A STUDY OF CARTOGRAPHIC GENERALIZATION BASED ON ROUGH SET METHOD

Wenjing Li
College of Resources and Environmental Engineering
Wuhan University of Science and Technology
947 Hepin Road, Wuhan 430081, China

Zhiyong Lin
School of Remote Sensing and Information Engineering
Wuhan University
129 Luoyu Road ,Wuhan 430079,China

Yi Long
College of Geographical Science
Nanjing Normal University
Nanjing 210097 , China

Abstract: This paper introduces how to apply rough set classification and attribute reduction method in map generalization and provides solutions to several special cases arising in the map spatial data processing. Unlike the analysis in the past which focused solely on the attribute data of map, in this paper, rough set attribute decision table specific for the features of map data is created on the basis of the structurization of spatial information in the map data, thus to meet the need for map generalization. Based on existing map generalization methods and aiming to their deficiencies, this paper proposes map generalization methods and models by employing the rough set classification, analyzes the relationship between the classification using rough set technique and the principles of map generalization, and from theoretical basis to application, expatiates on the common ground of the two and the feasibility of map generalization using rough sets.

Keywords: map generalization; rough set; classification; attribute decision-making table

1. Introduction

Rough set theory, first proposed by Polish scientist Zdzislaw Pawlak in 1982, is a new mathematical tool to deal with imperfect knowledge [1]. In recent years, scholars in the fields of cartography and GIS gradually have introduced the theories and methods of rough set into the geo-information science and tried to develop the geo-rough space theory of spatial information. By use of the rough set theory, integrating other theories and techniques, and at the meantime considering the uncertainties of spatial information, many research achievements have been made towards map generalization, spatial data mining and so on [2,3].

Rough set theory is a data mining method based on the classification. It makes analysis of attribute items and examples in the attribute decision table, removes unimportant attributes and examples which has no effect on classification, and finally gets a reductional attribute decision table and the minimal decision rules. This process is in fact a reduction of original mass data. Map generalization is the process of mining the needed information or data under the relevant scales among mass spatial data, and the result of map generalization doesn't affect the original description of an area, so it is in essence a kind of data reduction. From the perspective of data reduction, rough set theory and map generation bear intrinsic similarities.

2. Relationship Between Rough Sets and Map Generalization

Rough set theory is not only a new mathematical tool to deal with imperfect knowledge but also an approach to settle fuzzy problems. Its principles and methods have already been introduced in details in many relevant literatures [4], this paper will give no future explanation.

In brief, map generalization refers to the generalization of small scale data from known large scale data when the scale becomes smaller. Here, generalization is not a simple selection but to retain the most important information and omit less important information based on the preservation of the most essential features of spatial objects. The goal of map generalization is to seek out the most important and most representative information and retain it during the generalization. How to decide the importance of map objects (geographical objects) is one of the key problems. The importance of a map object is decided by the real geographical object it represents, while the importance of a real geographical object is determined by spatial attributes and thematic attributes of it. In the face of mass geographic data, the automatic data processing by computer is the most possible solution. Taking rough sets as the new tool of data mining is able to solve this problem. The attribute reduction founded on rough set classification is able to extract the most important attributes describing map objects through calculating the degrees of importance of all attributes, which are impact factors that change the classification of real objects most.

In rough sets, the most important attributes are the ones that can best manifest the differences between real objects during classification. In the process of map generalization, the most important attributes are the ones that can best reveal the features of geographical objects, namely, which can distinguish geographical object and the distinction can best reflect the features of the object.

3. Rough Set Representation of Map Data

From the basics of rough set theory we know that the data representation of attribute decision table cannot be directly applied to map data and GIS data. Therefore, in order to handle map data by rough set method, a common ground of the two must be found. Here we can not only describe map data by the forms of map data itself and through corresponding modification of rough set, but also represent map data as relational table based on the rough set attribute decision table.

