Babenko N.I.

Kherson **Physics and Mathematics Lyceum of KNTU and DNU**, Kherson, Ukraine

APPLICATION OF HIERARCHICAL CLUSTERING ALGORITHM FOR STRUCTURAL CHARACTERISTIC OF MOVING PHYSICAL OBJECTS

Approach to the development of management solutions using the cluster analysis can qualitatively improve the management system by moving objects through the adequate response to the impact of the key factors influencing the characteristics of physical objects.

The aim is to attempt to solve the problem of identifying key factors and physical signs of moving physical objects needed to make appropriate management decisions by using cluster analysis.

The article defines the types of clustering algorithms; the system of information parameters directly or indirectly characterizing the analyzed characteristics is emphasized, hierarchical and non-hierarchical cluster analysis methods are considered.

The research finding is the construction of tree diagram using the program STATISTICA 8, which gives the idea of possible clusters' number combining physical indicators under the dynamic changes of moving objects.

The advantage of cluster analysis usage is the use of factors relating to both internal and external environments of the physical properties' interaction of moving objects.

Keywords: cluster analysis, moving objects, STATISTICA 8.

Introduction. Characteristics describing properties of the moving physical objects are measured in different scales and dimensions. Although they have differences they are interrelated and interdependent due to they characterize the condition of the system in general. The solution of the problem of structuring of the properties of the moving physical objects in general comes down to clustering in n-dimensional space.

Under the conditions of moving physical objects characteristics of dynamic changes of their properties become ambiguous and inhomogeneous. It makes oneself evident in fact that arising conditions characterized by changes of their properties have occasionally become totally new. This fact makes a request to improve efficiency of the managerial decisions on a prediction of characteristics of the travelling physical objects and their structuring. Since existing methods of the estimation of the fact impact are not always satisfied the changing conditions it is difficult to find out the characteristics impact and determine key among them. For that matter there has been a methodical problem of acceptance of the managerial decisions based on the key figures and structuring facts in making an assessment of the condition of the travelling objects.

Analysis of the articles concerning this issue shows that efficient functioning of the moving objects demands understanding of different interrelations between factors that influence on value and results of the system's functioning in general. The cluster analysis is considered to be very useful for these purposes. Theoretical methodological aspects of the application of the cluster analysis are widely presented in the scientific publications (1-4).However the issues of application opportunities of cluster analysis methods in studies of phenomena and processes in dynamics haven't still found deep observation.

The aim of the research paper is an attempt to solve problems of determination of key factors and physical features of the travelling objects that are necessary for making adequate decisions via the application of the clustering procedure.

Presentation of the basic material

Cluster analysis belongs to the class of multivariable methods intended for formation relative Relatively remote groups of homogeneous objects based on measured characteristics and allow to classify multivariate observations each of which is described by set of output variables $x_1, x_2, x_3, \ldots x_m$. It can be used either for analysis of the structure in data characterizing phenomena and processes

according to the correlation matrix or for analysis of objects and subjects by levels which are described with equitable attributes. The main advantage of clustering is the data compression ability.

The task of cluster analysis is that on the ground of data contained in X to split up a set of objects G into m clusters Q_1, Q_2, Q_m such as each object G_j will belong to only one subset and that objects in the same group will be similar. The tree structured method is the most widespread to form clusters whereby distances between objects in multidimensional space are determined by squared Euclidian distances. Ward's method [5] can be used for evaluation distances between clusters which applies analysis-of-variance procedures that minimizes sum of squares at each step of clustering. However the method minimizes sum of squares so it can be applied only for small-sized clusters and data sets.

In cases where the number of clusters "K" is known the K-means clustering is used. The purpose of this is to form the "K" clusters that are located father apart from each other. This method is considered to be a variant of analysis of variance pointing on minimization of location of elements inside cluster and maximization distances between clusters. In this approach the objects move from one cluster to others to get significant result.

Different clustering algorithms can be divided into following models:
– Distribution-based clustering;
– Density-based clustering;
– Hierarchical clustering;
– Lattice clustering.

As far as the aim of the paper is structuring the characteristics of moving physical objects it is necessary to highlight the system of informative parameters. Initially-formed a set of characteristics is modified with the help of the removal operator of less informative key parameters. The sets of objects determined by cluster analysis can be interpreted as quality variable at the heart of which is lied a qualitative variable. This variable is fixed from observations and establish hypothesis under the number of clusters.

