# ER-MODEL OF CLASSIFICATION SYSTEM OF MULTIDIMENSIONAL DATA IN VECTOR GIS

*MESURA V.I.* *ZAVARZIN A.V.*

*Peculiarity of geographical data classification is also shown in the possibility of cartographical presentation of a process and results of the analysis. The given work contains one of the approaches to formalization of the automatic classification process of the multidimensional spatial data in GIS environment with the help of ER-model.The described model is used in software GisCluster. GisCluster developed by authors may function together with MS IIS web-server for distributing functions of spatial data classification, and also visualization of a process and results of the analysis through the Internet.*

## Model of classification of the multidimensional spatial data

The multidimensional spatial data is the set of the objects put simultaneously both in geographical, and in attributive spaces of factors. With the classification of such complex objects it is often necessary to take into account both spaces of factors. Besides that, geographical components existence allows to use additional opportunities of cartographical visualization.

Dealing with the problem of spatial data classification, authors came to a conclusion, that it is important not only to show the results of the analysis by cartographical or classical means, but also to have an opportunity to observe the process of classification.

During the construction of the visualization subsystem of the software GisCluster for multidimensional spatial data classification, the authors developed the model, allowing to combine means of visualization of the process and the results of analysis (cartographical and classical), algorithms, the aims of classification, initial factors, layers types of vector map etc.

The standard way of knowledge domain formalization is development of ER-model. For its construction it is necessary to display types of entires, their attributes and types of relationalships between them. In standard notation of Information Engineering (IE) the developed model with the help of Erwin software, is represented as follows:



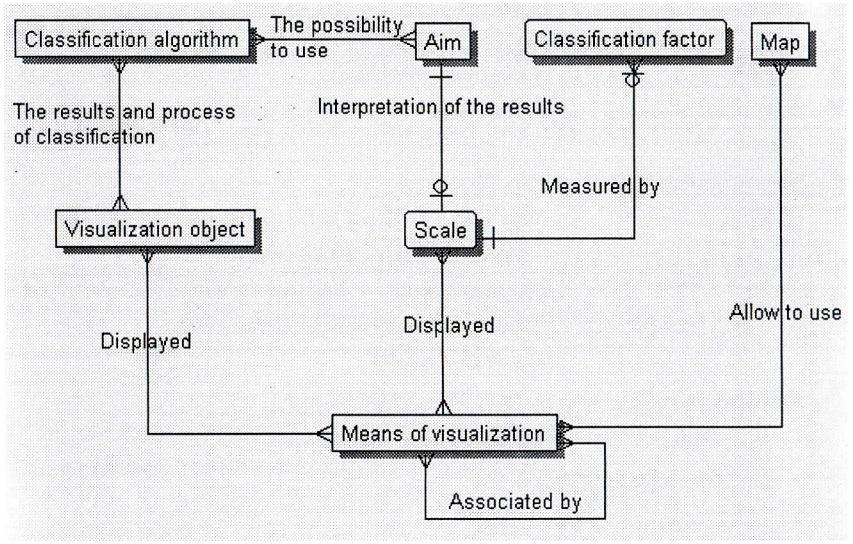


Figure 1. ER-model

For simplification of ER-diagram perception attributes of entires are not shown. Standard attributes of each entires are unique identifiers and names of instances. Additional attributes have entires "Means of visualization" and "Classification algorithm".

The attribute "Time" of " Classification Algorithm" entire signifies estimation of time complexity for a class algorithms. It is used for a choice of algorithm, when we have the big volumes of the data.

For distinction of cartographical and classical means of visualization the attribute "Map" is entered. The attribute "Multiregularity" specifies the possibility of visualization means to display simultaneously some factors. The attribute "Importance" or "Perceptibility" is set in expert mode and shows a degree of preference to use this or that means of visualization.

Let's give some examples of entires' instances.

| № | Name | Instances |
|---|---|---|
| 1 | Aim of classification | estimate, typology |
| 2 | Scale | name; rang; absolute |
| 3 | Algorithm of classification | hierarchical, core |
| 4 | Map layer type | point; line; polygon |
| 5 | Visualization object | cores; classification scheme; interim result; classification quality; hierarchical process |
| 6 | Factor | the amount of null voting papers; voting papers without marks; percentage of voices given for the winner of elections |
| 7 | Means of visualization | unique value; graduated color; dot density; graduated symbol; anamorphosis; isoline; bariers; label; bar chart; scatterplot; quality; line diagram; dendrogram |

Table 1. Some possible instances of entires

The most important instances of entires are objects of visualization, i.e. these things are shown during the process of algorithms working and after getting of results.

Relationships establish semantics of the model. Example, for each algorithm of classification it is possible to visualize the objects. Objects of visualization may be represented differently for a process and results of algorithms (for example, it is possible to visualize the cores of classes as dots in multidemensial space, using the scatterplot, or in geographical space using a map such as "unique value"). Thus relationships between means of visualization of the same objects are established. So, there is a relationship between the scatterplot and a map such as "unique value". While choosing an object on the scatterplot, the appropriate geographical object is marked on a map, and vice versa.Practical application of a model

Formalization has allowed to realize in system GisCluster necessary functions of visualization separately and correctly use them, depending on settings and users' data.

During the realization of classification, there is always initial data (attributes and vector map), and also the aim of classification is put. The model allows to determine:

• possible and most preferable ways of cartographical visualization of the given factor through definition of a factor scale and selection for this factor possible means



of visualization (relationship "measured by" and " displayed") with a restriction, put by the type of a map (relationship "allow to use");

• possible and most preferable ways of non cartographical visualization of the given factor;

• cartographical and classical ways of simultaneous visualization of several factors;

• set of possible and recommended algorithms of classification for the given aim of classification (relationship "the possibility to use");

• objects of visualization process and results of classification for the data with the use of certain algorithm (relationship "the process and results of classification");

• ways of visualization of exact step or result for the algorithm;

• presence and essence of relationship between classical and cartographical means of visualization.

Answers to the given questions can be received automatically with the help of typical SQL-queries formed beforehand (meaning, that ER-model is represented in a relational database).

With the use of SQL the first query, for example, looks as follows:

```
SELECT Visual.*
FROM Factor, Visual, Visual_Scale, Visual_Map
WHERE Factor.ID_Scale=Visual_Scale.ID_Scale
AND Visual_Scale.ID_Vis=Visual.ID
AND Visual_Map.ID_Vis=Visual.ID
AND Visual.Map=1 AND Factor.ID='ned'
AND Map.ID='polygon'
AND Visual.Multi =0
ORDER BY Visual.Imp DESC
```

The model may be easily supplemented with new algorithms of classification, with new means of visualization etc. In terms of ER-model it corresponds with introduction of new instances of entires, for relational model of data - with addition of records in existing relations.

**REFERENCES**

1. Thomas M.Connoly, Carolyn E.Beg. Database systems. A practical Approach to design, implementation and management. - University of Paisley, 2000, 1109 p.

*Mesura V.I. vimes@vstu.vinnica.ua*
*Zavarzin A.V. gisa@inbox.ru*