The typical example of using rough set to preserve the data form of map and directly describe geographical information is to process raster map by rough set. As shown in Figure 1, if each pixel of raster graphics is seen as a member of rough set universe, all pixels of raster graphics can be deemed as a rough set universe with grey values of each pixel as its attribute values and the types of pixels as decision values. In this way, the graphic information of map can be directly represented by rough set. The lower approximation of the set is the pixel set surely of a type of geographic element, the boundary region, as the pixel set that may or may not belong to the element, corresponds to the actual boundary of a space, and the upper approximation, being the pixel set that may belong to the element set, corresponds to all possible pixels of the space. The method that directly represents the attributes of spatial data by rough set takes pixels of raster graphics as basic units and determines which type of geographic element it belongs to according to the grey values and other attribute values of pixels. At present, the method is still limited to the raster graphics processing and the result of processing is often seen as image recognition.

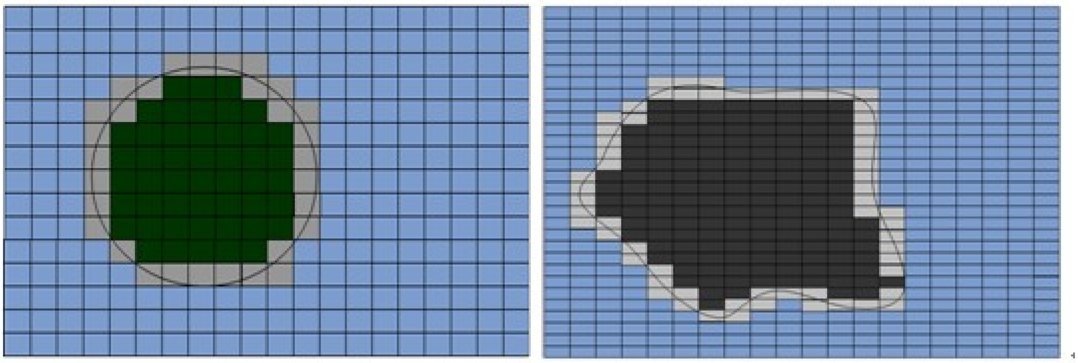

Fig.1 Raster images described by rough set method

Another approach to describe map data by rough set is based on relational table, which is very convenient for processing map attribute data, but for spatial data, especially the representation of the structural feature of space, which can’t be applied directly. Naturally, many present researches aim to the analysis of map attribute data by rough set. However, both spatial data and attribute data are necessary for every space; it is improper to investigate them separately, or that only one is highlighted and the other is disregarded. It is especially the case for map generalization. In this paper, the spatial data and attribute data are integrated to make study of map generalization, and the method adopted is to structurize map objects.

3.1 Structurization of Map Objects

It is a special and important step to structurize map objects in the process of map generalization. Through the structurization of space, all kinds of spatial information that can't be described in the rough set attribute decision table formerly is converted into the forms that can be represented by attribute decision table.

In order to generalize the spatial information, a global evaluation must be carried out to spatial objects, that is, to obtain the distribution rules and structural relations of spatial objects. This step is completed through the structurization of spatial data. Map data structurization describes the hierarchies of spatial object to get the state of its global distribution and region features. The result of structurization can not only evaluate the structural features of spatial objects on the whole, but also represent the adjacency between related objects; it not only has regard to global structure but also considers region features. The processing of data set by rough set need to establish corresponding attribute decision table, but listing coordinates as attribute-values of the table can't represent the structural features of spatial objects, so we need another method to represent spatial data, and this method is the structurization of spatial data.

3.2 Representation of Map Data by Attribute Decision Table

In rough set theory, the objects in universe is represented by attribute decision table, which is both a kind of special but important knowledge representation system and a kind of special information table. It indicates how to decide when certain conditions are met. Therefore, attribute decision table is also an information system $S = \langle U, R, V, f \rangle$; $R = C \cup D$ is a set of attributes, and subsets C and D are called condition attribute set and decision attribute set respectively. If decision attributes in a decision table contain only one attribute, the table is single decision; or else it is multi-decision. A multi-decision table can be converted into a single decision table.

