The Cluster of City Buildings Based on the SOM Neural Network

Z. Wang 1, Q. Liu 1
1. School of Resources and Environment, University of Electronic Science and Technology of China, 611731 Chengdu, China

E-mail: moniquewang@foxmail.com

Abstract. The clustering of buildings on the city map is an important field of cartographic and the main step to resolve issues related to the scale of the city map transformation process. The development of computer technology promotes the studies of cartographic and modelling algorithms. In order to achieve automatic clustering of map elements, the domestic and foreign scholars have proposed numerous automatic clustering algorithms. Although each algorithm works for specific problems, all algorithms are based on the graphical analysis transformation of building polygons on the map. By reading and analyzing current development of domestic and foreign researches, we find out that map building clustering needs more inspiring interactive features and automatic integrated decision-making simulation functions. It needs to improve the visualization software and integrated automatic intelligence, increase the proportion of the batch to support implementation of automated cartographic generalization to a higher level.

1. Introduction

Recently, with the growth of geographic information industry, the social demand of digital map is also increasing rapidly. Therefore, it is facing a higher challenge for the manual cartography. Map generalization is an important study area of cartography. It refers to the clustering of various elements, the operations such as mergers and simplification. The map generalization is essential to make a suitable map that can satisfy the requirements of the appropriate scale and actual needs when the map scale changes. The realization of the automatic map building clustering is the main step to provide conditions for the consolidation and simplification of cartography. It is not only conducive to streamline manual drawing operations and increase the productivity [1], but also it provides a technical support for scientific research in the field of cartography.

2. Methodology

2.1. SOM neural network

SOM refers to Self-Organizing Mapping, and SOM neural network algorithm is an unsupervised type of neural network algorithm [2]. It has been proposed by Teuvo Kohonen in 1981. The SOM algorithm only needs a few sample data and it implements automatic data clustering through clustering algorithm parameter settings [3]. Therefore, the algorithm has a high degree of intelligence. Typically, we use the competitive learning and self-organizing clustering feature of the SOM neural network to find the inherent similarity of the target experimental data. Without the requirement of human priori knowledge
constraints or parameters, the algorithm can achieve automatic clustering. Clustering results are completely in accordance with the internal data similarity, and it is not subject to external constraints. Moreover, it can reflect the real characteristics of the data very well. The two-dimensional structure of SOM neural networks [4] is shown in Figure 1.

![Figure 1. Structure of SOM neural networks](image)

The method of using SOM network algorithm to achieve the building clustering is proposed by Boyan Chen [5]. This method uses the coordinates of the gravity centres of the buildings to characterize the feature of building positions. By using the property of SOM network, buildings with close coordinates of gravity centres can be clustered as one group.

2.2. Introduction of building index

According to the understanding and recognition of building polygon features, this paper sets up 3 building indexes to extract characteristics of the buildings on the vector map [6]. After expressing different factors to characterize different features of the buildings, the effects of various factors on the course of realization of SOM clustering have been analysed and discussed.

2.2.1. Building area index. The building area index is defined and it refers to a single building polygon area. We directly use the ESRI's packed area calculation application programming interface to call the calculations. The building area index is added during the clustering process, as community buildings in the standardized town are always of similar building areas. Thus, standardized buildings will be clustered together in the artificial drawing process. The clustering process based on the building area index is shown in Figure 2.
2.2.2. Building density index. We make the current building’s center of gravity as the center of a circle and calculate the total area of the buildings within a radius $R$ of the circular. The definition of building density index is the area ratio, and the expression is shown in the formula 1. In the formula 1, the $Dens$ refers to the density index, the $SAT$ refers to the total area of the internal buildings in the circle, and $R$ represents the radius of the circle. In addition, the definition of radius $R$ is set in accordance with the scope of the map sheet in this paper. With the specific dimensions related to the expression shown in the formula, the density calculating method is shown in the formula 1.

$$Dens = \frac{SAT}{\pi \times R^2}$$  \(1\)

The radius formula is shown in formula 2, where the $MapS$ represents the map sheet range. The calculation results are obtained by the ESRI GIS .axMapControl1.MapScale interface.

$$R = \frac{5.0 \times MapS}{1000.0}$$  \(2\)

2.2.3. Building shape index. Building shape index refers to a flat rate of building polygons. The expression of building polygon shape index is shown in formula 3, where $l$ and $S$ are the perimeter and the area of the building polygon shape, respectively. The more flat of the polygon is, the smaller of the shape index of the building will be. There are three types of building polygon shapes, including compact type, circle and expanded form. Circle is the most special example of polygons, and its shape index is 1. If the shape factor is less than 1, we call the building is compact. If the shape factor is more than 1, the building is refer to the expanded form. Even if two polygons have the same building area, their shape factor could be very different. Introducing the shape index aims at improving the limitations of the area index. If the sizes are similar between the buildings and the differences between the shape factors are large, we need to consider whether to implement clustering though the building shape index. Besides, the influence of the shape index is shown in Figure 3.

$$Shp = \frac{l}{2(\pi \times S)^{1/2}}$$  \(3\)
2.2.4. The improved algorithm of building clustering. The experiment of building clustering using building index is divided into two steps. A preliminary clustering is implemented in accordance with the gravity centers of buildings. Then the secondary fine clustering is implemented based on the preliminary clustering result by considering other three index parameters. Before proceeding fine clustering, we need to normalize all three indexes [7]. This section carries on separate experiments with different indexes. According to the results of the experiments, the suitability of the indexes for building SOM clustering implementation is determined. The experimental flow chart is shown in Figure 4.

