Topological Relations between Random Areal Objects within GIS

Min DENG[1, 3] Wenbao LIU[2] Xiaoyong CHEN[3]
1. Department of Geomatics, Institute of Remote Sensing Information and Engineering,
Wuhan University, 229 Road Luoyu, Wuchang, Wuhan, 430079, P.R.China
dengmin208@263.net

2. Department of GI Science, Institute of Geo-informatics and Engineering,
Shandong University of Science and Technology, Taian, Shandong, 271019, P.R.China
liu_wenbao@yahoo.com

3. Geoinformatics, School of Advanced Technology, Asian Institute of Technology
P.O BOX 4, Klong Luang, Pathumthani 12120, THAILAND
xychen@ait.ac.th

ABSTRACT
There is inevitably error or uncertainty in spatial data, which are used to represent reality world, it will causes a suspicion for correctness of topological relations obtained by reasoning from observation data. In this paper, we firstly describe the geometric and statistical properties of areal object, and analyzing the effects of positional uncertainty on representation of topological relations between areal objects. Then, the concept of uncertain topological relations is defined, and the function of probability computation of a point falling inside an areal object is given. Finally, a new approach to determining random topological relations is presented.
Key words: topological relations; random object; positional error

1. Introduction

One of the most fundamental properties of spatial objects in the real world is topological relationship, which have been widely investigated in GIS in recent years (Egenhofer et al, 1991; 1995). Existing works are based on crisp sets and involve general topology such as algebraic and point-set topology. However, there are inevitably errors or uncertainties in spatial data, which are used to represent reality world (Goodchild and Gopal, 1989; Guptill and Morrison, 1995 Burrough and Frank, 1996; Shi and Liu, 2000). Those errors or uncertainties cause a suspicion for correctness of topological relations obtained by reasoning from observation data (Chen, 1996; Winter, 2000). In this paper, we study the effect of random errors on the description of topological relations between areal objects represented by spatial data with random positional error. The questions, which will be answered in this paper, are: (i) what the classifications of topological relations are; (ii) what the adaptive graphical structure for representing spatial objects with errors is; and (iii) what randomness of value of the element in 9-intersection model is. Sequentially, our focus is on how to describe areal objects with random errors, and how to define the concept of random sets. We defined the point-set interior, boundary and exterior of random areal objects. Apparently, three topological components of random areal objects are random sets, and their intersection also has two possible values, *i.e.*, ϕ and $\neg\phi$. With different values, their topological relations will be changed. In order to measure such random topological relations, a new form of the well-known 9-intersection model is developed. Further, a metric indicator, *i.e.*, relative probability, is defined. And take the topological relation with its relative probability being maximum value as the resulting answer. Finally, this paper ends with an applied example and elaborates its realization approaches to above-mentioned theoretic method under a GIS environment.

2. Representing uncertain graphic data

2.1 Changes of graphic structure

Definition 1 For any random line, all its vertex should satisfy Degree (N)$\geq$1, in which Degree (N) denotes the connective degree related with vertex N. If having Degree (N)=1, the vertex N is boundary point of random line too.

Definition 2 If there is a chain between two nodes, we call the two nodes connective. Furthermore, if all pairs of nodes in a planar graph are connective, the graph is connective.

Any graph in the plane, *G*, is composed of node, edge and face, and the numbers of these elements satisfy the following formula

$$f + n - e = c + 1 \tag{1}$$

Where *f*, *n*, *e* are the numbers of the face, node and edge, and *c* is the number of connective branches of *G*. If the planar graph is connective, *i.e.*, *c*=1, the formula (1) is simplified as

$$f + n - e = 2 \quad (2)$$

The expression is the famous Eula formula, which is often used for check of topological inconsistency. Below we apply it to analysis changes of graphic structure under the effect of uncertainty. In figure 1(a), it satisfies the expression: $f + n - e = 2$, while it doesn't satisfy in figure 1(b). This is mainly because uncertainty leads to a break of connectivity. From the view of spatial relations between spatial entities, there is a quantitative change in figure (a) and (b), and they are represented as 'Overlap' and 'Disjoint' respectively. In fact, topological relation between A1 and A2 is 'Meet'. Hence, it is very important to measure this kind of uncertainty.