Tab.1 Attribute decision-making table of rough set

Set	Condition attributes			Decision attribute	
U	X_1	X_2	X_3	Y_1	Y_2

The attribute decision table is generally a two-dimensional table (see Table 1) describing an object in a row and an attribute of an object in a column. Attributes are divided into condition attribute and decision attribute. The objects in universe are classified according as their attribute values.

The analysis of geographical objects by rough set is to list spatial objects in rows and attribute data of these objects in columns. This shows that the map generalization based on rough set theory takes spatial objects as map objects. The attribution information for describing spatial objects falls into two parts: information about objects and relational information between objects. The information about objects includes: semantic information (type, class and significance etc.), geometrical information (location, length, area and shape etc.), thematic attribute and quantity and quality index and so forth. The relational information between objects includes: semantic relation (administrative membership, thematic significance etc.) and spatial relation (adjacency, relevancy, degree of node, inclusion, connectivity relation and neighborhood relation etc.). All kinds of information about spatial objects are described in rough-set attribute decision table, as shown in Table 2. In this way, various analysis results of spatial objects are obtained by direct operation of rough sets. In Table 2, the description of every spatial object includes not only attribute data but spatial data as well. Thus, spatial objects can be evaluated and selected scientifically and reasonably during map generalization by taking full account of all kinds of information.

Tab.2 Attribution-decision making table of spatial objects

Set	Information of entities			Information between entities		expand attributes
Entities	semantic information	geometrical information	thematic attribute and so forth	semantic relation	spatial relation	Other special attribute information

4. Rough Set Based Map Data Processing

4.1 Map Data Preprocessing

Usually, the original map data in the database can’t be directly used for knowledge acquisition, and they need preprocessing, which comprises of imperfect information completion, redundant noise elimination, conversion between quantitative data and qualitative data, and data discretization etc.

4.1.1 Data Discretization

The rough set method need to discretize information in the decision table (e.g. integer type, character string, enumerated) and correspondingly grade the map data. Map data grading has a close relation with the application of map. Since spatial data includes a mass of statistical and measured data and continuous-value data, the discretization of spatial data is the key link of preprocessing in the map generalization by rough set. To put it simple, the discretization of rough set is divided into two steps: the first step is to convert quantitative continuous data (e.g. floating point data representation) into discretized data representation; the second phase is to abstractly merge discrete data to get the discrete value of higher level, which is equivalent to a process of data classification. Due to the particularity of map representation, the result of division by mathematical methods will be eventually converted to practical grading line. For example, if the grading line calculated mathematically is 990, it should be set as 1000 in the application of map, for the reason that the commonly-used grading line is the multiples of 5 or 10 and adjacent series increase by geometric series. This is also for the purpose of meeting people's need for visual discrimination of visualized maps.

4.1.2 Imperfectness of Data

Imperfectness of data is a concept in a broad sense. Compared with an ideal information system, except for the situation that all attribute values are known and unique, we encounter the problem of imperfect data. The complicated spatial objects and limited epistemic means of people bring difficulties to get the perfect geo-information. The imperfect data in the map system for the main has two classes: missing data and absent data. The missing attributes are existent but being missed for some reason and they will be obtained in the future; the absent attributes are unavailable. Generally, imperfect data are completed by statistical methods, but for map data processing, its region features should be taken into full consideration and human intervention is needed to a certain extent to ensure the reasonableness of its geographical connotation.

4.1.3 Other Preprocessing Issues

Map data has a very rich source, which brings convenience to the representation and expression of geo-information, but difficulties to the analysis and processing of data. How to gather data with different formats, different scales and for different purposes, eliminate the conflicts between them, and get the needed information to the maximum extent is a problem demanding prompt solution. For example, there might be many tables in the map database and attributes relevant to an object might be saved in different tables, so how to gather these attributes to describe this object is a such problem; also, due to the fact that epistemic means are different, the same attribute of the same object may have different descriptions, for example, a river sometimes is abstracted as an element of river system and sometimes an administrative boundary.