The improved algorithm is implemented by the secondary development of GIS based on C#. First of all, the improved algorithm needs to load the attribute information before the official building clustering. The attribute information includes various information of buildings and the details of the roads [8]. Second, the information is written into the attribute table of the map’s shp files. After creating the initial clustering attribute and a secondary attribute clustering result, the grouped information is written among the attributes of each building. Based on the demand of the experiment, five new properties, including other indexes information besides the coordinate information are created for the building attribute table. After exporting the property sheet shp file, a simple txt file with only three new indexes is created. Secondary clustering algorithm reads the data from the generated txt file which can be considered as input of the SOM neural network.

![Flowchart](image)

**Figure 3.** The building shape index

**Figure 4.** The flow chart of clustering

3. Results
We choose areas in downtown Boston with intensive-distribution buildings for experiment (Source: http://gis.dallascityhall.com/homepage/shapezip.htm). The size of cell of competitive layer in SOM
neural network is set to 40 * 40, depending on the amount of buildings, and the total iteration is 10000. The clustering results of the build-up area is much better than the normal areas, benefitting from the suitable property of centre coordinate factor to describe feature distribution, especially under the road’s constraints. However, from the results shown in Figure 5(a), we can see that the buildings of the eastern and northern part of the map sheets are clustered together due to the inconspicuous effects on gravity factor from the existing road constraints.

In the following experiment, we add area of the building for re-clustering. Clustering is to divide finely based on the original groups. The number of iterations and the size of competitive layer’s cell remains unchanged. Results are shown in Figure 5(b). Compared with the building area index in Figure 5(a), the buildings have been divided well and excessive clustering results have been avoided. However, mixed clustering results have been generated in the southeast of the building map. The reason for this is the areas of these buildings are relative close. During the regrouping, that random assignment of values would have a certain impact on the original results.

As to the results of density factor, the fine clustering results are still based on the performance of the centre of buildings’ gravity coordinates. We remain the parameter settings unchanged. The results are shown in Figure 5(c). As we can see, the building density is indeed performing better in built-up areas. Especially in the eastern and northern part of the map area, blocks with the similar buildings densities are clustered together. The blocks next to them with different densities are automatically grouped. The clustering itself maintains the initial clustering effect that use the centre coordinates of the buildings during the overall clustering process.

To avoid the limitation of area index, shape index is put forward. Shape index is the flat-ratio of building polygon in essence, and its clustering result is shown in Fig 5(d). In this map, we get local fine clustering for the building polygon with shape index.

![Figure 5.](image-url)
4. Conclusion
Clustering is achieved by utilizing the internal similarity of data [9]. This paper uses the morphological features of building polygon to get fine building clustering result. In order to show the different features of building polygon adequately, we define some indexes by describing building polygon status, including area index, density index and shape index in detail. Then the 3 indexes are used to get a fine building clustering in vector map by using self-organizing neural network algorithm respectively. The results show that in the context of a wide range of initial clustering, fine clustering is feasible and can meet the needs of manual cartography [10].

5. References
[1] Liu Q and Cheng B 2006 J. Geo-spatial Information Science 9 206
[2] Vesanto J, Alhoniemi E 2000 J. IEEE Transactions on Neural Networks 11 586
[3] Hagenbuchner M, Sperduti A and Ah Chung Tsoi 2003 J. IEEE Transactions on Neural Networks 14 491
[4] Bação F, Lobo V and Painho M 2005 J. Computers & Geosciences 31 155
[5] Cheng B, Liu Q and Li X 2013 J. Communications in Computer and Information Science 399 296
[6] Hequn Y, Yinning Y and Shengmao Z International Conference on Geoinformatics (IEEE) pp 1-5
[7] Argyrou A 2009 Advances in Self-Organizing Maps (Springer Berlin Heidelberg) pp 19–27
[8] Zhang LJ, Zhang F 2012 J. Advanced Materials Research 479-481 640
[9] Shekhar S and Xiong H 2008 Encyclopedia of GIS (Springer Nature) pp 68–71
[10] Sester M 2008 Self-Organizing Maps for Density-Preserving Reduction of Objects in Cartographic Generalization Self-Organising Maps (Wiley-Blackwell) pp 107–120