

# Use of Rough Sets Theory in Point Cluster and River Network Selection

Jia Qiu<sup>1</sup>, Ruisheng Wang<sup>1</sup>, Wenjing Li<sup>2,3</sup>

<sup>1</sup>Department of Geomatics Engineering, University of Calgary, Calgary, Canada

<sup>2</sup>School of Resource and Environmental Engineering, Wuhan University of Science and Technology, Wuhan, China

<sup>3</sup>Department of Geography, University of California at Santa Barbara, Santa Barbara, USA

Email: [jqiu@ucalgary.ca](mailto:jqiu@ucalgary.ca), [ruiswang@ucalgary.ca](mailto:ruiswang@ucalgary.ca), [liwenjing@wust.edu.cn](mailto:liwenjing@wust.edu.cn)

Received 25 April 2014; revised 20 May 2014; accepted 30 May 2014

Copyright © 2014 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

In this paper, we applied the rough sets to the point cluster and river network selection. In order to meet the requirements of rough sets, first, we structuralize and quantify the spatial information of objects by convex hull, triangulated irregular network (TIN), Voronoi diagram, etc.; second, we manually assign decisional attributes to the information table according to conditional attributes. In doing so, the spatial information and attribute information are integrated together to evaluate the importance of points and rivers by rough sets theory. Finally, we select the point cluster and the river network in a progressive manner. The experimental results show that our method is valid and effective. In comparison with previous work, our method has the advantage to adaptively consider the spatial and attribute information at the same time without any *a priori* knowledge.

## Keywords

Rough Sets Theory, Map Generalization, Point Cluster, River Network, Progressive Selection

---

## 1. Introduction

Point and line are the two most common objects to represent the spatial features and phenomena on the map, e.g. we use points to denote the village on small scale map, the point of interests (POIs) of the searching results in web or mobile mapping, we also employ lines to represent the road, river, as well as the region border on nearly all the maps. Generally, maps are generated to different scales for specific purposes. One of the most important techniques for scale transformation is selection, which can be used to reduce the complexity of map contents, emphasize the interesting phenomena, and delete the unimportant details while maintaining the geometrical and topological relations between objects. Acquiring geodata is much easier than it once was; especially as the

attribute information becomes increasingly available. This change gives the object selection an even wider implication—we should take into account not only the spatial characteristics of the objects, but also the attribute information during the process. Currently, most of the methods for object selection merely consider the attribute information. In considering of this, we proposed a novel method of using Rough Sets to combine the spatial and the attribute information together to evaluate the importance of spatial objects for progressive selection.

Rough sets is proposed by Z. Pawlak in 1990s, which can handle the incomplete and inaccurate data without relying on any *a priori* knowledge. In recent years, some researchers have tried to apply rough sets to cartographic generalization. Song *et al.* [1] employed the attribute reduction to mine the knowledge of river network selection. Li *et al.* [2]-[4] proposed a method based on rough set to generalize the point cluster as well as the river network. However, there are some limitations of their method, especially for applying the rough set to the non-decisional attribute information table. Our work is different from these two, we propose a strategy to assign decisional attribute for the information table, and use the dependency of attributes in rough sets to evaluate the importance of spatial objects for progressive selection.

The remainder of the paper is organized as follows: Section 2 surveys the related work. Section 3 outlines the main idea of object selection based on rough sets theory. Section 4 presents the experiments of the point cluster and river network selection. And Section 5 discusses the future application of our method.

## 2. Related Work

Research of point and line object selection in map generalization has a long history ranging from the traditional paper map to the electronic map on mobile device and web. Yan and Weibel [5] employ the Voronoi diagram for the point cluster selection. In their method, they constructed an equation taking into account the thematic importance as well as the local density to evaluate selection possibility. Unfortunately, this method lacks the ability to determine the weight of these two kinds of factors. Cai and Guo [6] introduced the Kohonen net to implement the selection of point set. The method can maintain the distribution feature of points but doesn't take the attribute information into account. De Berg *et al.* [7] proposed a method for the simplification of dot maps. They formalized the problem—derive coarse dot map from the fine one—by employing the concept of “ $\varepsilon$ -approximations.” And they can then compute good approximations of the derived and source maps by a heuristic method. The approach could be applied to the map with large amount of points while only consider the geometric information. Bereuter *et al.* [8] [9] proposed the concept of Real-time generalization for the points on the mobile device. They use the quadtree to manage the points with its “hierarchical subdivision structure” and “progressive levels of detail”. By employing the indices of the quadtree as auxiliary data, they refined the algorithms of selection, simplification, aggregation and displacement for the points. Their method focuses mainly on the efficiency and the distribution character of the points; merely taking into account the attribute information.

There are also many approaches have been developed regarding river selection. Ai *et al.* [10] [11] implemented an algorithm by employing the Voronoi diagram to calculate the catchment of the hydrographic network, and use the area of the catchment as a main factor to evaluate the importance of rivers, but the experimental results is not satisfying. Battenfield *et al.* [12]-[14] proposed a series methods for the generalization of National Hydrographic Dataset (NHD). Their approaches classify the hydrographic data by considering geographic factors—terrain and the climate, which include mountainous, hilly or flat, and dry or humid. Then, according to the characteristic of different types of hydrographic data, a researcher can set the processing sequence as well as the parameters of algorithms to generate coarse version of the fine hydrographic data. Hydrographic generalization results based on this method turn out to be quite effective. Sen *et al.* [15] [16] presented a method based on k-means clustering and Self Organizing Maps (SOM) to implement the generalization of the hydrographic data. This method enjoys the advantage that takes account of semantic information directly.