4.2 Special Situations of Decision Attributes of Map Data

Decision attribute value is an important part of rough set attribute decision table. Decision attributes need being used for data preprocessing, attribute reduction and decision-making. A decision attribute table often composes of several condition attributes and a decision attribute; however, for map data, other situations may emerge: one situation is with multiple decision attributes and another is without obvious decision attribute (Table 3).

Tab.3 Attribution-decision making table without decision-making information

U	c_1	c_2	c_3	c_4
x_1	1	1	1	1
x_2	1	2	2	1
x_3	2	2	1	1
x_4	3	3	3	3

For the first situation, two methods can be employed: first, generate a new attribute decision sub tables for each decision attribute with all condition attributes, and if there are *n* decision attributes, there will be *n* attribute decision sub tables; second, generate a new decision attribute by combining all decision attributes. Both of the two methods can convert the multi-decision information to be the single decision information and then carry out a series of calculation. When there is no obvious decision attribute, assume that the decisions of each example are different, and perform attribute reduction using discernibility matrix to obtain the simplest logic rules (Table 4). The result can be used for the automatic grading and classification of map attribute data either.

Tab.4 discernibility matrix without decision-making information

	x_1	x_2	x_3	x_4
x_1	0			
x_2	c_2, c_3	0		
x_3	c_1, c_2	c_1, c_3	0	
x_4	c_1, c_2, c_3, c_4	c_1, c_2, c_3, c_4	c_1, c_2, c_3, c_4	0

4.3 Structurization of Map Data

Structure recognition is a problem that must be solved for map data analysis, especially for map generalization. It should be said that map data structurization is a continuation of preprocessing by rough set method, aiming to the unique spatial data in the map system. With respect to vector graphics, spatial data is expressed by the line drawing information based on points, lines and polygons; but in the map database, it is saved by a series of (x, y) coordinates, and this data storage mode can’t be described by the attribute decision table of rough set. Meanwhile, spatial features of map objects can’t be obtained through these seemingly independent coordinates. Likewise, for raster graphics, spatial features and spatial relations of geographical objects can’t be obtained from location information and grey values of each pixel. How to make computer “understand” these graphics and spatial relations between them and then extract spatial features that are easily understood and calculated by computer is the problem to be solved by the structurization of map data.

Since the structurization of map objects is for the purpose of analyzing the distribution and structural features of map objects, the rules which must be considered during structurization are how to preserve contour shape, density comparison and texture structure of map objects. The result of the structurization of spatial objects can’t be directly used in rough set method, because it is still graphic information, which must be converted to attribute values that can be processed in attribute table. A spatial object has at least three basic space attribute data: own spatial semantic information, location information, and shape feature information. Other spatial information may also be included, for example, if a point has no shape information, Voronoi diagram is used for describing the ability of space occupancy by points. Linear objects often contain some special semantic information, such as whether a river is open to navigation. Topological adjacency relations can be established between connected area objects.

5. Map Generalization Model Based on Rough Set

The structurized map data can be saved in the attribute table as attribute values, which provide a basis for the processing by rough set. Non-spatial information attribute table can be created through discretizing map attribute data; in addition, through structurization, spatial data can also create a spatial information attribute table. By doing so, the existing map data can be so converted that it can be processed by rough set. And then analyze by use of attribute reduction to derive the decision rules for map generalization. By integrating other rules of map generalization, map objects are processed according to classification decision rules; check whether the results meet various requirements and output visual results (see Figure 2).