Euclidian distance matrix represents similarity and difference of physical exponents under dynamic changes of moving physical objects. The smaller this value is the higher degree of similarities of exponents and combinations in cluster. And conversely the larger value is the bigger difference between physical exponents on dynamic changes of moving objects.

Methods of cluster analysis can be divided into two groups: hierarchical and non-hierarchical. The point of hierarchical methods is to successively merge smaller clusters into bigger or to split bigger clusters into smaller. Agglomerative and divisive approaches are used for these purposes.

Agglomerative hierarchical methods are characterized by successfully merger of initial data and diminution of clusters. At the first step similar items are merged into one cluster. At later steps the merger should be continued until all items are clustered into a single cluster.

Hierarchical method was chosen to build clustering algorithm of structuring logical characteristics of moving physical objects. Thereby there was a goal that items from the same cluster were similar in other words there was performed minimization of distances while items from another cluster had to be efficiently different in other words the distance between clusters was maximum.

The hierarchical process of merger is presented in dendrograms. An example of dendrogram is shown in Fig.1.

a)   б)

Fig.1. Hierarchical clustering dendrogram: a). initial point position;
b). Single-link clustering dendrogram..

A graph algorithm of clustering was based on the construction of the minimal spanning tree joined all points. There were eliminated the biggest edges from a graph after its construction. The remaining components would be clusters (Fig.2.).

*Fig.2. Minimal spanning tree and two clusters formed after elimination
of the biggest edge.*

The cluster analysis algorithm developed due to these positions is shown in Fig.3.

The usage of agglomerative hierarchical clustering for creation of an internal image of the system with maximum information about data density will allow to solve hard tasks in determination of key factors of moving objects.

The result of the research was the building of dendrogram with the help of STATISTICA 8 which gives an idea of number of clusters that merge physical characteristics under the dynamic changes of moving objects. It was chosen distance, speed and speedup of moving objects in a dynamic medium, motion drag and hydraulic and dynamic pressure as analyze characteristics.

Presence of abrupt jump of average data and changes of the dynamic factors can be interpreted as the number of clusters that exist in an investigated merge so at the step where value of a coefficient increases abruptly the grouping process of new clusters ought to be stopped otherwise there will be merged clusters that are far away from each other.

Presence of abrupt jump of average data and changes of the dynamic factors can be interpreted as the number of clusters that exist in an investigated merge so at the step where value of a coefficient increases abruptly the grouping process of new clusters ought to be stopped otherwise there will be merged clusters that are far away from each other.

