix A whose roots form the spectrum of the matrix $y_1, y_2, ..., y_n$. Among all the eigenvalues of the matrix A there is a maximum y_s, an eigenvector corresponding to this number is the vector whose components form the desired set of weight coefficients.

The eigenvector can be found by the iteration method [11]. Since matrix A is a real symmetric matrix of order ***n***, the eigenvalues of this matrix are real numbers, and the eigenvectors $a_1, a_2, ..., a_n$ form an orthogonal basis. Then any arbitrary n-dimensional vector **f** can be uniquely decomposed in this basis: $f=a_1a^*_1+a_2a^*_2+ ... +a_na^*_n$, where a^*_s are the eigenvectors of the matrix A.

Multiplying **f** by A_m, we obtain $A_mf = a_1A_ma^*_1 + ...$. If y_1 is the largest eigenvalue of the matrix A, and ***m*** is sufficiently large, then terms that do not contain *y1* can be neglected. Then among the terms denoted by many points, only terms with the coefficients y_{m2}, $my_{(m-l)2}$, ..., y_{m3}, etc. are contained. In this case $A_mf/y_{m1}==a_1a^*_1+...$, where the non-written terms contain y_{m2}/y_{m1} ... etc.

Since $| y1 | > | y2 |, | y1 |> | y3 |$, then for $m \to \infty$

$$\lim_{m \to \infty} \frac{A_m f}{y_{m1}} = a_1 a_1 . \quad (7)$$

Denoting the L-th component of the vector A_mf in an arbitrary fixed basis by V_mL, and the L-th component of the vector **a** in the same basis by a_L and defining the limits

$$\lim_{m \to \infty} \frac{a_1 V_{(m-1)L}}{y_{(m-1)1}} = a_1 a_L, \quad \lim_{m \to \infty} \frac{a_1 V_{mL}}{y_{m1}} = a_1 a_L , \quad (8)$$

can be written

$$\lim_{m \to \infty} \frac{V_{mL}}{V_{(m-1)L}} = y_1 . \quad (9)$$

Thus, the numbers $V_{m1}, V_{m2}, ..., V_{mn}$ for large **m** are proportional to the numbers $a_1, a_2, ..., a_k$, and the vector A_mf is approximately proportional to the eigenvector corresponding to the eigenvalue y1. Finding the necessary set of weighting factors is as follows. We take an arbitrary vector **f** and compose the sequence A_1f; A_2f; A_3f; ..., A_mf, ... Starting from some m, the row of the matrix A_mf is approximately proportional to the row of the matrix $A_{m-1}f$. The coefficient of proportionality is an eigenvalue of y_1, and the vector $A_m f$ itself is an eigenvector whose components are the desired set of weight coefficients.

By arranging the coefficients in descending order | a_1 |> | a_2 |> ...>, we get system of attributes by significance.

Evaluation of the quality of industrial products

Since any industrial product in the process of assessing its quality is characterized, as a rule, by the set of characteristics defined above, the use of linear convolution of the characteristics with the weights found for them determines only belonging to one of several similar classes of objects (for example, the class of cars- passenger, sports, etc.). However, further, as a rule, there arises a need to assess the quality of objects within a given class on a set of similar characteristics. At the same time, for unbiasedness and consistency of the summary for information signs it is necessary to fulfill the following conditions [12]:

1. The limiting relative error of the total assessment should lie within the limits determined by the values of the maximum and minimum errors in the characteristics

2. The total estimate should take into account, in a certain way, the importance of each feature or group of similar characteristics.

Such conditions are satisfied by the empirical criterion of the kind proposed in this article:

$$J_k = \frac{\sum_{l=1}^{N_k} \left(\frac{1}{\theta_l}\right)^{p_l - 1}}{\sum_{l=1}^{N_k} \left(\frac{1}{\theta_l}\right)^{p_l}}, p_l \geq 1, \tag{10}$$

the value of which is determined on the basis of estimates $\theta_l, \ (l = \overline{1.N_k})$ characteristics *k*-th object from the class under consideration. It is the relative estimate of the *i*-th characteristic, which is defined as module of the ratio of the difference between the optimal for this class value of the i-th characteristic. It is the actual value of the *i*-th characteristic for the given object to the optimal value of the i-th characteristic of the class in question, and the number of features N_k and their list are the same for all objects in the class under consideration.

Relative evaluation lies in the range [0,1] and its smaller value corresponds to greater proximity to the optimal value p_l of the characteristic in this class of objects.

By choosing the appropriate value p_l, taking into account the quality of one or another characteristic, and also the degree of importance of characteristic, it is possible to make purposeful the change in the values of the total criterion in the range from 0 to 1. Moreover, if all the terms have the same (approximately the same) limiting relative error, then the sum has the same (or approximately the same) limiting relative error [12]. In other words, in this case the accuracy of the sum (in percentage terms) is not inferior to the accuracy of the terms. However, with a large number of terms, the sum is, as a rule, much more accurate than the individual members of the sum due to a mutual compensation of errors, so the true error of the sum only in exceptional cases coincides with or is close to the marginal error.

