

Using Rough set Represent the Uncertainty in GIS Spatial Data

Jin ZHANG

Dep. Surveying and Mapping of Taiyuan University of Technology 030024

E-mail: zjgis@sina.com

Abstract: To spatial object operations, it is very important to ascertain the future situation of spatial objects, such as , in map generalization, multiresolution presentation and spatial object classification and aggregation. Changing the situation of spatial objects , here situation of spatial object means the position, attribute and shape, will induce the uncertainty in spatial presentation. Here we use rough set handle these problems. A rough set is an extension of the standard mathematical set. In this extension, an uncertain set is represented by its upper and lower approximation. If the data point is in the lower approximation, we are sure that it is in the set. If it is not in the upper approximation, we are sure that it is not in the set.

Key words: Rough set uncertainty GIS multiresolution

1. Introduction

In recent years, representing uncertainty in spatial data has become more and more of a concerning. Since rising numbers of decisions are based on (information obtained from) spatial data and user confidence in this often computer-processed data is usually very high. It is getting increasingly important to specify how large the uncertainty in this data is, and, consequently, how large the uncertainty in the information obtained from this data is.

This problem has often been approached with fuzzy sets. A fuzzy set is an extension of the standard mathematical set idea, where each data point has an associated membership value, which expresses the likelihood of membership of the data point. This mapping from data points to likelihood is called the membership function. If this membership function is not obvious, it can be very hard to determine. In those cases, [Ola Ahlqvist,98 et al] anticipate that a rough set based approach is more appropriate, since there will be no need to determine a membership function, or even resort to an arbitrary one.

A rough set is also an extension of the standard mathematical set idea. [Pawlak,94] has succinctly described: In the rough set theory each vague concept is replaced by a pair of precise concepts called its lower and upper approximations; the lower approximation of a concept consists of all objects which surely belong to the concepts, whereas the upper approximation of concept consists of all objects which possibly belong to the concept. In this extension, namely, an uncertain set is represented by its upper and lower approximation. If the data point is in the lower approximation, we

are sure that it is in the set. If it is not in the upper approximation, we are sure that it is not in the set.

This paper introduces the rough set and how to calculate the upper and lower approximation. On this basis, we study the rough classification for GIS data. A basic model calculating the uncertainty of map generalization is also given by author.

2 Rough Sets

Formally, let U be a set; let R be a set of equivalence relations imposed on the universe U . The knowledge base K is defined as $K=(U,R)$; and the concept X in K is defined as a subset of U , i.e., $X \subseteq U$. Further, The upper approximation of the concept X in K under a given equivalence relation R is defined as :

$$X^- = \{x: [x]_R \cap X \neq \emptyset\}$$

And the lower approximation of the concept X is defined as

$$X_+ = \{x: [x]_R \subseteq X\}$$

Here $[x]_R$ represents the equivalence class of x under the given relation R . If $X_+ = X^-$, then the concept X is considered to be precise; otherwise, we say the concept is vague. For a vague concept, these two approximations under the given relation in the knowledge base K can be calculated and obtained.

A rough set is a pair (X_+, X^-) of standard sets, the lower and upper approximation. X_+ , the lower approximation, is always a subset of X^- , the upper approximation. The meaning of these two sets is that if a data point lies in X_+ , we are sure that the point is in the rough set, if a data point lies in $X^- - X_+$, we are unsure whether or not the point is in the rough set, and if a data point is outside X^- , we are sure that the point is not in the rough set. These sets can contain either individual points, or continuous areas; we will use the term 'area' below. We will often call $X^- - X_+$ the area of uncertainty of a rough set. As opposed to rough sets, standard sets are often called crisp, a term that also applies to a rough set with an empty area of uncertainty. Conversely, a rough set with an empty lower approximation and a non-empty area of uncertainty can be called completely *rough*.

Here is a example for calculating the X_+ and X^- :

$$\text{Let } U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$$

We define the equivalence relation $R = \{Y_1, Y_2, Y_3, Y_4, Y_5\}$

$$Y_1 = \{x_1, x_5\} \quad Y_2 = \{x_2\} \quad Y_3 = \{x_3, x_4\} \quad Y_4 = \{x_6\} \quad Y_5 = \{x_7, x_8\}$$

$$\text{If We define one classification as: } X_1 = \{x_1, x_3, x_7\} \quad X_2 = \{x_2, x_4\}$$

? Calculating X_{1+}, X_{1-} and X_{2+}, X_{2-}

$$\text{As } Y_1, Y_2, Y_3, Y_4, Y_5 \not\subseteq X_1 \quad \text{hence} \quad X_{1+} = \emptyset$$

$$\text{Also } Y_1 \cap X_1 \neq \emptyset \quad Y_2 \cap X_1 \neq \emptyset \quad Y_3 \cap X_1 \neq \emptyset \quad Y_4 \cap X_1 = \emptyset \quad Y_5 \cap X_1 = \emptyset$$

So $X_1^- = Y_1 \cup Y_3 \cup Y_5 = \{x_1, x_5, x_3, x_4, x_7, x_8\}$

Same reasons, we also get $X_2^- = \{x_2\}$ $X_2^- = \{x_2, x_3, x_4\}$

3 Rough Classification

[Ola Ahlgvist,98] described rough classification and also expressed two fundamental types of uncertainty:

- Uncertainty of spatial location: If a class has an upper approximation that is larger than its lower approximation (which may even be empty), uncertainty about the spatial location of that class has been expressed.
- Uncertainty of attribute value: If a certain area is assigned to the upper approximation of more than one class, it is no longer certain to which class that area belongs. Thus, uncertainty of attribute value has been expressed.

