RANDOM TOPOLOGICAL RELATIONS BETWEEN LINE AND AREA IN GIS

Deng Min [a*] **Li Chengming** [b] **Chen Xiaoyong** [c]

[a] Department of Geomatics, Institute of Remote Sensing Information and Engineering,
Wuhan University, 430079, No.229, Road Luoyu, District WuChang, Wuhan, P.R.China
dengmin208@263.net

[b] Chinese Academy of Surveying and Mapping, 100039,
No.16, Road Beitaiping, District Haidian, Beijing, P.R.China
cmli@casm.ac.cn

[c] Geoinformatics, School of Advanced Technology, Asian Institute of Technology
PO BOX 4, Klong Luang, Pathumthani 12120, THAILAND
xychen@ait.ac.th

KEY WORDS: Data quality, Topological relations, Uncertainty

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 observation data. In this paper, we firstly describe the geometric and statistical properties of random line and area, and analyzing the effect of error in spatial data on representation of topological relations between spatial entities. Then, we present the concept of uncertain topological relations sets in order to represent random topological relations between spatial data with random errors. Finally, a new approach to determining random topological relations is presented.

1. INTRODUCTION

There is inevitably error or uncertainty in spatial data, which are used to represent reality world [Blakemore, 1984; Chrisman, 1982; Guptill, 1995], it causes a suspicion for correctness of topological relations obtained by reasoning observation data [Winter, 2000; Worboys, 1995]. While most of spatial queries in GIS applications are based on topological relations between two objects, so it is obvious that a wrong answer, even a mistaken decision will be made if description of topological relations is not correct. Topological relation, as one of the most fundamental property of spatial objects, has been investigated in GIS recent years. Most of researches are based on crisp set and involve common topology such as algebraic and point-set topology [Egenhofer and Franzosa, 1991; Egenhofer and Herring, 1991; Egenhofer and Mark, 1994]. How to formalize the topological relations between two random objects is still an open question.

In this paper, we consider random error of spatial data, and assume that positional error of a node conform to a standard normal distribution, then to derive and compute positional error of any point on a line segment is based on stochastic process [Shi and Liu, 2000]. Furthermore, it is well-known that an areal object is composed of a sequence of ordered line segments, here, an areal object is regarded as a random field, and positional error of any point on its boundary line can be derived similarly. On the basis of this, we will analyze the effect of random error on description of topological relations between line and areal object in following three aspects: ① classes of topological relations; ② changes of graphical structure; ③ Randomness of value of element in 9-intersection model. Sequentially, we focus on how to describe random topological relation, defining the concept of uncertain topological relations sets. Finally, this paper ends with an approach of determining uncertain topological relations by means of relative probability as an indicator.

2. GEOMETRIC DESCRIPTIONS OF RANDOM LINE AND AREA

2.1 Definition of Random Line and Area

A mass of line entity existed in reality world such as road, river, administrative boundary and so on, are all expressed with line entity in GIS. These entities are composed of a sequence of ordered line segment, which is defined by two endpoints. The error of point coordinates acquired by measuring, scanning, digitalization etc., will lead to location uncertainty of line entity. Here, it is called random line. Let L_{12} a random line segment in 2-D space, is denoted by its two end points Z_1 and Z_2. Let $Z_1=(x_1,y_1)$, $Z_2=(x_2,y_2)$, defining following operation rule: 1) $tZ_i=(tx_i,ty_i)$ 2) $Z_1 \pm Z_2=(x_1 \pm x_2,\, y_1 \pm y_2)$, then random line segment can be expressed as

$$L_{12}=\{Z_1+t(Z_2-Z_1),\ 0 \le t \le 1\} \tag{1}$$

Obviously, L_{12} is a connective close set, or denoted as a close interval $[Z_1, Z_2]$. Furthermore, the line segment not including the end points is denoted as an open interval (Z_1, Z_2). So the random line element $L_{1,2,\cdots,n}$ composed of n random points $\{Z_i, 1 \le i \le n\}$, are expressed as a union set, *i.e.*,

$$L_{1,2,\cdots,n}=\cup_{i=1}^{n-1}[Z_i, Z_{i+1}] \tag{2}$$

any integer $i\,(1 \le i \le n-1)$, the relation between random line element and its vertex satisfy the following formula

$$L_{1,2,\cdots,i} \cap (Z_i, Z_{i+1})=\phi \tag{3}$$

Here, if random line point $Z_1 \ne Z_n$, then $L_{1,2,\cdots,n}$ is called simple random line. If having $Z_1 = Z_n$, it is a close polygon in geometry. Having $L_{1,2,\cdots,i} \cap (Z_i, Z_{i+1}) \ne \phi$, then $L_{1,2,\cdots,n}$ is a self- intersection random line. In this paper, we only discuss simple random line and area.