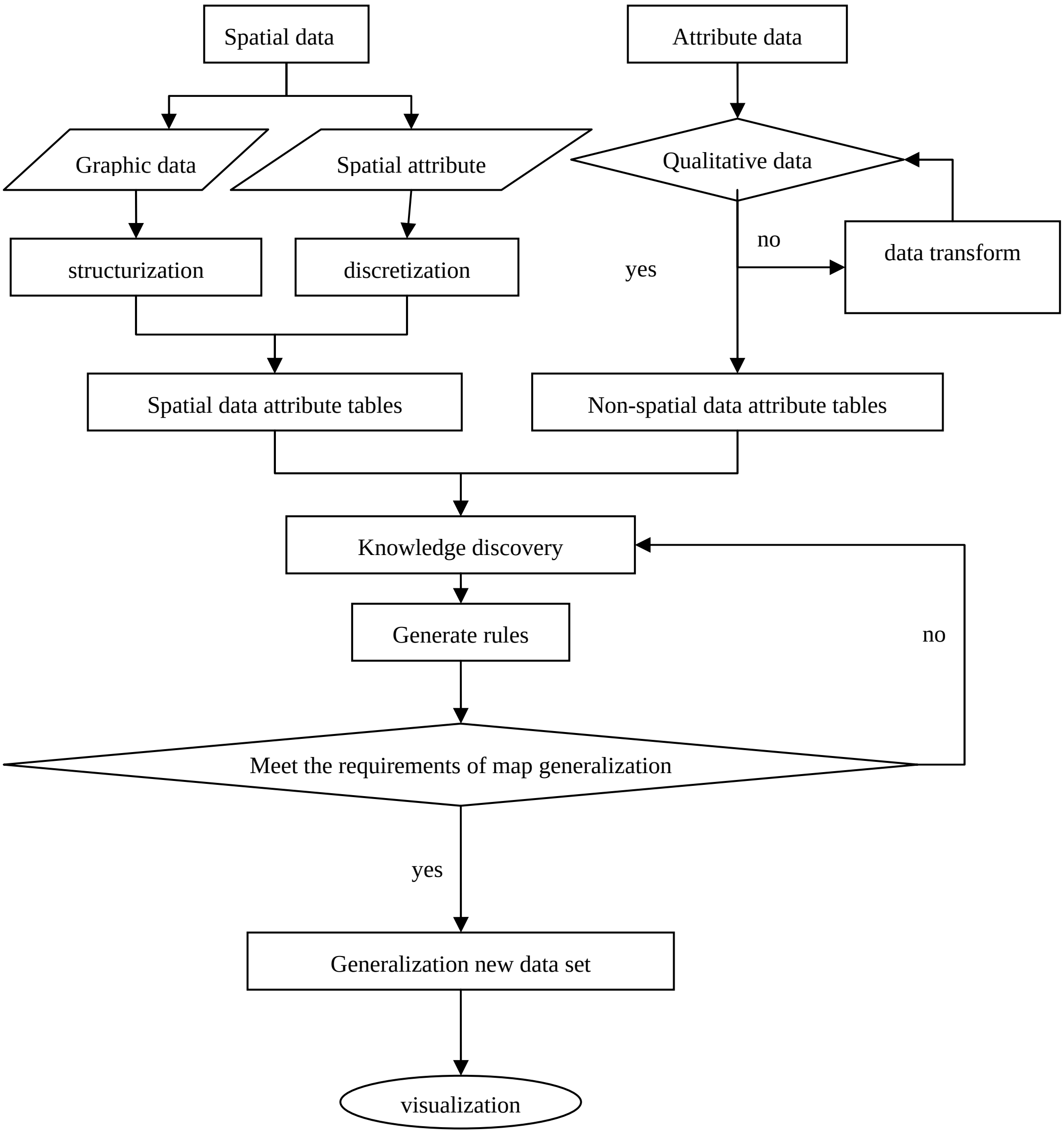

Fig.2 model of the map generalization based on rough set

For point objects, structurize point cluster through convex hull algorithm or Voronoi diagram, and thus integrate structural information and thematic attribute information of points to create an attribute decision table, see Table 5.

Table 5 contains the location information of points, the controlled area in Voronoi diagram, the semantic information and other spatial structure information as well as thematic attribute information. Also, the decision attributes can include in both the spatial information and attribute information.

Tab.5 Attribution-decision making table of points

Point cluster	Spatial infomation					Attribute information	
set	Attribute c_1	Attribute c_2	Attribute c_3	Attribute c_4	Attribute c_5	Attribute c_6	Attribute c_7
U	Voronoi area	Adjacency density of points	Convex hull level	Special Position Description	Other spatial features	Semantic information	Other thematic attributes
Point v_i							

Attribute decision table like this can be created for line objects either, see Table 6.