In that paper [Ola Ahlgvist, 98], the author also gives the error matrix extension. One of the classifications, call it A, has each of its classes associated with a column in the matrix. The other classification, B, has each of its classes associated with a row in the matrix. Each of the elements of the matrix then contains the area of the intersection of the two corresponding classes. This matrix has three properties: the row-sum, column-sum and total-sum properties: The sum of all the elements in a row is exactly the area of the class from classification B associated with that row. The sum of all the elements in a column is exactly the area of the class from classification A associated with that column. The sum of all the elements in the matrix is exactly the total area covered by the whole classification. Congalton (1991) also describes the various accuracy measures that can be computed from such a matrix.

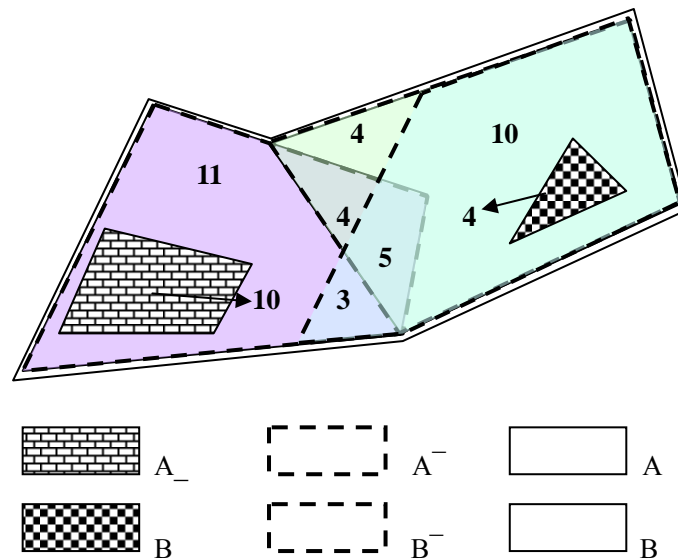


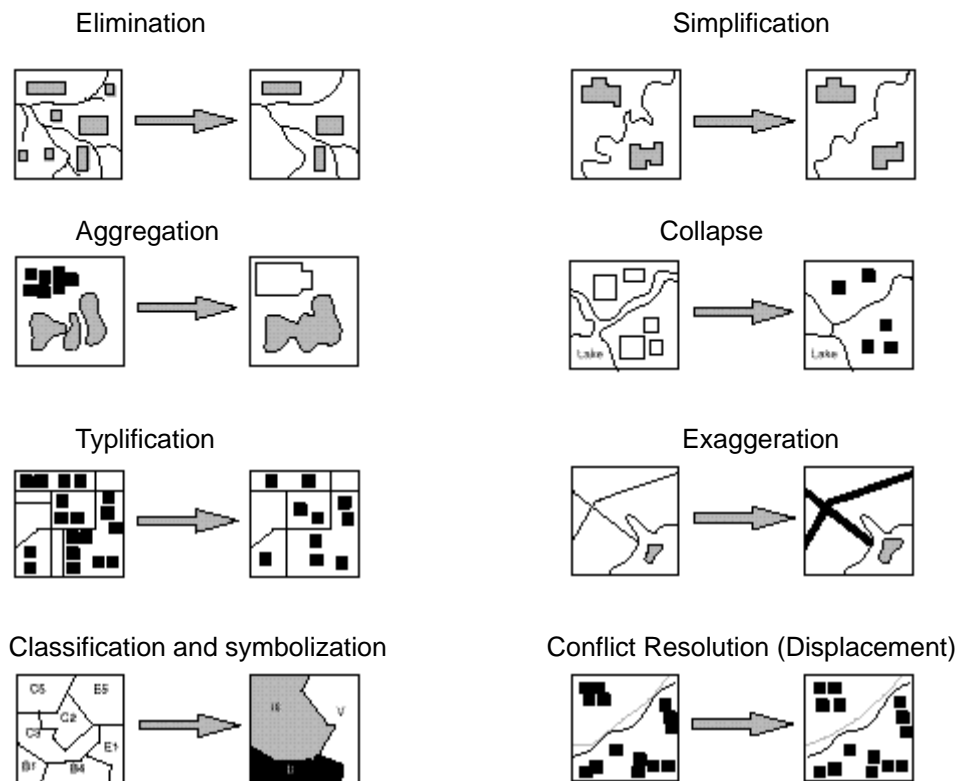
Table1 an extend error matrix

	A ₋	A ₋ -A ₋	B ₋	B ₋ -B ₋	Areas
A	10	15	0	8	29
B	0	8	4	15	22
Totals	10	23	4	23	

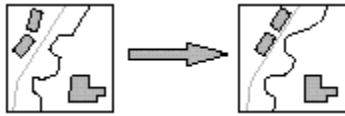
Table 1 is the extended error matrix for the situation as shown in figure 1. As we can see, however, the row-sum and total-sum properties are not equal, since some parts may be counted double, as we may be unsure to exactly which class they belong (and therefore have classified them in the uncertainty areas of more than one class). In the example matrix, the figures in the column labeled 'areas' are the actual areas of the classes, not the sums of the elements in the row, and the overall area can be obtained from the sum of that column, not from the sum of the last row.

4. Representing Map Generalization Classifications Uncertainty Using Rough Set

We often use nine arithmetic operators to generalize the map data. These nine arithmetic operators are: Elimination, Simplification, Aggregation, Collapse, Typification, Exaggeration, Classification and symbolization, Conflict Resolution (Displacement) and Refinement (In follow figures).[Barbara P. Buttenfield, 91]



Refinement



All map data (in 1:500 to 1:50000 scale) can be divided as nine main classes according to national map standards. These nine main classes are: Surveying control points, Settlements, The factory and mining building, Transport installations, Pipe lines, River, boundary, Terrain and Vegetation. We can fine divided these nine main classes into next fine classes till capturing the objects in map. Now we have the map objects, map scale and map generalization nine arithmetic operators. Also we have the rules applying nine arithmetic operators. Now we can construct set U, it is the all map objects. Also we can construct the equivalence class. Equivalence class one is scale series, other is the nine arithmetic operators rules. Based on these, the uncertainty of map generalization can be approached. Of course we must establish the operator rules, such as simplification.