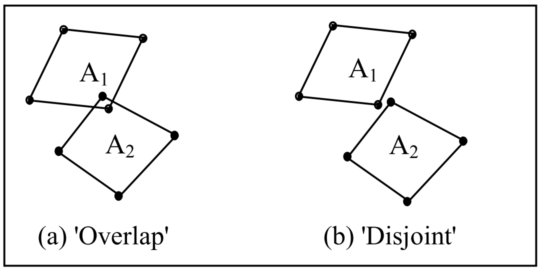

Figure 1 Effect of uncertainty on graphic structure

2.2 Modeling randomness of graphic data

For uncertainty of positional data, there is large numbers of research documents (Shi and Liu, 2000). Here, we only focus on spatial entities with definite boundary locations. In general, its positional uncertainty is major from digitalization, scanning and measuring. Therefore, we may assume that uncertainty of graphic data is random error. In spatial databases, a point, P_i, is represented as (x, y) in 2-dimensional space, a line or arc feature L_i as $P_1, P_2, \cdots, P_n$, a polygon feature, A_i, consisted of $L_1, L_2, \cdots, L_n$. So, positional uncertainty of a point will further affect correctness of line and areal features containing the point. Geometrically, any point in a line segment is represented:

$$P_m = m\, P_i + (1\text{-}m)\, P_{i+1} \quad (3)$$

Where, m is a parameter of splitting ratio, and its value equals $P_i P_m / P_i P_{i+1}$. That is to say, any point in a line segment is a linear combination of its two adjacent endpoints. Furthermore, we may represent any boundary point in areal feature by the formula (3). For simplicity, we assume that positional uncertainty of a point complies with normal distribution. In mathematical, a random point may be regarded equally as a random variable. According to the formula (3), any point in line and areal feature will also comply with normal distribution. So, a line feature will be regarded as a normal random process, an areal feature as a normal random field.

3. Describing topological relations between random areal objects

3.1 Invariant of topological relations types

Base on 9-intersection model (Egenhofer et al, 1991), topological relations between areal objects without error or uncertainty in 2-D space can be distinguished into 8 kinds. For random areal objects, though data describing their spatial position have inevitably error or uncertainty, separable topological relations are still 8 kinds based on 9-intersection model, that is, 'Disjoint', 'Meet', Overlap, 'Contains', 'Covers', 'Contained-By', 'Covered-By', 'Equal'.

3.2 Describing topological relations between random areal objects

As a whole, types of separable topological relations between random areal objects are the same. But uncertainty will be possible to change topological relations. As in figure 1, under the effect of positional uncertainty, topological relations between A_1 and A_2 are described as one of 'Overlap', 'Meet' and 'Disjoint'. Sometimes uncertainty does not any effect on description of topological relations. In figure 2, topological relations between A_1 and A_2 are both 'Overlap'.

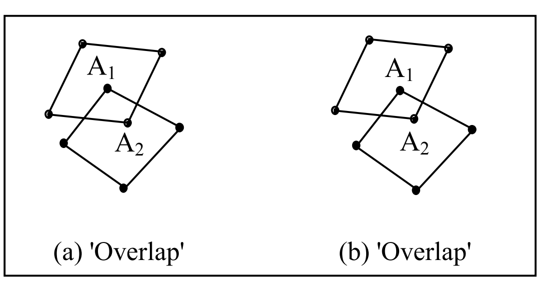

Figure 2 Effect of uncertainty on topological relations

Shi and Liu (2000) built a kind of g-band model of random line object based on stoma process. And Liu (1998) has set up g-donut model of random areal object based on stoma field in his postdoctoral report. Here, in order to describe such uncertain topological relations, we firstly define the region consisting of exterior boundary of g-donut as exterior region, denoted as g^+, the region consisting of interior boundary of g-donut as interior region, denoted as g^-. Similar to 9-intersection model by Egenhofer (1991), we define following 9-intersection models respectively,