In addition to mainly focus on the geometrical information and the topological information, some researches about the generalization of river network address the thematic information. Dahinden and Sester [17] developed a method for the river network selection based on its importance. But the importance of the rivers is not determined by geometric or topologic features. It is in fact linked with “the fact and the frequency of their use”. The generalization result of this method is not stable, which highly related to the crowd sourcing information—the data in public repositories, and parameters of the assessment approach. Tinker *et al.* [18] also implement a hydrographic generalization based on the thematic information—the local mean annual precipitation.

Since the thematic information is acquired far more easily nowadays, we need to develop tools for the map generalization that can handle the spatial information (the geometrical, topological characteristic) and the attribute information (the thematic, semantic data) at the same time. In this paper, we propose a selection method based on the rough sets. Our method can integrate the spatial information and attribute information together to evaluate the importance of the spatial object, which is the criterion for the object selection. In addition, we carry out the selection process in a progressive manner to maintain the distribution consistency of spatial objects.

### 3. Objects Selection Based on Rough Sets Theory

The selection of point and line objects is based on their importance that evaluated by the spatial information and attribute information. The basic idea of employing rough sets theory to calculate the importance of geospatial objects is to find the dependency between attributes. Thus, we will introduce the concepts of rough sets theory, then, outline the method for information table construction, and propose the progressive selection algorithm.

#### 3.1. Basic Concepts of Rough Sets

According to [19], “the starting point of the rough sets theory is the indiscernibility relation, generated by information about objects of interest”. The indiscernibility relation is defined as an information table (also called an information system or an attribute-value table) in rough sets theory. A model of the information table can be defined as:  $K = (U, A, V, f)$ . Where  $U$  is a finite set of objects named as universe;  $A = C \cup D$  is a finite set of attributes, where  $C$  and  $D$  are conditional attribute subset and decisional attribute subset; for each attribute  $a \in A$ ,  $f_a : U \rightarrow V_a$  is the function for designating the value for  $a$ . An indiscernibility relation  $I(B)$ , where  $B \subset A$ , is  $I(B) = \{(x, y) \in U \times U : f_a(x) = f_a(y), \forall a \in B\}$ .  $U/I(B)$  in short  $U/B$ , is the family of all indiscernibility of  $I(B)$ . “Another important issue in data analysis is discovering dependencies between attributes” [19]. For subsets  $C \subset A$  and  $D \subset A$ , if  $k = \gamma_C(D) = |POS_C(D)|/|U|$ , then  $D$  depends on  $C$  in degree  $k$ ,  $0 \leq k \leq 1$ .  $POS_C(D)$  is the positive region of  $U/D$  with respect to  $C$ , which consists of all the elements of  $U$  that can be uniquely classified to blocks of  $U/D$  by subset  $C$ .

Take **Table 1** as illustration, six entities  $U = \{e_i | i \in [1, 6]\}$  are included. The conditional attributes and decisional attribute are  $C = \{c_1, c_2, c_3\}$  and  $D = \{d\}$  separately. According to the above mentioned definitions,  $U/C = \{\{e_1\}, \{e_2\}, \{e_3\}, \{e_4\}, \{e_5\}, \{e_6\}\}$ , and  $U/D = \{\{e_1, e_4, e_5\}, \{e_2, e_3, e_6\}\}$ . Let subset  $P = \{c_1, c_2\}$ , then  $U/P = \{\{e_1, e_2, e_3\}, \{e_4, e_6\}, \{e_5\}\}$ . In this case, we could conclude that the  $POS_C(D) = 6$ , and  $POS_P(D) = 1$ , which can be interpreted that all elements of  $U$  can be uniquely classified to the partition  $U/D$  by means of  $C$ , while only 1 can be classified by means of  $P$ . Then the decisional attribute  $d$  totally depends on attribute set  $C$ , and 1/6 depends on attribute set  $P$ . For each attribute  $c_i \in C$ , the importance of  $c_i$  for the decisional attribute  $D$  can be defined as  $\sigma_{(c_i, D)} = (\gamma_C(D) - \gamma_{C-c_i}(D)) / \gamma_C(D)$ , therefore, the importance of  $c_3$  is  $1 - 1/6 = 5/6$ . That means without attribute  $c_3$ , only 1/6 elements in  $U$  can be correctly classified, then the importance of attribute  $c_3$  is 5/6.

By employing the dependency of attributes, we can calculate the importance of each conditional attribute if we can construct an information table for the spatial objects which are to be selected.