Simplification refers to the process of eliminating unneeded detail from a map, and geometrical, structural, and procedural knowledge can be profitably applied to the problem. General, three main types of simplification are considered in this section: point simplification, line simplification, and feature elimination. For point and feature simplification, we can use rough set establish the uncertainty model of map generalization. To elaborate on point simplification, assume that a homogeneous set of point symbols is used to represent the distribution of individual objects on a detailed map. Simplification enables several features to be represented by a single symbol on map produced at a smaller scale. If quantitative data are used, then the weight associated with a point may be used to aid in making a decision. If the data are nominal scaled then the process becomes more purely geometrical in the sense that each objects in the class has a weight of one and decisions are made on the basis of location.

Feature elimination is the another type of simplification. For feature elimination, a decision must be made about whether to display an object, given the purpose and scale of the intended map. Two major criteria can be used to provide guidance about feature retention: geometry and attributes.

In such situation, we can construct follow rough set elements:

$U = \{\text{All points in map}\}$

$R = \{X_1, X_2, X_3 \dots\}$

$X_1 = \{U \text{ in Scale series}\}$

$X_2 = \{U \text{ in Point distributive rules}\}$

$X_3 = \{U \text{ in Geometric elimination rules}\}$

$X_4 = \{U \text{ in Feature point elimination rules}\}$

...

Based on above elements, we can calculate the X_{-} and X^{-} .

5. Conclusions and suggestions

This paper studies that applying rough set represent the uncertainty of classification and map generalization. A rough set is an extension of the standard mathematical set idea. If a class has an upper approximation that is larger than its lower approximation (which may even be empty), uncertainty about the spatial location of that class has been expressed. If a certain area is assigned to the upper approximation of more than one class, it is no longer certain to which class that area belongs. Thus, uncertainty of attribute value has been expressed. Also we suggest using rough set to establish the uncertain model about map generalization. Of course, this is only first step. Next, we will construct the practice rules set according to the map generalization rules.

6. Acknowledgements

This paper is funded by Open Research Fund Program of LIESMARS under grant No.WKL(98)0301 and Nature science Fund of Shanxi Province under the grant No.991027

7. References

- [1] Barbara P. Buttenfield, Robert B. McMaster and Herbert Freeman, Map generalization: Making rules for knowledge representation, Longman house, Burnt Mill, Harlow,1991,pp.87-102
- [2] Congalton, Russel G., A review of assessing the accuracy of classifications of remotely sensed data. Remote sensing environment, 1991(37), pp.35-46
- [3] Pawlak Z, Rough set, International Journal Computer and Information Science, 1982(11), pp.341-356
- [4] Ola Ahlqvist, Johannes Keukelaar and Arim Oukbir, Using rough classification to represent uncertainty in spatial data, 10th colloquium of the spatial information research center, University of Otago, New Zealand, Nov. 1998