$$\Im^+ = \begin{bmatrix} (g_1^+)^0 \cap (g_2^+)^0 & (g_1^+)^0 \cap \partial(g_2^+) & (g_1^+)^0 \cap (g_2^+)^- \\ \partial(g_1^+) \cap (g_2^+)^0 & \partial(g_1^+) \cap \partial(g_2^+) & \partial(g_1^+) \cap (g_2^+)^- \\ (g_1^+)^- \cap (g_2^+)^0 & (g_1^+)^- \cap \partial(g_2^+) & (g_1^+)^- \cap (g_2^+)^- \end{bmatrix}$$

And

$$\Im^- = \begin{bmatrix} (g_1^-)^0 \cap (g_2^-)^0 & (g_1^-)^0 \cap \partial(g_2^-) & (g_1^-)^0 \cap (g_2^-)^- \\ \partial(g_1^-) \cap (g_2^-)^0 & \partial(g_1^-) \cap \partial(g_2^-) & \partial(g_1^-) \cap (g_2^-)^- \\ (g_1^-)^- \cap (g_2^-)^0 & (g_1^-)^- \cap \partial(g_2^-) & (g_1^-)^- \cap (g_2^-)^- \end{bmatrix}$$

Where, $\Im^+$, $\Im^-$ are both a 3×3 matrix; while g_i^+, g_i^- (i=1, 2) are exterior region and interior region of g-donut of areal object A_i. Parameters $(g_i^+)^0$, $\partial(g_i^+)$ and $(g_i^+)^-$ represent the interior, boundary and exterior of g_1^+ respectively. Correspondingly, $(g_i^-)^0$, $\partial(g_i^-)$, $(g_i^-)^-$ represent the interior, boundary and exterior of g_i^- respectively.

Obviously, all possible topological cases will be included in the interval [$\Im^+$ $\Im^-$]. Here, it should be noticed that symbol [ ] is not continuous real number between $\Im^+$ and $\Im^-$ in mathematics. It only denotes all possible topological relations from $\Im^+$ to $\Im^-$, or from $\Im^-$ to $\Im^+$ under continuous positional changes of spatial objects. For example, if $\Im^+$ means 'Overlap', $\Im^-$ 'Disjoint', then all possible topological relations between A_1 and A_2 are 'Overlap', 'Meet' and 'Disjoint'. Similarly, if $\Im^+$ means 'Overlap', $\Im^-$ 'Overlap', then all possible topological relations between A_1 and A_2 are only 'Overlap', that is, positional uncertainty has not any effect on description of topological relations between A_1 and A_2.

4 Measuring uncertainty of topological relations between random areal objects

4.1 Uncertainty of taking value of elements in 9-intersection model

Change of topological relations from one case to another case is caused by changes of taking value of one or several elements in 9-intersection model. Take figure 1 as an example, all possible topological relations between A_1 and A_2 are 'Overlap', 'Meet' and 'Disjoint', and their corresponding 3×3 matrix are respectively:

1) 'Overlap' $\begin{bmatrix} \neg\phi & \neg\phi & \neg\phi \\ \neg\phi & \neg\phi & \neg\phi \\ \neg\phi & \neg\phi & \neg\phi \end{bmatrix}$

$\Downarrow$

2) 'Meet' $\begin{bmatrix} \boxed{\phi} & \boxed{\phi} & \neg\phi \\ \boxed{\phi} & \neg\phi & \neg\phi \\ \neg\phi & \neg\phi & \neg\phi \end{bmatrix}$

$\Downarrow$

3) 'Disjoint' $\begin{bmatrix} \phi & \phi & \neg\phi \\ \phi & \phi & \neg\phi \\ \neg\phi & \boxed{\neg\phi} & \neg\phi \end{bmatrix}$

Taking values of above bounded elements will change with different topological relations, which only is caused by positional uncertainty. While positional uncertainty of spatial object is often measured by computing its probability within certain spatial ranges, so it is practicable to measure uncertainty of topological relations between random areal objects.