**Table 1.** Information table of rough sets.

$U$	$c_1$	$c_2$	$c_3$	$d$
$e_1$	$(v_{11}) T$	$(v_{12}) T$	$(v_{13}) N$	$(v_{1d}) F$
$e_2$	$(v_{21}) T$	$(v_{22}) T$	$(v_{23}) H$	$(v_{2d}) T$
$e_3$	$(v_{31}) T$	$(v_{32}) T$	$(v_{33}) VH$	$(v_{3d}) T$
$e_4$	$(v_{41}) F$	$(v_{42}) T$	$(v_{43}) N$	$(v_{4d}) F$
$e_5$	$(v_{51}) F$	$(v_{52}) F$	$(v_{53}) H$	$(v_{5d}) F$
$e_6$	$(v_{61}) F$	$(v_{62}) T$	$(v_{63}) VH$	$(v_{6d}) T$

### 3.2. Information Table Construction

Attribute information always can be easily transformed to quantitative indexes. While the spatial information sometimes cannot be obtained directly, we need use structuralization to extract it. Structuralizing spatial objects means to recognize and quantify the topological relations between objects and the spatial distribution features of objects. There are many methods for the structuralization of map objects, such as convex hull [20], constrained TIN [21], Voronoi diagram [22] and so on.

#### 3.2.1. Structuralization of Point Cluster

Take the hotel distribution of a city in **Figure 1(a)** as example. All the data in our experiment is vector data that can be displayed at any scale. First, the nesting level of the convex hull is constructed, which is shown in **Figure 1(b)**. This information can represent the boundary structure of points since points on border have a greater impact on general structure than the inner points. Second, the Voronoi diagram, shown as **Figure 1(c)**, is employed to represent the affection domain of each point. The larger the area is, the more important the point tends to be. Third, the local density, shown as the gray area of **Figure 1(c)** and Equation (1), can be derived from the Voronoi diagram. The smaller the local density is, the more opportunities the point to be retained during selection.

$$D_i = (m+1) / \left( S_i + \sum_{k=1}^m S_k \right). \quad (1)$$

where  $D_i$  is the local density of point  $i$ ,  $S_i$  is the Voronoi area of point  $i$ ,  $m$  is the count of the neighbours of point  $i$ , and  $S_k$  is the Voronoi area of point  $i$ 's neighbour point  $k$ .

In addition, we take into account the relation between the road and the point, shown as **Figure 1(d)**. It is more convenient for the travellers to stay the hotel near by the main road; then we prefer to retain the point near to the main road. By employing these methods, we can quantify the spatial information of the point cluster—represent the border and inner of points by the level of convex hull, affection domain and local density by Voronoi area, and the location by distance to the road.

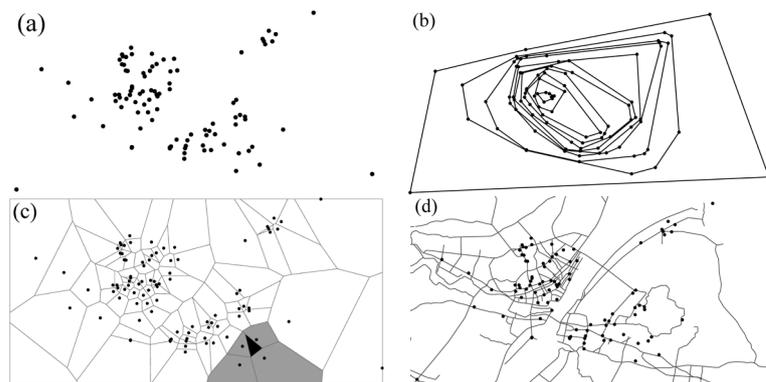
#### 3.2.2. Structuralization of River Network

The pattern of the river network is more complicated than the point cluster. In this paper, we focus on the tree-like river network, which consists of a trunk river and some tributaries. **Figure 2(a)** is an example of the river network which consists of 14 rivers. The blue river 1 is the main stream and others import into it.

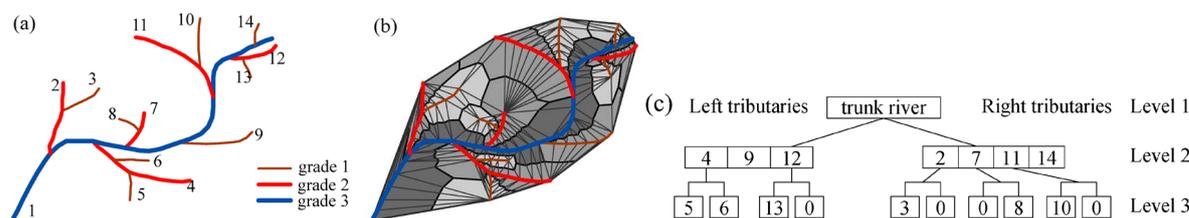
First, the length of each river should be taken into account to evaluate the importance of the rivers. Long rivers are always regarded as the geometrical skeleton of the network. Second, we use Horton code [23] as the grade to represent the hierarchy relationship between the rivers, which shown as **Figure 2(a)**. No rivers import into rivers with grade 1. The red rivers, with grade 2, have the rivers with grade 1 directly imported into. And the blue river with the largest grade has many rivers with grade 2 directly imported into. We couldn't remove the river before those rivers directly import into it are removed. Third, the level of the rivers showing in **Figure 2(c)** is employed to denote the tree-like structure of the rivers in this paper. The level is much like the binary tree, each node has at most two child nodes, distinguished as "left" and "right", and nodes with children are parent nodes. In the river network, for each river the left tributaries are regarded as its left node, and right tributaries are the right node. Rivers in the same node are siblings to each other. For example, in this river network, river 4, 9, 12 are siblings to each other, river 1 is the common parent of these rivers, and river 4 has the left child river 5, the right child river 6. Deletion of the river with a higher level is more likely to destroy the tree-like structure than the deletion of a river with lower level. Then we prefer to select rivers with a high level. Fourth, the catchment of the rivers [11], showing in **Figure 2(b)**, is employed as an index as well. We use the Voronoi diagram of the rivers, which generated by TIN, to partition the space, and regard the Voronoi area as the catchment of the river. Furthermore, the catchment of the parent river owns all its children's catchment. This index is a synthetical index which could represent not only the hierarchy relation between rivers, but also their intervals and local density. In addition, we introduce another index—a location metric—to reflect the location of rivers related to their parents in the tree-like structure. The index is shown as Equation (2).