2.2 Geometric Constraints of Random Line

For any digital line stored in GIS after data procession and compression procedures, the data used to express the location of line can be assumed to satisfy the following conditions:

1) The distance between any two points is far larger than upper limit of the maximum error, namely,

$$Z_iZ_j >> \delta \quad (1 \le i, j \le n-1)$$

2) The distance of any point to any segment, *d*, is far larger than the maximum error up limit, namely,

$$d(Z_i, Z_{i-1}Z_{i+1}) >> \delta$$

2.3 Statistical Description of Random Line

We supposed that: (1) the coordinates vectors $Z_i=(X_i\ Y_i)^T$ and $Z_{i+1}=(X_{i+1}\ Y_{i+1})^T$ are independent and the same accuracy; (2) random point Z_i ($i=1,2,\cdots,n-1$) complies with the two dimensional normal distribution

$$Z_i \sim N_2(\mu_{Z_i}, \textstyle\sum_{Z_iZ_i})\,(i=1,2,\cdots,n-1) \tag{4}$$

Here, $\mu_{Z_i}=(\mu_{X_i}\ \mu_{Y_i})^T$ is mathematical expectation of any random point coordinate vector $Z_i=(X_i\ Y_i)^T$ ($i=0,1$) and $\sum_{Z_iZ_i}$ is its covariance matrix, having

$$\sum\nolimits_{Z_iZ_i}=\begin{bmatrix} \sigma_{X_i}{}^2 & 0 \\ 0 & \sigma_{Y_i}{}^2 \end{bmatrix} \ (i=0,1) \tag{5}$$

For any point in random line, its coordinate components can be expressed as

$$\begin{cases} X_i(t_i)=(1-t_i)X_i+t_iX_{i+1} \\ Y_i(t_i)=(1-t_i)Y_i+t_iY_{i+1} \end{cases} \ (0 \le t_i \le 1) \tag{6}$$

It means that $X_i(t_i)$ and $Y_i(t_i)$ is a linear combination of normal random variables X_i, X_{i+1} Y_i and Y_{i+1} separately, they hereby satisfy the normal distribution too. Hence, the random line is composed of infinitely normal random variables $X(t)$ and $Y(t)$, where $\{t_i, 1 \le i \le n-1\}$ and $t_i \in [0,1]$.

3. IMPACT OF UNCERTAINTY ON TOPOLOGICAL RELATION BETWEEN LINE AND AREA

3.1 Invariant of Topological Relations Types

Base on 9-intersection model, topological relations between line and area without error or uncertainty in 2-D space can be distinguished into 19 types. For random line and random area, though data describing their spatial position have inevitably error or uncertainty, the topological relation types separable is still 19 kinds based on the 9-intersection model.

3.2 Change of Graphic Structure

Definition 1. For any random line, all its vertex should satisfy Degree(N) ≥ 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

$$f+n-e=c+1 \tag{7}$$

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 (7) is simplified as

$$f+n-e=2 \tag{8}$$

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, because of the positional error, the graphic structure composed of the random line *L* and random area A such as (a) and (b) is different. In figure 1(a), only numbers of face, node and edge are differs, while keeping the connectivity. But in figure 1(b), not only the numbers of graphic components differ, but the connectivity also changes.

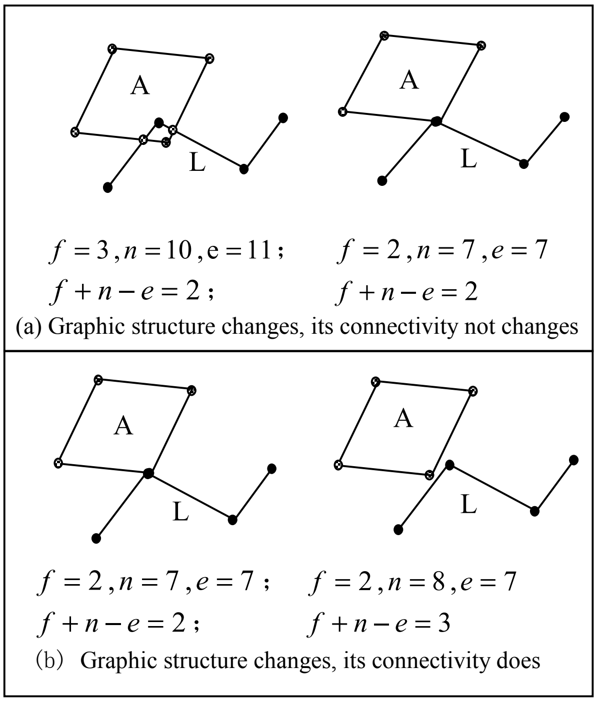

Figure 1 The effect of uncertainty on graphic structure

3.3 Uncertainty of Value of Element in 9-intersection Model

Topological relations between random line (L) and random area (A) are described based on the 9-intersection model, which is defined by intersection of L's interior (L^o), boundary (∂L) and exterior (L^-) and A's interior (A^0), boundary (∂A) and exterior (A^-), *i.e.*,