Obtained results of the cluster analysis permit to range physical values under the dynamic changes of moving objects on the one hand in order of reliance on the other hand in order of efficiency. It should be emphasized that measures of efficiency characterize tactic development providing an opportunity of operable influence on proceeding process. Degree of reliability reduces passing from one cluster to other. Data of the firs cluster is considered to be the most reliable. According to efficiency of values of dynamic displacement the most efficient it is considered to be the final cluster merging the groups of similar physical values with common properties.
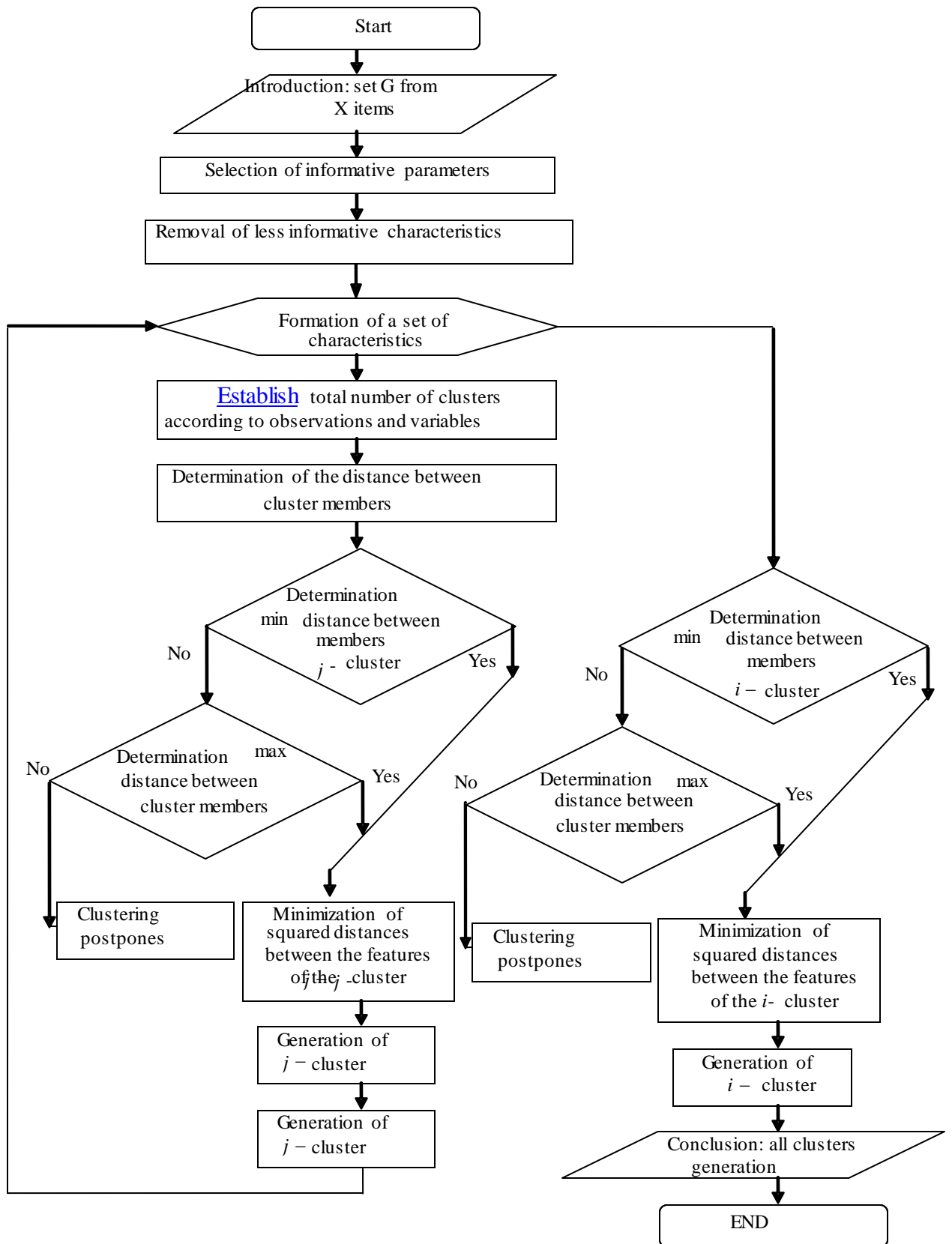
```
                          ┌─────────────────────┐
                          │       Start         │
                          └─────────────────────┘
                                    │
                     ╱──────────────────────────────╲
                     │  Introduction: set G from     │
                     │          X items              │
                     ╲──────────────────────────────╱
                                    │
                     ┌──────────────────────────────┐
                     │ Selection of informative parameters │
                     └──────────────────────────────┘
                                    │
                     ┌──────────────────────────────┐
                     │ Removal of less informative characteristics │
                     └──────────────────────────────┘
                                    │
              ◄─────────────────────┴──────────────────────►
                     Formation of a set of
                        characteristics
```

Start

Introduction: set G from X items

Selection of informative parameters

Removal of less informative characteristics

Formation of a set of characteristics

Establish total number of clusters according to observations and variables

Determination of the distance between cluster members

Determination min distance between members j - cluster

No   Yes

Determination max distance between cluster members

No   Yes

Clustering postpones

Minimization of squared distances between the features of the j -cluster

Generation of $j -$ cluster

Generation of $j -$ cluster

Determination min distance between members $i -$ cluster

No   Yes

Determination max distance between cluster members

No   Yes

Clustering postpones

Minimization of squared distances between the features of the i- cluster

Generation of $i -$ cluster

Conclusion: all clusters generation

END

Fig.3. Algorithm of agglomerative hierarchical clustering of moving physical objects.

