Point Set Generalization Based on the Kohonen Net

CAI Yongxiang  GUO Qingsheng

Abstract  Point set generalization is one of the essential problems in map generalization. On the demands analysis of point set generalization, this paper proposes a method to generalize point sets based on the Kohonen Net model; the standard SOM algorithm has been improved so as to preserve the spatial distribution properties of the original point set. Examples illustrate that this method suits the generalization of point sets.

Keywords  multipoint objects; map generalization; Kohonen Net; spatial distribution

Introduction

There are so many multipoint objects that form region scenes in maps, such as residential areas, islands and islets, and a group of small lakes. Point set generalization is the most essential problem in map generalization, which does not concern amalgamation but selection, and it must preserve the spatial distribution characteristic of the original objects while reducing them in number. Some algorithms for point set selection have been proposed. For example, with the help of convex hull algorithm, multiple embedded convex polygons have been constructed, which transform the selection of multipoint objects to the simplification of convex hulls\(^1\); the fractal transformation models of the Square Root Law based on fractal theory have been proposed to select grouped points\(^2\); the method of point cluster simplification based on Delaunay triangulation and the Voronoi diagram model has been researched by AI Tinghua et al\(^3\). Lately, the neural network technique has been introduced into map generalization to solve classification problems, such as classifying Britain’s residential areas\(^4\), distinguishing high density areas and then realizing object amalgamation\(^5\), clustering streets based on spatial characteristics\(^6\) and typification of buildings\(^7\), etc.

Selection and typification of point sets are researched based on the Kohonen Net model in this paper, and the standard SOM algorithm has been improved so as to preserve the spatial distribution properties of the original point set. The structure of this paper is listed as follows. Section 1 is the introduction of this paper. Section 2 discusses the generalization principles of point sets. Section 3 introduces the basic principles of the Kohonen Net. Section 4 focuses on the method of point set generalization based on the Kohonen Net. Examples are given in Section 5. Finally, a summary is presented in Section 6.

1  The principles of multipoint object generalization

If the real area of a certain object in the real world
is very small, then this object can only be represented by means of its spatial position. So many distributed points can represent small grouped objects, such as grouped islands and islets in a certain region, isolated houses along rivers or canals and so on. A point set refers to point objects of the same geography in a certain range of distribution, while scattering but having a certain distribution structure and pattern. A point set may be a relative whole or a part of the entire research area. In our research, the attributes of the points are not considered, i.e., all the points in a point set are equivalently important and have no differences in size.

There are mainly three distribution types: scattered distribution along line elements, scattered distribution in a region and regular distribution.

The generalization of point sets includes selection and typification. Selection means to select some points from the original point set, and typification means to use typical points instead of the original point set. These typical points may not be in the original point set. Certainly, some points may also be selected from the original point set, such as the points in the outline of the original point set. The generalization principles are listed as follows:\[1,3,9\].

1) The outline shape must be preserved. As the abstraction goes further, a point set may construct a virtual region in spatial cognition. The boundary of the virtual region is the outline of the point set.

2) The relative density and texture structure of the point set must be preserved. After generalization, the spatial distribution characteristics of the original point set must be preserved. The main indexes are relative density and texture structure.

2 The basic principles of SOM

The self-organizing map (SOM), namely, the Kohonen Net, was proposed by Teuvo Kohonen in 1981. It simulates brain mapping of sensory input and can effectively create the spatially organized ‘internal representations’ of various features of input data. The topological relationship among data points is optimally preserved by the mapping\[^{10}\]. The network structure of the Kohonen Net usually has two layers: input layer and output layer. The neurons of the output layer may be set as the nodes of the 1D, 2D or high-dimensional grid, and the input layer is interconnected with the output layer. Suppose there are \(D\) nodes in the input layer, in accordance with the \(D\) dimension properties of the input samples, which is described by input vector \(X = [x_1, x_2, \ldots, x_D]^T\); the \(M\) nodes of the output layer, in accordance with \(M\) neural cells, are connected in parallel to all the input nodes via different weight vectors \(W_i(i = 1, 2, \ldots, M; j = 1, 2, \ldots, D)\); the weight vectors are adapted according to the input point set during the self-organizing learning procedure. For the \(i\)th neuron in the output layer, its input is \(u_i = \sum_{j=1}^{D} w_{ij} x_j\), and its output is \(V_i = f(u_i)\). As a neuron stimulant function, \(f()\) usually adopts the linearity function.