4.2 Function of computing probability

Although we know all possible topological relations by analyzing $\Im^+$ and $\Im^-$, it is often needed to computing the probability as some topological relation t_i. In 2.2 section, we mentioned how to represent an areal object in current GIS databases. From viewpoint of set theory, an areal object will be an ordered set consisted of a series of elements. Therefore, when position of an

areal object is of uncertainty or randomness, its elements will be of uncertainty or randomness. Here, we call such a set as random set. In this paper, probability is used to represent error or uncertainty of element in a random set.
For an areal object ***A***, when a point ***s*** is lying in its boundary, then the probability of point ***s*** falling inside ***A*** is 0.5, and the probability of ***s*** not falling inside ***A*** is also 0.5.
Therefore, we define a random set ***A***,

$$p\{s \in A\} = 1 - \tfrac{1}{2}\max\{H(s,a),\ a \in A\} \tag{3}$$

Where, ***s*** is a point element, and ***A*** is a random set. $p\{s \in A\}$ is the probability of ***s*** belonging to random set A, $H(s,a)$ the probability function of determining whether ***s*** and ***a*** are the same point.
For any point $a(x_a, y_a)$ of falling in boundary of ***A***, its probability density functon is defined as

$$f(x, y) = \frac{1}{2\pi\sigma^2}\exp\left[-\frac{(x-x_a)^2+(y-y_a)^2}{2o^2}\right] \tag{4}$$

Here, parameter σ is the standard error of $a(x_a, y_a)$, so the probability that any point falls in the equal density error circle is

$$\begin{aligned} p(d) &= \iint_{C_d}\frac{1}{2\pi\sigma^2}\exp\left[-\frac{(x-x_a)^2+(y-y_b)^2}{2\sigma^2}\right]dxdy \\ &= \int_0^d \frac{\rho}{\sigma^2}\exp\left(-\frac{\rho}{2\sigma^2}\right)d\rho \\ &= 1-\exp\left(-\frac{d^2}{2\sigma^2}\right) \end{aligned} \tag{5}$$

Where, C_d: $(x-x_a)^2+(y-y_b)^2 \le d^2$, *d* the radiu of error circle, So we define $H(s,a)$ as

$$\begin{aligned} H(s,a) &= 1 - p(d) \\ &= \exp\left(-\frac{d^2}{2\sigma^2}\right) \end{aligned} \tag{6}$$

here, *d* is the distance between the point elements $s(x_s, y_s)$ and $a(x_a, y_a)$. So, the probability function of computing any point ***s*** belonging to some areal object ***A*** is

$$p\{s \in A\} = 1 - \tfrac{1}{2}\exp\left(-\frac{d^2}{2\sigma^2}\right) \tag{7}$$

4.3 Determining uncertain topological relations

Because there is random error in spatial data, the topological relation ($\hat{t}$) derived by spatial data in GIS sometimes does not coincide with real topological relation (t). But sometimes it is difficult to determine the real topological relation, so the uncertain topological relations can be expressed with confidence probability of t as $\hat{t}$. In addition, we may obtain all possible topological relations by analyzing $\Im^+$ and $\Im^-$, noted as $\hat{t}_1, \hat{t}_2, \cdots, \hat{t}_n$. Hence, we define uncertain topological relations as follows:

$$p(t \mid \hat{t}_i) = \begin{cases} p_{i1}, & t = \hat{t}_1 \\ p_{i2}, & t = \hat{t}_2 \\ \vdots & \\ p_{in}, & t = \hat{t}_n \end{cases} \tag{8}$$

Where, $p_{ij}(1 \le j \le n)$ is the probability when $t = \hat{t}_j$, while $\hat{t}_j$ is a relation existed with some probability $p_{ij}(p_{ij} > 0)$. Equivalently, formula (17) is expressed as

$$t_i = p_{i1}/\hat{t}_1 + p_{i2}/\hat{t}_2 + \cdots + p_{in}/\hat{t}_n \tag{9}$$

