DATA VISUALIZATION OF CLASTERIZED BY DYNAMICALLY-INTERVAL SELFORGANIZED MAP

O.V. Hodych, T. P. Mazepa
Lviv Polytechnic National University, Information Systems and Networks Department

Visualization of the cluster structure of high-dimensional data is one of the central research topics of data analysis and knowledge engineering. Self organized Kohonen maps are an efficient data clustering technology, which enable their visualization. However, the practical use of Kohonen maps points to a number of disadvantages:

1. static, pre-defined structure of the neural lattice, which leads to inadequate approximation of the data in conditions of the wrong choice of number of elements and the topology of the neighbourhood;

2. the mapping Φ, that is implemented by a trained Kohonen map, provides a "point" approximation of the data, and therefore, there is always an element of the best approximation making it difficult to identify the false classification;

3. the inability to finish Kohonen maps; in practical tasks there is a need to clarify the data model that in the case of Kohonen maps leads to a complete restructuring of the display Φ.

Dynamically-interval safe organized map (DISM) was developed to overcome these shortcomings. The ideas of the process of self-organization of Kohonen maps and the bases of interval analysis. Using interval weight vectors in the elements of the DISM allows to construct hypercube areas of the data space for their modeling. The construction method of interval vectors uses the hypothesis of λ-compactness that distinguishes technology DISM among other studies.

Among the disadvantages of the DISM in comparison with Kohonen maps is the lack of a structured two-dimensional grid of its elements. In the case of Kohonen maps, it provides the possibility to visualize the cluster structure of the simulated data in the plane, that facilitates the analysis of high-dimensional data.

Taking into consideration the topicality of data visualization, the purpose of research described in the article is to overcome this drawback. The objectives of the study are to develop a method of visualization of the cluster structure of the data simulated DISM and analysis of the adequacy of this method for high-dimensional data.

The developed method allows us to make a qualitative step in the direction of solving the issue of data visualization of the simulated DISM. The experiments indicate to adequate visualization of the cluster structure of high-dimensional data. At the same time, it seems as for Kohonen maps, the developed method does not possess the ability automatic binding of small clusters into larger clusters, that complicates the data analysis. To demonstrate this shortcoming is used the Iris dataset for which the map of DISM heightsis made.

It is easy to verify that using only this image is quite difficult to determine how many clusters are present in the data. Knowing that the Iris dataset contains three clusters, the height map can be divided into two boundaries are more intense color. Indeed, these boundaries delineate three clusters of data. In our previous study was developed by semi-automatic method of consolidation of data clusters for the Kohonen map. In further studies it is planned approbation of this method maps the DRIVE.

Additionally, it is needed to pay attention to the execution time of the developed method of DISM visualization in comparison with the technology of Kohonen maps. Taking into consideration the characteristics of the learning algorithm of DISM map, the acceptance time of Kohonen maps for the design of the DISM elements in two-dimensional space for all used data sets, was three times faster than building of Kohonen maps on the original datasets.

It is noticed that we have not performed an asymptotic estimation of algorithmic complexity of DISM teaching methods.This is one of the important tasks of further research and development of technologies for data analysis on the basis of the DISM.

The work solves the problem of visualization of the cluster structure of high-dimensional on the basis of data model received by dynamically-interval self organized map (DISM). The developed method visualization method uses Kohonen map to design elements of the DISM on the two-dimensional lattice in combination with the algorithm of U-Matrix for visualization of data clusters.

Keywords – data visualization, data analysis,self organized map.