UDC 338.24+330.42

USE OF CLUSTER ANALYSIS IN THE PROCESS OF CLASSIFICATION OF ECONOMIC OBJECTS

Norik Larisa, candidate of economic sciences, associate professor,
associate professor of department of higher mathematics and economic-mathematical methods,
Simon Kuznets Kharkiv National University of Economics, Ukraine,
Melentieva Irina, teacher of Kharkiv College of Trade and Economics
Kyiv National University of Trade and Economics, Ukraine

***Abstract* – The basic task of classification of economic objects is described. The review of basic types of tasks of classification is given. The example of realization of cluster analysis is resulted in the task of classification of industrial enterprises on the indexes of results of their activity**

***Keywords* – Task of classification, cluster analysis, procedures of cluster analysis, measure of likenesses, function of distance, hierarchical clusterization**

In an economy at treatment of large volumes of information there is a task of choice, organization and use of sorting methods, based on the algorithm of classification. The basic task of classification consists in breaking up of great number of information on categories or classes so that all of elements into every class had an enough body of general signs, which allows ignoring them differences [1].

The economic tasks of classification can part on three types:

breaking up of information (enterprises, regions, countries, indexes of basic descriptions of objects) on the set classes;

determination of descriptions of classes (key factors) from these classes and including in them of research objects;

creation of classes (categories) on the basis of set of objects and some information about them.

For the tasks of the first type descriptions of information and classes, and also criteria of taking of information, must be obviously set to one or another class [2].

In the classic type of task of such type formulated as follows:

let there is information about classes $I(G_1, G_2,..., G_n)$ from set of classes $G_1, G_2,..., G_n$, information about the objects of research $I(P)$ from the set of objects $P = \{P_1, P_2,..., P_m\}$ and $P = \{P_1, P_2,..., P_m\}$ $P = \bigcup_{i=1}^{n} G_i$.

A task consists of that on information about classes $I(G_1, G_2,..., G_n)$ and on description of objects $I(P)$ to build an algorithm, as a result of application of which it is possible it would be to draw a conclusion about belonging of this object to one of classes.

The tasks of the second type in economic researches are examined, when classification in accordance with the set criteria is already executed and classes are created. It is necessary on the basis of the known information about description of classes, their amount and about the attributes of incoming in them objects to calculate descriptions of classes and expose connections between the objects of class [2].

The tasks of classification of the third type suppose that basic data (attributes, descriptions) are described and set criteria of classification. The amount of classes (or categories) to which basic data will be attributed in accordance with these criteria of classification is beforehand unknown thus.

In economic researches a task of breaking up of aggregate of objects on a category (classes) in accordance with the beforehand known (set) criteria is one of basic. Such task can be attributed to the tasks of classification of the first type and name the task of multidimensional classification.

A cluster analysis, factor analysis and multidimensional scaling, behave to the basic

methods of multidimensional classification. Advantage of cluster analysis is in that he allows to execute breaking up of objects not on one parameter, but on the whole set of signs. In addition, a cluster analysis lays on no limitations of the examined objects and allows examining basic data of arbitrary type.

Let $P = \{P_1, P_2, ..., P_m\}$ it is a set of objects from some aggregate.

For a set P there is a set of vectors of measuring $X = \{X_1, X_2, ..., X_m\}$, which describe a set P. Set X it can be presented how m points in n-measured euclidean space.

The task of cluster analysis consists in that on the basis of information, contained in a set X, to break up the set of objects P on k clusters (classes). Every object P thus must belong to the only to one class. Into a class objects are similar between itself, objects which belong to the different classes are heterogeneous.

The decision of task of cluster analysis is breaking up, which satisfies some criterion of optimality.

Among procedures clusterizations select hierarchical, parallel and successive (tabl. 1).

Table 1

Procedures of clusterization

Type of procedure	Contents of procedure
Hierar-chical	Association in a separate class most the close located objects of classification, and then – more remote
Parallel	Optimization of breaking up in obedience to the chosen function of quality will be realized and clusters are created on principle of determining the locations of most closeness of points. The concept of standard points is applyed in this case
Succes-sive	In part of vectors of measuring is the beginning applyed in metrical space, after other, which get out in accordance with the set rule in some vicinity of previous part of information until all of selection will not be outspent. Thus the selection of accumulation of points is made by finding of fashions

The concept of homogeneity of objects of cluster is set by some the distances, which characterizes the degree of closeness (likeness, similarity) of objects. There are different types of measures of likeness: coefficients of similarity; coefficients of connection (correlations); indexes of distance are in metrical space, norm.

The different methods of cluster analysis are today applyed: method of complete linkage, method of maximal local distance, method of weighted pair-group average, Ward's method, centroid method [3].

We will consider the application of cluster analysis on the example of classification 19 industrial enterprises of Ukraine (E1, E2, …, E19). Each of enterprises is an object which is described by 83 indexes of results of their activity (financial indexes, indexes of efficiency of investments, use of capital assets, labour resources, indexes of expenses et al). Comparison of distance between these objects will reflect their likeness with each other. Economic sense of the similar understanding of likeness means that enterprises are considered the more so alike, than less distinction is between indexes which they are described by.

At the calculation of distance it is necessary to set the scale of measuring. Standardization of information is therefore made in the software environment of package of STATISTICA [3].