Because of the effect of random error or uncertainty, real topological relations will be one of all possible topological relations. Here, we present concept of relative probability as an indicator to determine real topological relations. Let topological relation between random areal objects described in GIS be $\hat{t}_i$, its real topological relations *t* may take one of $\{\hat{t}_j, 1 \le j \le n\}$. Here, let

$$\hat{t}_i = \{\alpha_{11}^i, \alpha_{12}^i, \alpha_{13}^i, \alpha_{21}^i, \alpha_{22}^i, \alpha_{23}^i, \alpha_{31}^i, \alpha_{32}^i, \alpha_{33}^i\}$$
$$\hat{t}_j = \{\alpha_{11}^j, \alpha_{12}^j, \alpha_{13}^j, \alpha_{21}^j, \alpha_{22}^j, \alpha_{23}^j, \alpha_{31}^j, \alpha_{32}^j, \alpha_{33}^j\}$$

Where, α_{kl}^i $\alpha_{kl}^j (1 \le k, l \le 3)$ are vales of elements in the 9-intersection model, and the index *i*, *j* are separately row and column of corresponding 9-intersection model. Then, we define

$$p_{ij} = \frac{\prod_{k,l=1}^{k,l=3} p(\alpha_{kl}^{j})}{\prod_{k,l=1}^{k,l=3} p(\alpha_{kl}^{i})} \tag{10}$$

Here, p_{ij} is called relative probability of topological relation. It satisfies the following properties: (1) $0 < p_{ij} < +\infty$; (2) $p_{ij} = (p_{ji})^{-1}$; (3) $p_{ij} = 1$ if and only if $j = i$.

Thus, we can compute all relative probabilities in formula (9) by means of (10). Furthermore, let

$$p_{ic} = \max\{p_{i1}, p_{i2}, \cdots, p_{in}\} \tag{11}$$

Therefore, We can determine that the real topological relation between random areal objects is $\hat{t}_c$.

5. Conclusions and outlooks

We analyze the effect of uncertainty on classification and description of topological relations in detail, presenting the concept of relative probability as an indicator to measure and determine uncertain topological relation between areal objects. It is very important to make a decision and analysis. Further work is to focus on how to represent and measure uncertainty of spatial query and spatial analysis caused by uncertain spatial relations.

References

[1] Burrough PA and Frank AU (1996) Geographic objects with indeterminate boundaries. Taylor and Francis, London.

[2] Chen X (1996) Spatial Relation between Uncertain Sets. *International Archives of Photogrammetry and Remote Sensing*, XVIII congress. Vienna, pp.105-110.

[3] Goodchild MF and Gopal S (1989) Accuracy of spatial databases. London: Tayloy & Francis, pp. 107-113

[4] Guptill SC and Morrison JL (1995) Elements of Spatial Data Quality. Oxford: Elsevier Scientific.

[5] Egenhofer M and Franzosa R (1991) Point-set topological spatial relations. *International Journal of Geographical Information Systems*, Vol. 5(2), pp.161-174

[6] Egenhofer M and Herring J (1991) *Categorizing binary topological relationships between regions, lines and points in geographic databases*. In: A Framework for the Definition of Topological Relationships and an Approach to Spatial Reasoning within this Framework, edited by Egenhofer M, Herring J, et al. Santa Barbara, CA, pp.1-28.

[7] Egenhofer M and Mark D (1995) Modeling conceptual neighborhoods of topological line-region relations. *International Journal of Geographical Information Systems*, Vol. 9 (5), pp.555-565

[8] Shi W and Liu W (2000) A Stochastic Process-Based Model for Positional Error of Line Segments in GIS. *International Journal of Geographic Information Sciences,* Vol.14 (1), pp.51-66.

[9] Winter S (2000) Uncertain topological relations between imprecision regions*. International Journal of Geographical Information Systems,* Vol.14 (5), pp.411- 430.