Tree Diagram for 11 Variables
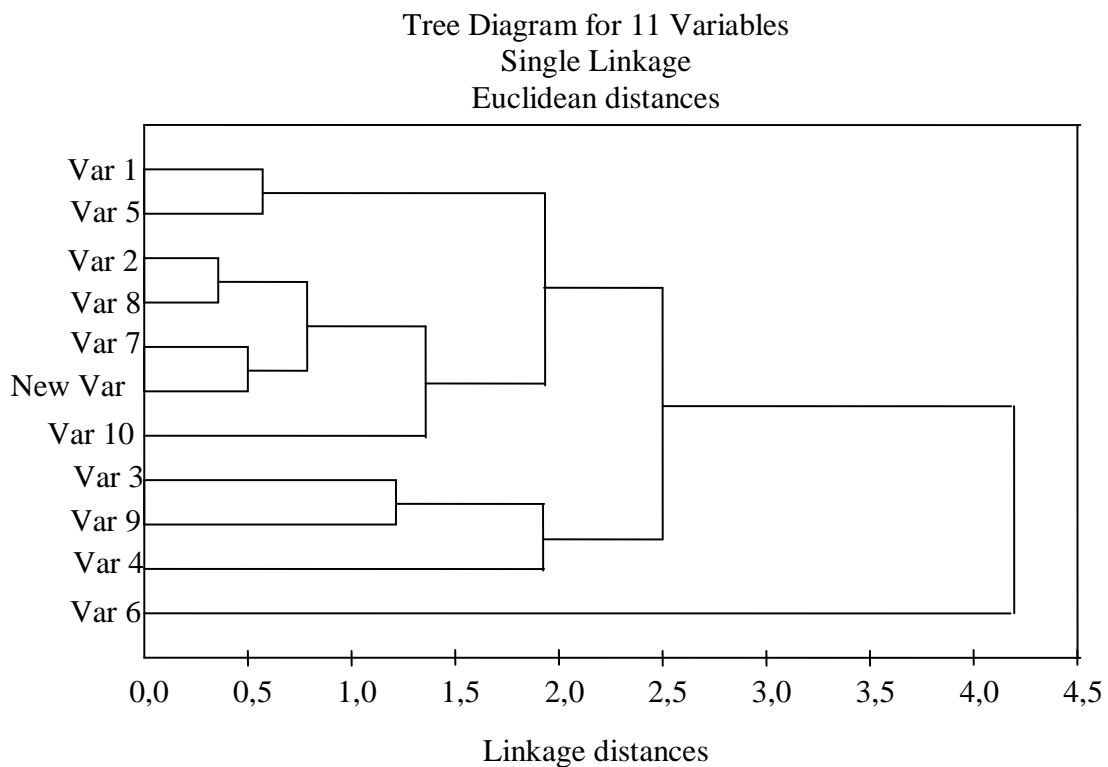Single Linkage
Euclidean distances

Linkage distances

Fig.4. Clustering of moving physical objective features.

Conclusion: The approach to management decision making with the help of cluster analysis permits to improve management system of moving objects in terms of adequate reaction to influence on key factors that affect characteristics of physical objects.

An application of cluster analysis permits to develop management decisions on which basis we can evaluate any strategic approaches of functioning physical objects. Besides, a choice of a shaped alternative is based on many factors that present integral or resultant characteristics. The main advantage of cluster analysis is in its using of factors relative to either internal or external mutual interference of physical properties of moving objects.

Informative representation usage of data density combined with the Euclidian distance matrix and minimization of the sum of squares at each step of clustering greatly improves its quality.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Мандель И. Д. Кластерный анализ/ И. Д. Мандель– М.: Финансы и статистика, 1988. – 176 с.

2. Брюханов В. В. Кластерный анализ как метод определения ключевых факторов/ В. В. Брюханов– ГОУ ВПО КГТЭИ, 2006. – С. 33-36.
3. Факторный, дискриминантный и кластерный анализ/[Дж. О. Ким, Ч. У. Мюллер, и др.]; Пер. с англ. – М.: Финансы и статистика, 1989. – 215 с.
4. Литвиненко В. И. Кластерный анализ данных на основе модифицированной иммунной сети / Литвиненко В. И. – УСим. – 2009. – С. 54-61.
5. Ward J. H. Hierarchical grouping to optimize an objective functions/ J. H. Ward // Journal of the American Statistical Association. – 1963. – 236 p.