The learning algorithm of the Kohonen Net is a series of iterative processes including competing, cooperating and updating. In the competing procedure, the neuron that has the nearest spacing between itself and the input vector will win. For different application purposes, it is possible to notice the position of the victorious neuron in the output grid or the updated weight vector. After biologic neurons are stimulated, the winner which is selected in the competing procedure not only strengthens itself, but also makes its neighboring neurons strengthened while restraining its ulterior neurons, which is the phenomena of on-center / off-surround, and is named cooperating. The victorious neuron is at the center of the topological neighborhood of the cooperating neurons. The neurons in the scope of the neighborhood are excited. The Gauss Function is a typical selection of topological neighborhood functions. In neighborhood initialization, the neighborhood includes almost all the neurons surrounding the winner. It declines exponentially with increasing iterative times. The updating is a self-adaptive procedure, adopting the transformation form of the Hebb Learning Rule to update the weight vector of neurons in the winner’s neighborhood.

The iterative processes finish till reaching the end threshold. As a result, the weight vector \(W_j\) of the
The $j$th neuron moves to vector $X$. With the procedure repeating, the updated vector weights tend to obey the distribution of the input vector, and are characterized by preserving the same topological neighborhood relationship with the input vectors.

3 Point set generalization based on the Kohonen Net

Once the SOM algorithm convergences, the characteristic mapping displays important statistical characteristics—simulating input space and density matching. In the neuron weight vector space, the higher input density region has higher density representation, and the lower density region has lower density representation, i.e., the relative density can be preserved.

The two important statistical features can be satisfied by two demands of point set generalization: one is reserving the density comparison of the original point set; the other is reserving the texture structure of the original point set. So it is reasonable to apply SOM to generalize the point set.

Point set generalization also needs to preserve the distributing outline. The mechanism of SOM learning signifies that when several vectors map the same neuron of the output space, its weight vector will move and converge to somewhere between these vectors. In other words, though the typification can preserve the relative density and texture structure of the original point set, its outline is hard to be preserved, and the phenomenon of ‘in-shrink’ may exist, i.e., it is possible for the point set to shift inward.

According to Gestalt principles in visual adjacency cognition, Ai Tinghua obtained the distribution range polygon of a point set by progressively stripping the outside triangles. This method preserves effectively the outline character of the point set. Referencing this method, the outline polygon is constructed to represent the distribution scope of the point set in this paper, and the Douglas-Peucker algorithm is used to simplify it. The self-organizing map can preserve the relative density and texture structure of the point set, and outline polygon simplification can preserve the distributive scope of the point set, so the two methods can be combined to generalize the point set. The detailed method is listed as follows: ① constructing the outline polygon, separating the original point set into outline points and inner points; ② using Douglas-Peucker algorithm to simplify the outline points; ③ using the inner points as input samples and using the Kohonen Net for characteristic mapping; ④ the union of results is the generalized point set. The ratios of the deleted outline points to all outline points are the same as the ratios of the deleted inner points to all inner points.

The simplification process of the inner point set is listed as follows.

1) Initialization. Setting some variables and parameters.

Sample dataset. Suppose there are $N$ points in the inner point set as input samples. Because we are considering position only, the spatial dimension of the samples is 2D, viz. The input layer has 2 nodes, which are the coordinates $x$ and $y$.

Output layer. The number of neurons in the output layer is determined according to the Square Root Law. Neurons are arranged in a 2D grid. Its array mode can be specified at will. The position of a neuron in the grid determines its topological neighborhood.

Weight vector space. Every neuron in the output layer connects with the input layer by a weight vector. All weight vectors construct the weight vector space. It is a 2D space, which is the same as the samples data space,

$$W_i = \{w_{i1}, w_{i2}\} \quad (i = 1, 2, \ldots, M).$$

Weight vectors need initialization in the beginning of learning, and we can set a random value in the scope of input vector distribution, or select $M$ samples from the samples space directly as the initial values of the weight vectors.

In addition, the initial learning rate $\eta_0$, maximum iterative times, the threshold value of $\sum_{i=1}^{M} \Delta W_i^2$, time constant $\tau_1$, $\tau_2$ and initial topological neighborhood $\sigma_0$ are determined by experience.