$$L_i = L/n. \quad (2)$$

where  $L_i$  is the location metric of river  $i$ ,  $L$  is the length of river  $i$ 's parent, and  $n$  is the amount of river  $i$ 's sib-



**Figure 1.** Structuralization of point cluster. (a) The hotel distribution. (b) the illustration of convex hull. (c) the Voronoi diagram and local density. And (d) the distance to the nearby road.



**Figure 2.** Structuralization of river network. (a) the illustration of the grade of the rivers. (b) the catchment area of the rivers. And (c) the level of the rivers.

lings in the tree. According to this definition, rivers that are siblings to each have the same location metric. The index is interpreted as follow: if two rivers import the same parent from the different sides, then the river with fewer siblings tends to be more important than the other river with more siblings regardless of all other indexes.

By the structuralization of river network, we can quantify the spatial information of the river network into several parameters—describe the tree-like structure by the Horton code (grade) and tree level, the catchment (or affection domain) by Voronoi area, and the location by location metric.

### 3.2.3. Information Table of Points and Rivers

After structuralization, we can construct the information table as **Table 2** and **Table 3** by adding the attribute information. In **Table 2**, the “Grade” and the “Number of Rooms” are attribute information of hotels. Grade represents the star-level of the hotel. This is also a synthetical index that manually generated according to the facilities, services, management, and *et al.* we can get this index via the internet. In addition, the number of the rooms can reflex the scale of the hotel immediately. In **Table 3**, we didn’t add any additional attribute information to this table. That is because the attribute information of the rivers is not easily obtainable. But we could add that kind of information to the table if it is available, such as the precipitation of the catchment.

According to rough sets, to calculate the importance of conditional attributes, a decisional attribute should be contained. In the case of spatial object selection, the decision is to decide which objects should be deleted and which should be selected. But in **Table 2** and **Table 3**, all attributes are conditional attribute—we cannot decide which object should be retained according only one of them. Therefore, we must assign the decisional attribute manually in order to conduct the selection operation. In general, we assign a fuzzy decisional attribute to each record by the value of all conditional attributes. The procedure is shown as below.

First, discretize all continuous attributes since the rough sets is designed to handle the discretized data. Second, normalize all values to positive kind—the large the better—by Equation (3) if they are negative ones. Third, multiply all values of conditional attributes to deduct a fuzzy importance (Equation (4)) of spatial objects. And last, assign a decisional attribute for all objects. The details will be presented in Section 4.

$$v'_{ik} = \max_{1 \leq j \leq n} \{v_{jk}\} - v_{ik} + 1. \tag{3}$$

**Table 2.** Information table for hotel selection.

$U$	$c_1$	$c_2$	$c_3$	$c_4$	$c_5$	$c_6$
	convex hull layer	local density	Voronoi area	distance to road	grade	number of rooms
$e_i$	$v_{i1}$	$v_{i2}$	$v_{i3}$	$v_{i4}$	$v_{i5}$	$v_{i6}$

**Table 3.** Information table for river network selection.

$U$	$c_1$	$c_2$	$c_3$	$c_4$	$c_5$
	length	grade	level	catchment	location
$e_i$	$v_{i1}$	$v_{i2}$	$v_{i3}$	$v_{i4}$	$v_{i5}$

where  $v_{ik}$  is the discretized value of the attribute  $k$  of object  $i$ .  $v'_{ik}$  is the value after normalization,  $n$  is the amount of objects, and  $\max_{1 \leq j \leq n} \{v_{jk}\}$  is the maximum value of attribute  $k$  among all objects before normalization.

$$W_{f,i} = \prod_{k=1}^m v'_{ik} . \tag{4}$$

where  $W_{f,i}$  is the fuzzy importance of object  $i$ ,  $m$  is the amount of the conditional attribute.

After we add the decisional attribute to the table according to the fuzzy importance of objects, we will get the information table with decisional attribute for objects selection. Then we can apply the rough sets theory to the information table for evaluating the importance of the spatial objects.

