Prognostication of Non-Stationary Temporal Rows by the Methods of Clusterization and Local Approximation

Boris Shamsha, Vitalij Ajvasov

Abstract- **In this paper we suggest the method of using the cluster analysis allowing the organization of clusters families and constructing local models of prognostication in every cluster.**

Keywords- **Forecasting, clusterization, local models**

I. INTRODUCTION

The results of the comparative analysis of clusterization methods efficiency are the foundation of the development of the prognostication algorithm. The essence of the comparative analysis consists in determination of clusterization algorithms efficiency with the presence of plenty of classes and features.

II. TASK SOLVING

The efficiency of popular clusterization methods which are present in many PPP was compared.

This choice is due to the necessity of solving the task of clusterization at the technical engineering level and to offer the grounded recommendations as to application of any method depending on the dynamic statistical characteristics of the non-stationary stochastic temporal rows. Eleven methods of clusterization were treated in the work. Using Kramer's conjugant coefficient (K_k), Hamming's coefficient (K_x) and the sum of inner-class dispersions considering all D features is suggested as quality characteristics.

It is shown that the increase of the member of classes along with good classification, the value of K_x is increased. Along with the increase of the number of features, K_x is decreased and it indicates bad classification. As the numbers of classis decreases and space dimension increases the quality of classification is decreased, K_k, K_x are diminished, coefficient D is increased. The dependence of quality criteria on space dimension is connected with the increase of the whole volume occupied by a cluster, enlarging distances on the absolute value. Due to this, general variation in the matrix of distances is increased. One should aim at processing low dimension arrays and select as many classes as possible - the results will be more reliable.The hierarchical algorithms of the agglomerative type have the best preciseness when the number of classes is great. When the number of classes is getting fewer the preciseness is going down. [1]

Algorithm K- average with a great member of classes is getting more variational than with a small one. This creates the opportunity to achieve much more higher classification quality. In various conditions the algorithm of dispersion increase minimization works well. The algorithm of the closest neighbour method gives bad results for the same cases. The algorithm of the far neighbour is considered to be a good one.

The centroidal method and the group-average method give identical results. As the number of clusters increases the efficiency of algorithms classification also increases. The agglomerative hierarchical cluster procedures appeared to be the most effective ones.

For referring the current information to any local model in real time we suggest using the discriminant analyses. Discriminant functions on the bases of many dynamic indicators are formal. The functions are orthogonal and allow to carry out canonical analysis. Discriminantfunctions for each cluster are determinated with the estimation of every function. [2]

III. CONCLUSION

As the result of our experimental researches we suggest to use the combined method of clusterization. It consists in consecutive joining up the closest elements at first and then the whole groups of elements which are more and more remote from each other.

REFERANCES

[1] Мандель И.Д. Кластерный анализ.— М.: Финансы и статистика. 1988.—176 с.

[2] Гуржій А.М., Шамша Б.В. та ін. Математичне забезпечення інформаційно-керуючих систем. – Харків: тов «компанія СМІТ», 2006. – 448с.

Boris Shamsha - Kharkiv National University of Radio Electronics, 14 Lenina Str., Kharkiv, 61166, UKRAINE Vitaliy Ayvazov - Kharkiv National University of Radio Electronics, 14 Lenina Str Kharkiv, 61166, UKRAINE.