2) Training. The process of training is listed as follows.

① Initialization. Select $M$ samples from the sample space as the initial values of weight vectors.
\( W_j(0), \quad j = 1, \ldots, M. \)

2. **Sampling.** Sample \( X \) from the sample point set, take its coordinates as nodes of the input layer.

3. **Selecting the best-matching neuron.** Based on the Euclidean distances between every neuron vector of the output layer and sample \( X \), the minimum distance defines the ‘winner’ \( c. \)

4. **Updating the weight vectors.** By defining a neighborhood set \( N_c \) around the winner neuron \( C \), all the cells within \( N_c \) are updated. \( \eta(n) \) is the learning rate; \( h_{j,i}(n) \) is the neighborhood function around the winner neuron \( C \). The width or radius of \( N_c \) can be time-variable, in fact, \( \eta(n) \) and \( h_{j,i}(n) \) change dynamically in terms of formula (5) and (3) in order to get the best results.

5. **Continuing step 2 till meeting the end condition.**

During the iterative process, if \( \sum_{i=1}^{M} \Delta W_i^2 \) is less than a certain threshold value, it illuminates weights tending to converge, and the iterative process can be finished. A maximum iterative number is set as an end condition.

After finishing the mapping, the set of weight vectors \( \{ W_j \} \) are the typificated results of the input vectors. If the set is replaced by the most adjacent sample points, these sample points are the selected results.

4. **Applications in point set generalization**

To illustrate the above approach, applications in point set generalization were researched. Fig.1 is a case of a point set generalization where there are 51 points in the original point set (see Fig.1(a)). As the original 70\% of points are reserved after generalization, Fig.1(b) shows the comparison of the generalized result using the Kohonen Net with the original dataset; the circle symbols are the generalized points and the black triangle symbols are the original points. Fig.1(b) shows that though relative density is preserved, the distribution outline has a certain distortion.

According to the above approach, the original point set is classified into a convex hull point set and an inner point set to be simplified, respectively. The convex hull point set includes 14 points (see Fig.1(c)).

The Douglas-Peucker algorithm is used to simplify it. When reserving 70\% of points, viz. 10 points, the simplified result of the convex hull is Fig.1(d), and the circle symbols signify preserved points. The inner point set includes 37 points; for reserving 70\% of points, viz. 26 points, the comparison of the typified result of the inner point set with the original dataset is shown in Fig.1(e), and the comparison of the selected result of the inner point set with the original dataset is shown in Fig.1(f); the union of the generalized results of the inner point set and the simplified results of the convex hull point set is the final result; the final typification result is shown in Fig.1(g), and the final selection result is shown in Fig.1(h). However, in fact, there are differences among them. Fig.1(g) shows the improvement relative to Fig.1(b). The scope of outline and relative density are preserved well. Fig.1(i) is the generalized result of points decreased to 70\%; Fig.1(j) is the generalized result of points decreased to 40\%, and Fig.1(k) is the generalized result of points decreased to 20\%. All these generalized results can preserve the characteristics of spatial distribution in a vision.

In order to test if relative density is preserved, the density values before and after generalization can be computed and compared. Relative density preserving means that the bigger the region density is before generalization, the bigger its density is after the generalization. The density of a point is denoted by the reciprocal of the area of a Voronoi polygon around it\(^[3]\). The density values before and after generalization are computed directly during point selection. In points typification, the density after generalization can be computed directly, the density before generalization is represented by the average density of the region which the typificated point represents. Fig.2 is the map of density contraction before and after the generalization in the points selection (contrast Fig.1(h) to Fig.1(a)). 36 points are selected, the \( x \)-axis is the points No., the \( y \)-axis is the density, the dashed line is the density distribution of 36 points after selection, and the real line is the density distribution of 36 points in the original data. Fig.2 shows that high or low density can all be preserved after the selection. The method can preserve relative density in the point set generalization.
In addition, we experimented with grouped points distributing along a special direction and succeeded. Fig.3(a) is the original point set. It includes 36 points distributing along a line. Its typification needs to preserve the characteristics of its relative density contrast, so the Kohonen Net is used directly. When reserving 70% of the points, 25 points are selected, the black triangle symbols are the original points in Fig.3(b), and the circle symbols are sample points selected randomly from the sample space in initialization. On account of selecting randomly, their distribution does not reflect the space distribution pattern of the origi-
nal points well. Using the Kohonen Net to train these sample points, weight vectors are adjusted to the optimal position. The contraction of the mapping result and the original dataset is listed in Fig. 3(c). It could reflect the density contrast relation in the original points. In Fig. 3(c), some mapped points refer to the original positions of the sample points, and some points are located among several sample points and represent these points. This is the typification of a grouped point set. If typificated points are replaced by the most adjacent sample points, we can obtain the selection result and it is listed in Fig. 3(d); Fig. 3(e) is the selection result and Fig. 3(f) is the typification result. There is no distinct difference in the distribution patterns but a little offset in space is caused in some points. When reserving 45% of the original points, viz. 16 points, Fig. 3(g) is the result of typification and Fig. 3(h) is the result of selection. You can find that the two kinds of generalization results tend to be consistent as the map scale becomes smaller. This example illustrates that the generalized point distribution preserves the original relative density contraction and the original distribution tendency.