## 4. Experiment

In this section, we use a hotel distribution of a city and a tree-like river network as the data set to test our algorithm. In order to maintain the distribution features of the objects, we proposed a progressive selection method. This turns out to be effective when comparing the selection results of automatic and manual.

### 4.1. Importance Evaluation of Map Objects

Take **Figure 1(a)** as illustration to explain the general procedure of importance evaluation. There are 96 points represent 96 hotels in total. According to Section 3, the information table of the hotel selection can be constructed as **Table 4**, the unit of local density, Voronoi area, and distance to road are  $km^{-1}$ ,  $km^2$ , and  $km$  separately.

First, we employ *k-means* algorithm to discretize the continuous attributes. In this process, the continuous attributes are classified into 4 categories, which approximate to the category of the grade of hotels. After normalizing the attribute to positive index according to Equation (3), the fuzzy importance of hotels can be calculated by Equation (4). Now, we can manually assign the decisional attribute of hotels by their fuzzy importance: sort descending all objects by their fuzzy importance; take the selection amount as boundary to set the decisional attribute as 3 to part of objects within this boundary, and 1 to part of objects outside the boundary, and 2 to the rest. This operation is rough due to the fact that we only need to know which objects have more opportunities to be selected. The selection amount is calculated by the ‘‘Radical Law’’ [24] shown as Equation (5).

$$N_2 = N_1 \sqrt{M_1/M_2} . \tag{5}$$

where  $N_1$  is the number of objects on the source map,  $N_2$  is the number of objects on the derived map,  $M_1$  is the scale denominator of the source map, and  $M_2$  is the scale denominator of the derived map.

Now, we can obtain an information table with decisional attribute. And the rough sets theory can be applied to calculate the weight of each conditional attribute according to Part 1 of Section 3. The importance of objects will be evaluated by Equation (6).

$$W_i = \sum_{k=1}^m (v_{ik} \times \sigma_k) . \tag{6}$$

where  $W_i$  is the importance of object  $i$ ,  $\sigma_k$  is the weight of the conditional attribute  $k$ ,  $v_{ik}$  is the value of attribute  $k$  of object  $i$  after discretization and normalization,  $m$  is the amount of conditional attributes. The value of  $W_i$

**Table 4.** Information table of the hotel selection.

ID	convex hull layer	local density	Voronoi area	distance to road	grade	number of rooms
1	10	0.0669	14.5825	0.0881	3	281
2	11	0.0504	33.4751	0.0603	1	63
.....	.....	.....	.....	.....	.....	.....
95	10	0.0569	18.4036	0.2629	1	80
96	11	0.0284	47.4667	0.1427	3	162

represents the importance of the hotel and can be regarded as the inference to conduct the selection operation.

## 4.2. Procedure of Progressive Selection

As the importance of each object has calculated, the information about which objects have greater opportunities to be retained during selection can be clearly obtained. But we cannot simply remove all objects with a relatively low importance. The removal of objects can change the local distribution where these objects are located, *i.e.* the Voronoi area, the local density, as well as their convex hull level. For example, we generalize the scale of the hotel to half of its source by deleting all points whose order is outside the selection amount (objects are sorted descending by their importance), and the result is shown as **Figure 3(b)**. We can observe that this operation has destroyed the distribution feature of the hotel—most of hotels located in the center of the map have removed.

Therefore, we proposed a progressive method for the selection of point clusters and river networks. To conduct the selection, first we set a mark for all the objects: “unset” denotes the object has not considered currently; “removed” denotes the objects should be removed after selection; and “retain” denotes not remove temporarily. The general procedure of the progressive selection is presented as **Figure 4**.

The first-order adjacent neighbours of points and rivers are shown as **Figure 5**. The neighbour of the point is derived from the TIN: those points directly connected with the point by the edge of a triangle. For the point  $P$  in **Figure 5(a)**, its neighbours are  $A$ ,  $B$ ,  $C$ ,  $D$ , and  $E$ . The neighbour of the river is derived from the level of the river network: those siblings directly adjacent to the river without any other rivers import to their parent between their influxes. For the river 5 in **Figure 5(b)**, its neighbours are river 4 and 6.

Due to the hierarchy structure, the progressive selection of river networks is slightly different from the process of the point cluster. We only use rivers in the leaf node of the tree to construct information table in terms of we cannot remove a river before all its tributaries are removed. The progressive selection procedure iteratively constructs the information table to calculate the importance of the spatial objects. In addition to taking into account the spatial information and the attribute information of spatial objects, it considers the dynamic characteristic of the selection process.

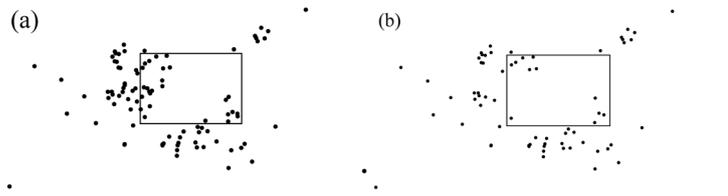
## 4.3. Point Cluster and River Network Selection

There are 96 points in the source map shown as **Figure 1(a)**. Then we generalize the map scale to a half and a quarter of the source scale, 68 and 48 points will be retained after generalization according to the Equation (6). The results are shown as **Figure 6**. In comparison with a selection strategy that removes points at one time, the progressive approach can retain the distribution characteristic of the point cluster properly.

