СЕКЦІЯ 13. ЗАСТОСУВАННЯ КОМП'ЮТЕРНИХ ТЕХНОЛОГІЙ ДЛЯ ВИРІШЕННЯ НАУКОВИХ І СОЦІАЛЬНИХ ПРОБЛЕМ У МЕДИЦИНІ

ABOUT THE POSSIBILITY OF THE CLUSTER ANALYSIS APPLICATION FOR THE COMPARISON OF TIME SERIES

Antonova I.V., Chikina N.A.
National Technical University
«Kharkiv Polytechnic Institute», Kharkiv

When analyzing time series describing the dynamics of the same statistical indicator characterizing the state of the studied system, the actual task is to compare these series. The task is carried out with the aim of identifying among them groups with «similar» indicator dynamics. In order to obtain more complete information about the possible states of the object under study, the authors considered several eventual approaches to solving this problem.

Let the system state S in each time point t be characterized by the indicators $x(t), y(t),..., z(t)$. Identification of possible system states can be carried out either on each of these indicators or on their any set depending on research purposes. The similar situation takes place, for example, when determining the risk factors totality or forming the risk groups of diseases development, etc.

Suppose that there are time implementations $X_j(t) = \{x_{ij}(t)\}, j = \overline{1,m}, i = \overline{1,n}$ for the indicator $x(t)$. It is known that the image of the time series in the clusters form is one of the best representations of its main characteristics. Therefore, the solution of this task was carried out by means of the cluster analysis as a method revealing an internal structure of the data set. The K-means and Hierarchical clustering methods were applied. For the real available time series, quite close results were obtained for these two methods.

The results of cluster analysis were checked for adequacy. For this purpose, a pairwise comparison of the available time series was carried out for simple methods to evaluate their proximity. In this case, the usual Euclidean distance was used. Also of interest was the application of the method of identifying the longest common sequence on a set of specified time series and sequences with a general trend. According to the test results, it can be concluded that the quality of the results of the cluster analysis is quite high when solving the problem of time series comparing.

Such researches show that the joint learning of two or more time series makes it possible to describe the object of study more adequately. In addition, it allows one to construct a short-term forecast of such characteristics of its dynamics as time trends, bursts, and their duration and amplitude.