Consequently, the main factor determining the change in the values of this sum in the range $[0 \div 1]$ is the choice of the value p_l corresponding to each object, which is achieved as follows. All characteristics of the object of this class are combined in accordance with their significance in a number of groups. Let, for example, for the *k*-th class of objects the following estimates be divided into groups (see Table 1).

Table 1

$\theta_1, \theta_3, \theta_6 \dots$	$\theta_2, \theta_9, \theta_{11} \dots \dots \theta_{N_k-1}, \theta_{N_k}$	$\dots \theta_{N_k-1}, \theta_{N_k}$
1st group	*j*-th group	*R*-th group

Thus, N_k the partial characteristics θ_l, $(l = \overline{1.N_k})$ are divided into R groups by $S_{j,}$ $(j = \overline{1,R})$ the elements in each partition group. It is known [13], that the total number of permutations in this case is equal to:

$$P_0 = N_k\,! = (\sum_{j=1}^{R} S_j)! \tag{11}$$

The number of permutations in which at least one of the particular characteristics falls from one's group to another, which in turn should be fined, is defined as:

$$P_B = (\sum_{j=1}^{R} S_j)! - \prod_{j=1}^{R} (S_j)!. \tag{12}$$

Then the proportion of this number of permutations from the total number is determined in the following way:

$$\Delta p = \frac{P_B}{P_0} = \frac{(\sum_{j=1}^{R} S_j)! - \prod_{j=1}^{R} (S_j)!}{(\sum_{j=1}^{R} S_j)!}, \tag{13}$$

and can serve as the value of the discrete step of the penalty *pl*. Moreover, if for some *l*-th partial characteristic, the estimate by which must be in the j-th group $(j = \overline{1,R})$, falls into the i-th group $(i = \overline{1,R})$, then the corresponding value *pl* is determined according to the expression

$$\begin{cases} pl = 1 + (i - j)\Delta p & i > j; \\ pl = 1, & i \le j. \end{cases} \quad (14)$$

The quantitative values of the complex criterion thus obtained can be interpreted in a more suitable system of estimates. Most often, interval scales are used for this purpose. The interval scale is introduced to within a linear transformation:

$$\{x_i\} \to \{x_i^{'}\} = ax_i + b \quad (15)$$

and involves transforming the estimates obtained from one scale into estimates on another scale using a linear transformation (15). Obviously, the smaller value of criterion (13) corresponds to the best quality of the given object in relation to other objects of the given class.

Conclusion

To solve the task of classifying industrial products on the basis of the statistical method, a procedure is proposed for the formation of a ranked scale of informative indicators of the quality evaluation of these products. The evaluation of the quality of a specific industrial product within the class under consideration is proposed to be carried out using the empirical criterion proposed in the article, which allows using the combinatorial approach to avoid the shortcomings of known methods (in particular, integral, differential, etc.).

REFERENCES

1. Bukharin S.V. Expert assessment of the quality and price of goods. - Voronezh: DONO (IMMiF), 2006 -160p.

2. Medunetsky V.M., Fedosovsky M.E. Quality and certification of industrial products. Teaching manual. - SPb. : NRU ITMO, 2013 – 44p.

3. Feigenbaum A. Product quality control. - M .: Economy, 1996.-471p.

4. Orlov A.I. Organizational economic modeling. Part II. Expert estimates.- M.: Publishing House of Moscow State University. Bauman. 2011.- 486p.

5. Litvak B.G. Expert assessments and decision making - M .: Patent. 1996. -271p.

6. Dobrov G.M. Expert estimates in scientific and technical forecasting. - K.: Naukova Dumka, 1974.-160 p.

7. Beshelev SL, Gurvich F. G. Mathematical methods of expert assessments. - M.: Statistics, 1980. - 263p.

8. Fomin Ya. A. Pattern Recognition: Theory and Applications. - 2nd ed. - M.: FASIS, 2012. - 429p.

9. Fomin Ya. A., Tarlovsky G. R. Statistical theory of pattern recognition. - M.: Radio and communication, 1986. - 624p.

10. Dubrovin V.I. Subbotin S.A. Boguslaev A.V. Yatsenko V.K. Intelligent tools for diagnosing and predicting the reliability of aircraft engines-Zaporozhye: Motor-Sich. 2003.-279p.

11. Samara A.A. Introduction to numerical methods. - SPb.: Lan publishing house, 2005. – 288p.

12. A.P. Mishin, I.V. Proskuryakov. Higher algebra. —M.: State Publishing House of Physics and Mathematics, 1962. -300p.