$$\Im_9(L,A)=\begin{bmatrix} L^0\cap A^0 & L^0\cap\partial A & L^0\cap A^- \\ \partial L\cap A^0 & \partial A\cap\partial A & \partial A\cap A^- \\ L^-\cap A^0 & L^-\cap\partial A & L^-\cap A^- \end{bmatrix} \quad (9)$$

According to point-set topology, the topological components of a random line L_i are defined as

$$\partial L=\{b,e\};\ L^0=L/\{b,e\};L^-=R^2-L \quad (10)$$

Their geometric meanings are illustrated in figure 2.

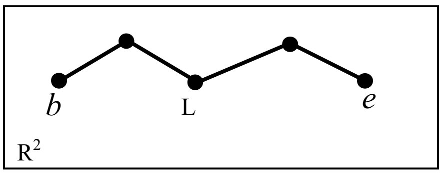

Figure 2 Topological definition of 2-D random line

We may see from the figure, that the two endpoints of random line L are its topological boundary, and the other vertex and all edges are its interior. But the measuring points described in GIS are disaccording with their real location, that is, ∂L_i, L_i^0 and L_i^- in formula (10) contain random error.Here, we define them random set. In this paper, probability is used to represent error or uncertainty of element in random set. The followings is its definition

$$p\{s\in A\}=\max\{H(s,a),\,a\in A\} \quad (11)$$

Where, s is a point element, A a random set, $p\{s\in \mathrm{A}\}$ is the probability of s belonging to random set A, $H(s,a)$ is probability function of determining whether s and a are the same point. Furthermore, the probability that s does not belong to random set A is expressed as

$$p\{s\notin A\}=1-\max\{H(s,a),\,a\in A\} \quad (12)$$

For any point $a(x_a,y_a)$ in random line L, its probability density functon is defined as

$$f(x,y)=\tfrac{1}{2\pi\sigma^2}\exp[-\tfrac{(x-x_a)^2+(y-y_a)^2}{2o^2}] \quad (13)$$

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

$$\begin{aligned} p(d) &= \iint_{C_d}\tfrac{1}{2\pi\sigma^2}\exp[-\tfrac{(x-x_a)^2+(y-y_b)^2}{2\sigma^2}]dxdy \\ &= \int_0^d \tfrac{\rho}{\sigma^2}\exp(-\tfrac{\rho}{2\sigma^2})d\rho \\ &= 1-\exp(-\tfrac{d^2}{2\sigma^2}) \end{aligned} \quad (14)$$

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

$$\begin{aligned} H(s,a) &= 1-p(d) \\ &= \exp(-\tfrac{d^2}{2\sigma^2}) \end{aligned} \quad (15)$$

here, d is the distance between the point elements $s(x_s,y_s)$ and $a(x_a,y_a)$

So, If uncertainty will be considered, then values of elements in the 9-intersection model are extended as $\phi/p_{ij}(\phi)$ or $\neg\phi/p_{ij}(\neg\phi)$, and satisfy the expression

$$p_{ij}(\phi)=1-p_{ij}(\neg\phi) \quad (16)$$

Where, $i,j(1\le i,j\le 3)$ is the number of row and collum in the 9-intersection model.

4. DETERMINING TOPOLOGICAL RELATIONS BETWEEN LINE AND AREA

4.1 Uncertain Topological Relation Set

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 to $\hat{t}$.In addition, topological relations based on the 9-intersection model is a definite classification. So the real topological relation t will only is in the neighbourhood of $\hat{t}$. Here, we define uncertain topological relations set as

$$p(t \mid \hat{t}_i) = \begin{cases} p_{i1}, & t = \hat{t}_1 \\ p_{i2}, & t = \hat{t}_2 \\ \vdots & \\ p_{in}, & t = \hat{t}_n \end{cases} \tag{17}$$

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

$$t_i = p_{i1}/\hat{t}_1 + p_{i2}/\hat{t}_2 + \cdots + p_{in}/\hat{t}_n \tag{18}$$

