CartoSom – Cartogram creation using self-organizing maps

Roberto HENRIQUES, Fernando BAÇÃO and Victor LOBO

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa
Campus de Campolide
1070-312 Lisboa
Portugal

KEYWORDS

Self-organizing maps, cartograms, ….

ABSTRACT

The basic idea of a cartogram is to distort a map. This distortion comes from the substitution of area for some other variable (in most examples population). The objective is to scale each region according to the value it represents for the new variable, while keeping the map recognizable. The use of cartograms is previous to the use of computerized maps and computer visualization. The first cartograms were created to show the geographic distribution of population, in the context of human geography (Raisz, 1934). Typically, cartograms are applied to portrait demographic (Tobler, 1986), electoral (House & Kocmoud, 1998) and epidemiological data (Gusein-Zade & Tikunov, 1993). Cartograms can be seen as variants of a map. The difference between a map and a cartogram is the variable that defines the size of the regions. In a map this variable is the geographic area of the regions, while in the cartogram any other georeferenced variable may be used.

According to NCGIA (2002) cartograms can be divided into three types: non-continuous, contiguous and Dorling Cartograms. Non-continuous cartograms are the simplest cartograms in which regions does not preserve topology. In this type of cartograms each region is allowed to shrink or expand creating gaps or overlaps in the result map. Contiguous cartogram preserves topology causing great distortion in shape. Since a good cartogram must be recognizable, this type of cartogram has to balance these two factors. A Dorling cartogram maintains neither shape, topology nor region centroids. This type of cartogram, instead of enlarging or shrinking regions, creates new objects with a uniform shape, usually a circle, with the appropriate size.

The Self-organizing map (SOM) (Kohonen, 1982) was introduced in 1981 and is a neural network particularly suited for data clustering and data visualization. The SOM's basic idea is to map high-dimensional data into one or two dimensions, maintaining the most relevant features of the data patterns. The SOMs objective is to extract and illustrate the essential structures in a dataset through a map, usually known as U-matrix, resulting from an unsupervised learning process (Kaski & Lagus, 1996).

In this paper we present a new algorithm to create cartograms based on the SOM. Usually, when building cartograms, areas with a high value on the selected variable "grow", occupying the geographic space made available by areas with smaller values. This grow/shrink process is focused on creating an equal-density map where high values will be represented by larger areas, and small values by smaller areas. In our

proposal the cartogram will be created based on the neurons movement during the learning process of the SOM.

An example of this process is shown in Figure 1, showing a simplified example using only two regions. The two regions are geographically identical but have distinct values for the variable ***p*.** Assuming variable ***p*** to be population, the region color represents the value of population (dark color corresponds to a higher value of population). In the following step (Figure 1b) we randomly generate points (represented with triangles) inside each region. The number of points generated is a linear function of the value of population. In the Figure 1c a two-dimensional SOM (4 x 4) is initialized. A constant neuron density is used to initialize the SOM, contrary to the usual practice, in which the neurons are randomly initialized in the input space. After the training phase (Figure 1d) the neurons are labeled (Figure 1e). Figure 1f represents the mapping of each neuron to its initial position. A space transformation is then performed based on the neurons position and label resulting on a population cartogram (Figure 1h).

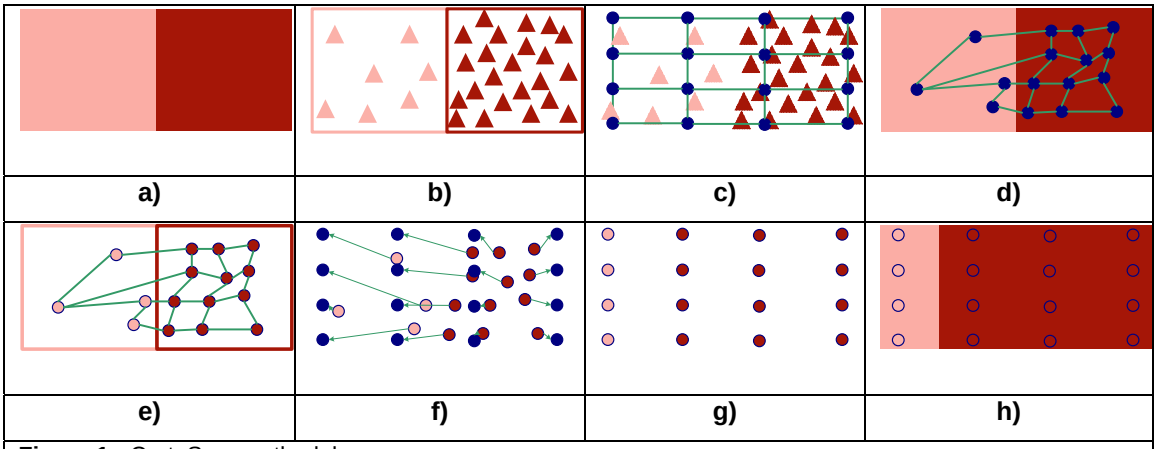

Figure 1 - CartoSom methodology

In this paper we analyze and discuss the problem of cartogram drawing based on the SOM. We propose a new cartogram drawing algorithm based on the SOM, and present the results of comparisons made with other methods using population data from the USA and Portugal.

REFERENCES

GUSEIN-ZADE S.; TIKUNOV, V. (1993) A New Technique for Constructing Continuous Cartograms. Cartography and Geographic Information Systems, 20 (3): 167-173

HOUSE, D.; KOCMOUD., C.(1998) Continuous cartogram construction. Proceedings of IEEE Visualization, 197-204.

KASKI, S.; LAGUS, K. (1996) Comparing self-organizing maps. in C. von der Malsburg, W. von Seelen, J. C. Vorbruggen, and B. Sendhoff, eds, Proceedings of ICANN96, International Conference on Artificial Neural Networks, Lecture Notes in Computer Science, vol. 1112, pages 809--814. Springer, Berlin, 1996 <http://www.cis.hut.fi/~sami/therest.html>

KOHONEN, T. (1982) Self-organizing formation of topologically correct feature maps. Biol. Cyb., 43 (1): 59-69.

NCGIA /USGS.Cartogram Central, 2002.
<http://www.ncgia.ucsb.edu/projects/Cartogram Central/index.html>

RAISZ, E. (1934) The rectangular statistical cartogram. The Geographical Review 24:292-296.

TOBLER, W. (1986) “Pseudo-Cartograms” The American Cartographer, 13,1 (1986), 43-50.