On the first stage the analysis of possibility of forming of clusters is conducted. By the method of hierarchical classification, rules of method of complete connection and measure of closeness of euclidean distance dendrogram is got (fig. 1).

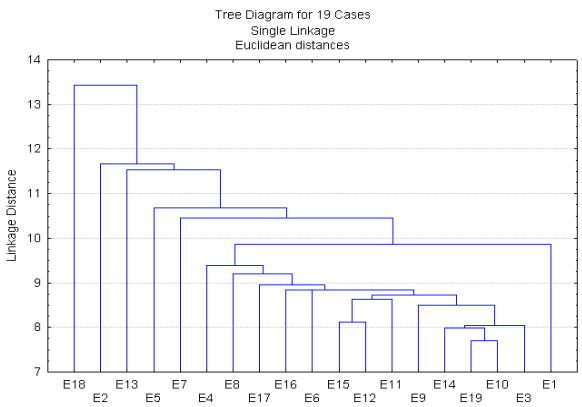

Fig. 1. **Dendrogram results cluster analysis**

With the purpose of determining the optimum amount of clusters it is necessary to build the graph of list of association of enterprises in clusters, setting aside distances on a vertical ax, and on horizontal is a step of association (fig. 2).

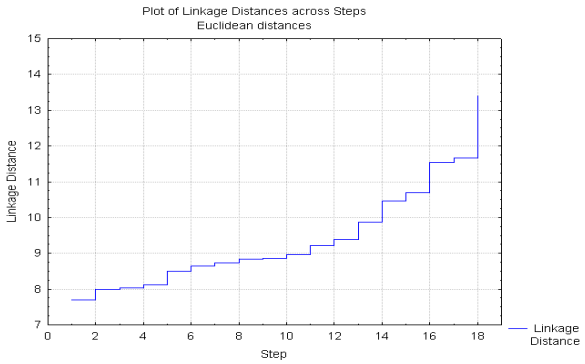

Fig. 2. **Graph of list of association of enterprises in clusters**

The optimum is consider such amount of clusters, which is equal to the difference of number of objects of supervisions and number of steps after which distance of association grows saltatory.

Coming from visual presentation of results, supposition is done, that enterprises are formed by five natural clusters. Verification of this supposition is further executed by the method of K-means and verification of meaningfulness of distinction between the got groups. The value of level of meaningfulness is got $p < 0{,}05$, that means meaningfulness of distinction between the got clusters. The mean values of all of indexes considerably differ from each other (fig. 3). It testifies to the high-quality breaking up on groups.

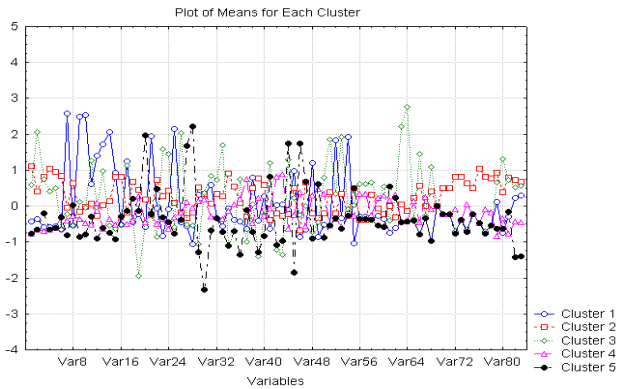

Fig. 3. **Linear graph of clusters**

In the first cluster the enterprises E2 and E6 entered with the high level of business activity of enterprise and efficiency of the use of capital assets of enterprise, but with the low level of efficiency of investments, in the second – E1, E3, E8, E13, E15, E17 with the stable enough results of activity, in the third – E7, E18 with the high level of profitability and efficiency of innovations, in fourth – E5, E9, E10, E11, E12, E14, E19 with the high level of expenses, in fifth – E4, E16 with the low level of mean values of all of groups of performance indicators.

Information 19 enterprises high-quality determine belonging to the classes and suitable for further research. Further for every cluster it is possible to calculate basic descriptive statisticians which can be applyed for the estimation of possible losses of enterprise.

Thus, a large value has a cluster analysis in the leadthrough of analytical researches due to possibility to transform the large volume of scalene information in a well-organized, compact kind. It is instrumental in the increase of level of evidentness and perception of results of analysis, and also provides a basis for prognostication.

References

1. Runkler T.A. Web mining with relational clustering /Runkler T.A., Bezdek J.C. // International Journal of Approximate Reasoning. – February 2003. – Vol. 32, № 2-3. – P. 217-236.

2. Bradley P. S. Constrained k-means clustering [Electronic resource] / Bradley P. S., Bennett K. P. Demiriz A. // Microsoft Research. MSR-TR-2000-65. 2000. – Redmond W. A. – The access mode is to the article: http://www.litech.org/~wkiri/Papers/wkiri.html

3. Khalafyan A. A. Statistica 6.0 Statistical analysis of data: textbook / A. A. Khalafyan. – M. : Binom, 2007. – 512 p.

Authors

Norik Larisa, associate professor, Simon Kuznets Kharkiv National University of Economics (larisa.norik@gmail.com),

Melentieva Irina, teacher, Kharkiv College of Trade and Economics Kyiv National University of Trade and Economics (imelenteva@list.ru).

The theses acted on January, 25, 2016.
It is published in an author release.