4.2 Description of Uncertain Topological Relations

As above mentioned, value of element in the 9-intersection model sometimes is uncertain, while the probability of taking some value is certain. Therefore, we can extend the 9-intersection model to describe such uncertain topological relations, *i.e.*,

$$\Im_9(L,A) = \begin{bmatrix} p_{11}(L^0 \cap A^0) & p_{12}(L^0 \cap \partial A) & p_{13}(L^0 \cap A^-) \\ p_{21}(\partial L \cap A^0) & p_{22}(\partial L \cap \partial A) & p_{23}(\partial L \cap A^-) \\ p_{31}(L^- \cap A^0) & p_{32}(L^- \cap \partial A) & p_{33}(L^- \cap A^-) \end{bmatrix} \tag{19}$$

As showed in figure 1(a), the two kinds of topological relations between *L* and *A* are described as

a1) Case 1

$$\begin{array}{c|ccc} & A^0 & \partial A & A^- \\ \hline L^0 & \boxed{\neg\phi} & \neg\phi & \neg\phi \\ \partial L & \phi & \phi & \neg\phi \\ L^- & \neg\phi & \neg\phi & \neg\phi \end{array}$$

a2) Case 2

$$\begin{array}{c|ccc} & A^0 & \partial A & A^- \\ \hline L^0 & \boxed{\phi} & \neg\phi & \neg\phi \\ \partial L & \phi & \phi & \neg\phi \\ L^- & \neg\phi & \neg\phi & \neg\phi \end{array}$$

Obviously, it is the change of taking value of $L^0 \cap A^0$ that causes the change of topological relation between *L* and *A* in Fig1 (a). For the same reason, two kinds of topological relations in Fig1 (b) can be described as:

b1) Case 1

$$\begin{array}{c|ccc} & A^0 & \partial A & A^- \\ \hline L^0 & \phi & \boxed{\neg\phi} & \neg\phi \\ \partial L & \phi & \phi & \neg\phi \\ L^- & \neg\phi & \neg\phi & \neg\phi \end{array}$$

b2) Case 2

$$\begin{array}{c|ccc} & A^0 & \partial A & A^- \\ \hline L^0 & \phi & \boxed{\phi} & \neg\phi \\ \partial L & \phi & \phi & \neg\phi \\ L^- & \neg\phi & \neg\phi & \neg\phi \end{array}$$

Similarly, cause of leading such uncertainty is the changes of taking value of $L^0 \cap \partial A$.

4.3 Probability Computation of Uncertain Topological Relations

Because of the effect of random error or uncertainty, real topological relations will be a kind 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 line and random area 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

$$p_{ij} = \frac{\prod_{k,l=1}^{k,l=3} p(\alpha_{kl}^j)}{\prod_{k,l=1}^{k,l=3} p(\alpha_{kl}^i)} \tag{20}$$

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

Here, we still take figure 1 as an example. After analyzing figure 1(a), the formula (20) can be simplified as

$$p_{ij} = \frac{p(\alpha_{11}^j)}{p(\alpha_{11}^i)} = \frac{p(L_1^0 \cap L_2^0 = \neg\phi)}{p(L_1^0 \cap L_2^0 = \phi)}$$

After analyzing figure 1(b), the formula (20) can be simplified as

$$p_{ij} = \frac{p(\alpha_{12}^j)}{p(\alpha_{12}^i)} = \frac{p(L_1^0 \cap \partial L_2 = \phi)}{p(L_1^0 \cap \partial L_2 = \neg\phi)}$$

Thus, we can compute all relative probabilities in formula (17) by means of (20). Furthermore, let

$$p_{ic} = \max\{p_{i1}, p_{i2}, \cdots, p_{in}\} \tag{21}$$

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

5. CONCLUSIONS

We analyze the effect of uncertainty on types and description of topological relations in detail, present to concept of relative probability as an indicator to measure and determine uncertain topological relation between line and area. The approach can be extended to discuss uncertain topological relations between two areal objects. Only the difference is in computation methods of probability.

REFERENCES

[1] Blakemore M, 1984. Generalization and Error in Spatial Databases. *Cartographica,* Vol.21, pp.131-139

[2] Chrisman N.R, 1982. A Theory of Cartographic Error and Its Measurement in Digital Databases. *In: Proceedings of Auto-Carto 5*. Crystal City, VA, pp.159-168.

[3] Guptill S.C and Morrison J.L, 2000. Elements of Spatial Data Quality. Oxford: Elsevier Scientific, 1995.

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

[5] Worboys M, 1995. GIS, a computing perspective. London: Taylor & Francis.

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

[7] 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.

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

[9] 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.