For the river network, the experimental data is shown as **Figure 7(a)** [25]. 99 rivers exist in the source map. Then we generalize the scale to half of the original by our method, which result is shown as **Figure 7(b)**. 70 rivers are retained in the derived map. **Figure 7(c)** [25] is a manual selection result in the same scale as **Figure 7(b)**, 68 rivers are retained. Comparing these two maps, we can observe that there is only a slight difference between automatic selection and manual selection (indicated by circles). The method based on rough sets and progressive selection can reflect the main structure of river network. It also outperforms the method proposed by Ai *et al.* [11] with the same data set, which is shown in **Figure 7(d)**.

## 5. Conclusion and Future Work

In this research project, we applied the rough sets theory to the map objects selection directly. We extracted and quantified the spatial information of objects, and manually assigned a decisional attribute for the objects according to their spatial and attribute information. By doing so, we constructed a complete information table for

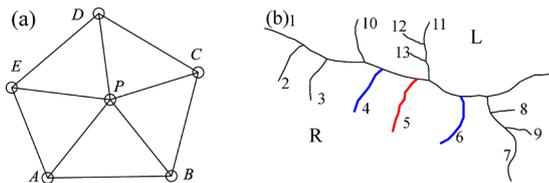


**Figure 3.** Result of select at one time. (a) the source map of hotel distribution. And (b) the hotel distribution with half of its source scale by selecting at one time.

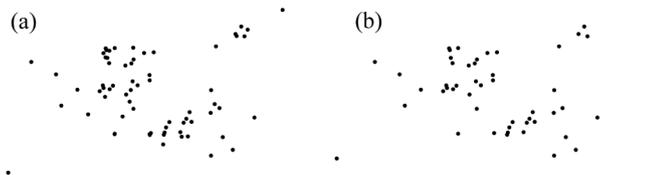
**Algorithm 1:** Progressive selection of map objects

- 1 Initialize the mark of all elements in  $U$  to ‘unset’,  $m$  to 0, and  $k$  to 1;
- 2 Construct the information table by elements in  $U$  whose mark is ‘unset’;
- 3 Calculate the importance of the elements in  $U$  by leveraging rough sets theory;
- 4 Sort descending all elements;
- 5 Get  $e_k$  from  $U$ ;
- 6 **if** mark of  $e_k$  is ‘unset’ **then**
- 7     set mark of  $e_k$  to ‘removed’;
- 8     set mark of all  $e_k$ ’s first-order adjacent neighbours (Figure 5) to ‘retain’;
- 9     set  $m = m + 1, k = k + 1$ ;
- 10    **if**  $m == N_1 - N_2$  **then**
- 11     end the procedure;
- 12    **if**  $k > N_1 - N_2$  **then**
- 13     set all objects whose marker is ‘retain’ to ‘unset’;
- 14     go to step 2;
- 15 **else if** mark of  $e_k$  is ‘retain’ **then**
- 16     set  $k = k + 1$ ;
- 17     **if**  $k > N_1 - N_2$  **then**
- 18         set all objects whose marker is ‘retain’ to ‘unset’;
- 19         go to step 2;
- 20     **else**
- 21         repeat step 5;