Fig. 4 is a trial which has an amount of relative points. The original dataset includes 213 points and is listed in Fig. 4(a); Fig. 4(b) is the result which preserves 70% of the original points, viz. 144 points. The result with a shrunk scale is shown in Fig. 4(c). It shows that the distribution scope is also the original
CAI Yongxiang, et al./Point Set Generalization Based on the Konhonen 227

Fig.4 Selection of a point set with spatial distribution remaining

polygon shape of the cognitive regions. The relative density is preserved well in view.

Parameters in the experiments are set as follows: initial learning rate is 0.1; maximum iterative time is 10 000; the threshold value of $\sum \Delta W_i$ is 0.000 001; $\tau_1$ is 2 000; $\tau_2$ is 3 000. The output layer arrays the 2D grid. In each row is a set of 12 cells and $\sigma_\sigma$ is a set of 10. In fact, the weights are stable with the iterative times till 7 000.

5 Conclusion

Kohonen Network mapping has the characteristics of approximate spatial distribution and relative density preservation. This paper combined the Kohonen mapping model with outline polygon simplification to generalize a point set to satisfy the demands of point set generalization. These demands include spatial outline shape preservation, relative density preservation and texture structure preservation. The examples show that the point set approach based on Kohonen mapping can preserve the characteristics of the original points’ spatial distribution. However, only single point set generalization is considered and these points are equally important. In order to solve practical problems, clustering must first be executed as multi-subgroup point sets are to be generalized, and the semantic characteristic of point objects and contexts should be taken into account.

References

[1] Wu Hehai(1997) Principe of convex hull and its applica-

tions in generalization of grouped point objects[J]. En
gineering of Surveying and Mapping, 6(1): 1-6

[2] Wang Qiao, Wu Hehai(1998) Research on fractal representa-
tion and automatic map generalization[M]. Wu
han:Wuhan Technical University of Surveying and Mapping Press

[3] Ai Tinghua, Liu Yaolin(2002) A method of point cluster simplification with spatial distribution properties preserved[J]. Acta Geodaetica et Cartographica Sinica, 31(2): 175-180

[4] Openshaw S, Blake M, Wymer C (1995) Using neuro-
computing methods to classify Britain’s residential areas [OL]. http://www.geog.leeds.ac.uk/papers/95-1/

[5] Allouche M K, Moulin B (2005) Amalgamation in cart-
ographic generalization using Kohonen’s feature nets[J]. International Journal of Geographical Information Science, 19(8-9): 899-914

[6] Jiang Bin, Harrie L(2004) Selection of streets from a network using self-organizing maps[J]. Transactions in GIS, 8(3): 335–350

[7] Hajholt P(1995) Generalization of build-up areas using Kohonen-networks[C]. Proceedings of Eurocarto XIII, Ispra, Italy

[8] Sester M (2005) Optimization approaches for generalization and abstraction[J]. International Journal of Geographical Information Science, 19(8-9): 871-897

[9] Guo Qingsheng(2002) Theories and methods of map automatic generalization[M].Beijing:Surveying and Mapping Press

[10] Simon H, Ye Shiwei (2004)Translation neural networks a comprehensive foundation (Second Edition)[M]. Beijing: China Machine Press