Tab.6 Attribution-decision making table of single line

line cluster	Spatial information					Attribute information
set	Attribute c_1	Attribute c_2	Attribute c_3	Attribute c_4	Attribute c_5	Attribute c_6
U	Line level	Length of line	semantic information	Special Position Description	Other spatial features	Other thematic attributes
Line v_i						

From table 5 and table 6 we see, through the structured processing of cartographic information, the structural information of spatial objects can be added to the attribute decision table as a data item to realize the generalization of spatial data and attribute data. Apart from the condition attributes listed in the tables, other attribute data may be selected according to different applied themes, so as to meet the needs of different applications. During the analysis and reduction of attribute decision table, which to be designated as condition attributes and which to be designated as decision attributes should be determined selectively, thus satisfying the needs of different thematic applications.

6. Conclusion

The decision rule reduction of rough set is a process of obtaining knowledge from data and a process of spatial data mining with self-adapting effect. This process of knowledge generation can reduce human intervention as possible and preserve the original features of objects. Rough set can not only generalize existing data in map database, but also determine which data are in need of being updated through decision rules. With respect to the calculation form, map database contains a large amount of relational data tables, which can be seen as attribute decision tables in the rough set. This provides convenience for the application of rough set method. As regards its efficiency, by the reduction of attributes and attribute values during knowledge discovery, rough set method can eliminate superfluous attribute information, raise the efficiency, and lower the error rate. This is very important for information systems with mass data, such as cartography.

Only in recent twenty years, the rough set theory is much accounted of and applied in the data decision analysis, machine learning and pattern recognition of computer. However, it has been introduced into the fields of cartography and GIS not for long, so it is a new topic to apply theories and methods of rough set into the field of map generalization, especially integrating spatial information and attribute information. How to make use of spatial data and attribute data of map generalization better for rough set and how to use rough set to meet the special requirements in the map analysis and processing still need further researches.

Acknowledgment

This work is supported by the National Science Foundation of China under Grant No.40701158 and No.40671154.

References

1. Pawlak, Z., *Some Issues on Rough Sets.* Transactions on Rough Sets, 2004. **1**: p. 1-58.
2. W. Li, *Research on Model and Methods of GIS Information Generalization Based on Rough sets,* doctoral dissertation, 2006, Wuhan University.
3. C. Di, *Spatial Data Mining and Knowledge Discovery.* 2001, Wuhan , Wuhan University Press.
4. G. Wang, *Rough Set Theory and Knowledge Acquisition***.** 2001, Xi'an: Xi'an Jiao tong University Press.
5. Roman Slowinski, D.V., *A Generalized Definition of Rough Approximations Based on Similarity.* IEEE Transactions on Data and Knowledge Engineering, 2000. 12(2): p. 331-336.
6. Inuiguchi, M., *Generalization of Rough Sets and Rule Extraction*. Transactions on Rough Sets, 2004. 1: p. 96-119
7. Pawlak, Z., *Rough Sets: Theoretical Aspects of Reasoning about Data.* 1991: Kluwer Academic Publishers.
8. H. Wu, *Structured approach to implementing automatic cartographic generalization.* Journal of Wuhan Technical University of Surveying and Mapping.

Vol 21(3), p. 277 - 295, 1996.

9. W. Zhang, G. Chou, *Uncertainty decision-making based on rough set.* 2005, Beijing: Tsinghua University Press.
10. Li, Z.,S. Openshaw. *A natural principle for the objective generlization of digital maps.* Cartography and Geographic Information System, 1993, 20(1): p. 19-29.
11. Jing, B.,C.Claramunt. *A Structural Approach to the Model Generalization of an Urban Street Network.* GeoInformatica, 2004, 8(2): p. 157-171.
12. CHEN, J., C. LI, Z. LI,C. GOLD. *A Voronoi-based 9-intersection model for spatial relations.* int. j. geographical information science, 2001, 15(3): p. 201-220.