**Figure 4.** Procedure of the progressive selection of map objects.

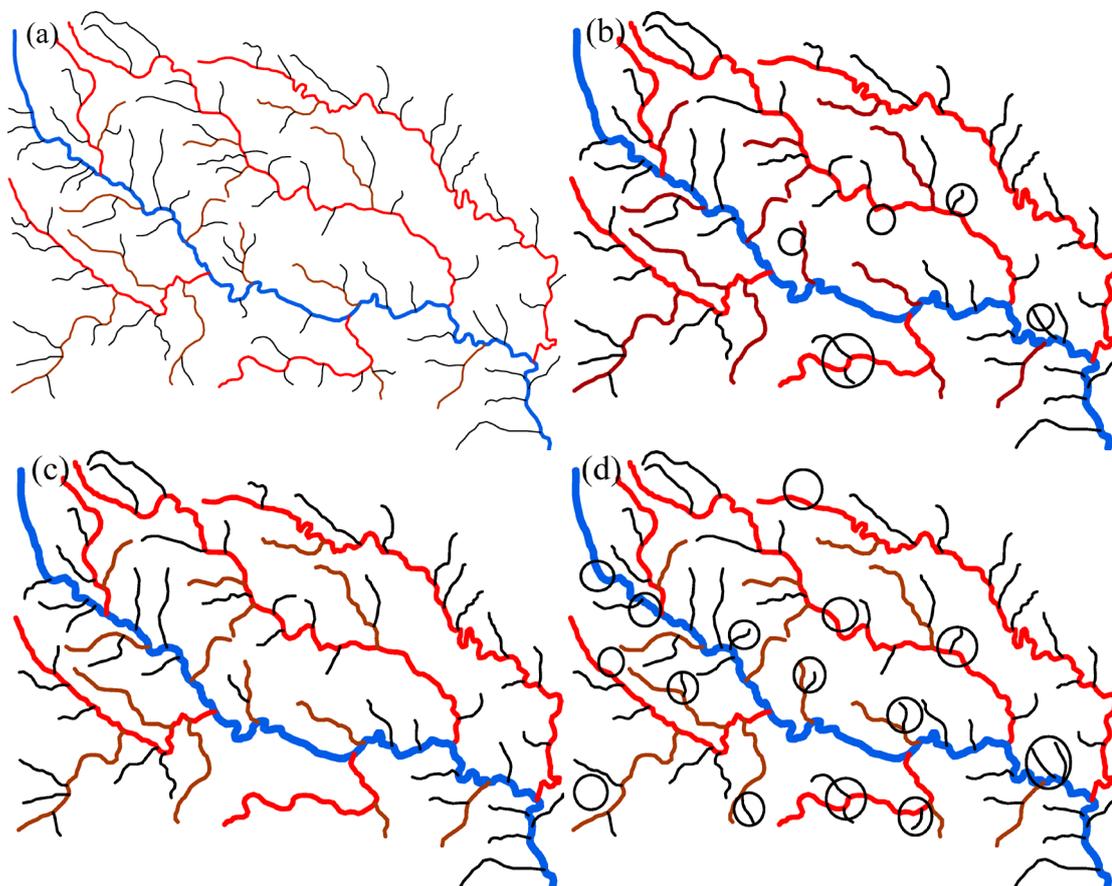


**Figure 5.** Illustration of neighbours of object. (a) the first-order adjacent neighbours of point. And (b) first-order adjacent neighbours of river.



**Figure 6.** Hotel selection results. (a) The hotel map with half of the source scale. (b) The hotel map with quarter of the source scale.

the rough sets application—evaluating the importance of spatial objects. In addition, we proposed a progressive selection procedure based on the importance. The method has the capacity to take into account the spatial information as well as the attribute information in the meantime without any *a priori* knowledge. And the use of



**Figure 7.** River network selection results. (a) the original map. (b) the map with half of the original scale selected by our method. (c) the map with the same scale as (b) selected manually. And (d) the automatic selection result by Ai et al.

progressive selection techniques can take into full consideration the change of the spatial information during selection process and maintain the distribution pattern of objects after selection.

In our method, the most important process is to quantify the spatial information of the objects. In other words, if we can quantify the spatial information of a kind of spatial entities, we then could evaluate the entities via the rough sets theory to perform a progressive selection. On the map, the road network and the contour cluster are also represented as lines. But their spatial characteristics are quite different from the tree-like river network. Many methods have been developed for recognizing the structured spatial information of these objects, e.g. the “stroke” [22] [26] or the “block” [27] of the road network, the “curve set” [28] of the map. How to apply our method to these two kinds of spatial features is a topic for further research.

## Acknowledgements

This work was supported by the National Science Foundation of China under Grant No. 41271449.

## References

- [1] Song, Y., Yu, D. and Shen, C. (2009) Knowledge Acquisition Model of Map Generalization Based on Granular Computing. *Proceedings of SPIE International Symposium on Spatial Analysis, Spatial-Temporal Data Modeling, and Data Mining*, 7492, South Africa. <http://dx.doi.org/10.1117/12.838333>
- [2] Li, W., Wang, H., Lin, Z., et al. (2009) A Study of Point Objects Generalization Based on Rough Sets Technology. *Proceedings of 17th International Conference on Geoinformatics*, Fairfax, 12-14 August 2009, 1-6.
- [3] Qiu, J. and Li, W. (2010) River Network Dynamic Selection with Spatial Data and Attribute Data Based on Rough Set. *Proceedings 18th International Conference on Geoinformatics*, Beijing, 18-20 June 2010, 1-5.

