HIERARCHICAL AGGLOMERATIVE CLUSTER ANALYSIS OF ONE-DIMENSIONAL ASYMMETRICALLY DISTRIBUTED DATA IN THE ENVIRONMENT OF MS EXCEL

Kaminsky R.M., Nych L.N.
Information Systems and Networks Department, Lviv Polytechnic National University, S. Bandery Str., 12, Lviv, 79013, UKRAINE

The purpose of this study is to develop an approach to the implementation of the method of hierarchical agglomerative cluster analysis by means of spreadsheet Microsoft Excel 2003 and its later versions. Implementation of this method in the Excel environment solved a number of problems. These are: forming of the Table "object -property" normalization of values and bringing them into a single interval, calculate the values of matrix similarity determining the numerical values of the parameters dendrogram, dendrogram construction and allocation of required clusters.

This technique was used to split the group into subgroups of personnel carrier in terms of speed of decision making. All the band members have been trained on the same scenario. The script is a series of test images with the desired object search. The objects were test images localized randomly. The samples of these values were provided for cluster analysis as follows. To form the table "object-property" as objects of thirteen operators were featured as well as eleven properties of statistical characteristics of individual samples were given. Quantitative characteristics were obtained using descriptive statistics and parameters were defined by distributions of a histogram. To calculate the similarity matrix the table "object-property" and its copy were used. The table is the original fixed characteristics of a specific operator, which were compared with the characteristics of all other operators using the Euclidean metric. To determine the parameters a dendrogram of flexible strategies by Williams-Lance was used. For auxiliary the dendrogram clearly identified three clusters, which are very good in terms of compactness hypothesis. Comparison of clusters was made by calculating the average value of the properties which have physical significance and relationship to efficiency.

This technique is a new, easy to learn, has great practical value and complements the spreadsheet Excel significantly.

Keywords – cluster analysis, table object-property, proximity matrix, dendrogram, clusters.