- [4] Li, W., Qiu, J., Wu, Z., *et al.* (2011) Application of Rough Sets in GIS Generalization. *Proceedings of 6th International Conference on Rough Sets and Knowledge Technology*, **6954**, 347-353. [http://dx.doi.org/10.1007/978-3-642-24425-4\\_45](http://dx.doi.org/10.1007/978-3-642-24425-4_45)
- [5] Yan, H. and Weibel, R. (2008) An Algorithm for Point Cluster Generalization Based on the Voronoi Diagram. *Computers & Geosciences*, **34**, 939-954. <http://dx.doi.org/10.1016/j.cageo.2007.07.008>
- [6] Cai, Y. and Guo, Q. (2012) Point Set Generalization Based on the Kohonen Net. *Geospatial Information Science*, **11**, 221-227. <http://dx.doi.org/10.1007/s11806-008-0091-y>
- [7] De Berg, M., Bose, P., Cheong, O., *et al.* (2004) On Simplifying Dot Maps. *Computational Geometry: Theory and Applications*, **27**, 43-62. <http://dx.doi.org/10.1016/j.comgeo.2003.07.005>
- [8] Bereuter, P. and Weibel, R. (2013) Real-Time Generalization of Point Data in Mobile and Web Mapping Using Quadtrees. *Cartography and Geographic Information Science*, **40**, 271-281. <http://dx.doi.org/10.1016/j.comgeo.2003.07.005>
- [9] Bereuter, P., Weibel, R. and Burghardt, D. (2012) Content Zooming and Exploration for Mobile Maps. *Proceedings of the AGILE'2012 International Conference on Geographic Information Science*, Avignon, 24-27 April, 2012
- [10] Ai, T., Ai, B. and Huang, Y. (2009) Multi-Scale Representaion of Hydrographic Network Data for Progressive Transmission Over Web. *Proceedings of 24th International Cartographic Conference*, Santiago, Chile.
- [11] Ai, T., Liu, Y. and Chen, J. (2006) The Hierarchical Watershed Partitioning and Data Simplification of River Network. *Proceedings of 12th International Symposium on Spatial Data Handling*, Vienna, 617-632.
- [12] Bittenfield, B.P., Stanislawski, L.V. and Brewer, C.A. (2013) Adapting Generalization Tools to Physiographic Diversity for the United States National Hydrography Dataset. *Cartography and Geographic Information Science*, **38**, 289-301. <http://dx.doi.org/10.1559/15230406382289>
- [13] Stanislawski, L.V. and Bittenfield, B.P. (2013) Hydrographic Generalization Tailored to Dry Mountainous Regions. *Cartography and Geographic Information Science*, **38**, 117-125. <http://dx.doi.org/10.1559/15230406382117>
- [14] Stanislawski, L.V. (2009) Feature Pruning by Upstream Drainage Area to Support Automated Generalization of the United States National Hydrography Dataset. *Computer, Environment and Urban Systems*, **33**, 325-333. <http://dx.doi.org/10.1559/15230406382117>
- [15] Sen, A. and Gokgoz, T. (2012) Clustering Approaches for Hydrographic Generalization. *Proceedings of GIS Ostrava 2012—Surface Models for Geoscience*, Ostrava.
- [16] Sen, A., Gokgoz, T. and Sester, M. (2014) Model Generalization of Two Different Drainage Patterns by Self-Organizing Maps. *Cartography and Geographic Information Science*, **41**, 151-165. <http://dx.doi.org/10.1080/15230406.2013.877231>
- [17] Dahinden, T. and Sester, M. (2009) Categorization of Linear Objects for Map Generalization Using Geocoded Articles of a Knowledge Repository. *Proceedings of 9th International Conference on Spatial Information Theory*, Aber Wrach'h, 21-25 September 2009.
- [18] Tinker, M., Anthamatten, P., Simley, J. and Finn, M.P. (2013) A Method to Generalize Stream Flowlines in Small-Scale Maps by a Variable Flow-based Pruning Threshold. *Cartography and Geographic Information Science*, **40**, 444-457. <http://dx.doi.org/10.1080/15230406.2013.801701>
- [19] Pawlak, Z. (1997) Rough Set Approach to Knowledge-Based Decision Support. *European Journal of Operational Research*, **99**, 48-57. [http://dx.doi.org/10.1016/S0377-2217\(96\)00382-7](http://dx.doi.org/10.1016/S0377-2217(96)00382-7)
- [20] Wu, H. (1997) Structured Approach to Implementing Automatic Cartographic Generalization. *Proceedings of 18th International Cartographic Conference*, Stockholm, 23-27 June 1997, 349-356.
- [21] Li, J. and Ai, T. (2010) A Triangulated Spatial Model for Detection of Spatial Characteristics of GIS Data. *Proceedings of 2010 IEEE International Conference on Progress in Informatics and Computing*, Shanghai, 10-12 December 2010, 155-159.
- [22] Liu, X., Zhan, F.B. and Ai, T. (2010) Road Selection Based on Voronoi Diagrams and “Strokes” in Map Generalization. *International Journal of Applied Earth Observation and Geoinformation*, **12**, S194-S202.
- [23] Horton, R.E. (1945) Erosional Development of Streams and Their Drainage Basins; Hydrophysical Approach to Quantitative Morphology. *Geological Society of America Bulletin*, **56**, 275-370. [http://dx.doi.org/10.1130/0016-7606\(1945\)56\[275:EDOSAT\]2.0.CO;2](http://dx.doi.org/10.1130/0016-7606(1945)56[275:EDOSAT]2.0.CO;2)
- [24] Topfer, F. and Pillewizer, W. (1966) The Principles of Selection. *The Cartographic Journal*, **3**, 10-16. <http://dx.doi.org/10.1179/caj.1966.3.1.10>
- [25] Zhu, G. (2004) *Cartography*. Wuhan University Press, Wuhan.
- [26] Thomson, R.C. (2006) The “Stroke” Concept in Geographic Network Generalization and Analysis. *Proceedings of*

12th International Symposium on Spatial Data Handling, Vienna, 12-14 July 2006, 681-697.

- [27] Gulgen, F. and Gokgoz, T. (2010) A Block-Based Selection Method for Road Network Generalization. *International Journal of Digital Earth*, **4**, 133-153. <http://dx.doi.org/10.1080/17538947.2010.489972>
- [28] Li, W., Qiu, J., Lin, Z., et al. (2013) Approach of Curve Bends Recognition and Contour Cluster Structuralization. *Acta Geodaetica et Cartographica Sinica*, **42**, 295-303.