Automatic Creating Techniques for Topological Spatial Relationship Based on the Grid Index

Cai Shaohua[1,2] Wang Jiayao[3] Jiang Hongtao[2] Liao Ning[2]

(LREIS, Institute Of Geography, Chinese Academy of Sciences Beijing 100101)[1]

(The Surveying and Mapping Center of The General Staff ,Beijing 100088)[2]

(Zhenzhou Institute of Surveying and Mapping, Zhenzhou 450052) [3]

Abstract Based on the grid index algorithm of node matching , this paper throughly analyzed the right-turn and recursive algorithm of polygon creating, the polygon inside node getting method , and the nested polygon identifying theory. The new algorithms are validated to be efficient to the topological relationship creating by the analysis.

Key words Geographical Information System, Topological Relationship, Grid Index , Minimum Bounding Rectangle

Topological spatial relationship is a mathematical method for definitely defining the spatial data structure. The topological data structure is essential to the spatial analysis and the application of GIS(Geographic Information System). It describes the relationship among the point, line and polygon. Automatic creating topology relation for the raw vector graphic data is all the way focused by the scholars of GIS. The key problems to create topological relationship are the speed, the efficiency, and the automatic level. The flow of topology creating is the cleaning of the arcs, node matching, polygon creating, getting the inside point of the polygon, and identifying the island relationship. By analyzing the flow of topology creating, we can find the main step that specially influence the efficiency is node matching. In order to solve these problems, the author developed a new algorithm Based on the grid index. The main thought of the algorithm is the special techniques to build the grid index of feature nodes. At the same time, this paper thoroughly analyzed the right-turn and recursive algorithm for polygon creating, the polygon inside point getting method, and the island polygon identifying theory. These new algorithms are validated to be efficient by the practice.

1 The Analysis of the Topologic Relationship Creation

In the early algorithm, the topologic relationship is inputted by handwork during the map digitations and graphic edition. In some improved algorithms, the creation of topologic relationship still need the inside polygon node to be inputted at first. Those methods not only waste plenty of time, but also bring with many encoding errors and input errors. Because the manual input works have all kinds of limitations, the automatic creating method substitutes the old ones.

The process of automatic creating topologic relationship is as the follows.

1) Node matching. Building the relationship between the points and arcs.
2) Creating polygons. Building the relationship between the polygon and the arcs.
3) Getting the polygon inside point.
4) Island polygon identifying. Building the nest tree among polygons.

By the above four steps, the spatial topologic relationship are basically formed. Among the above methods, the node matching is the key factor to the effectivity and speed of topologic relationship's creation. The followed contents mainly describe the Grid Index algorithm for node matching. To the polygon creating , only the key algorithms are described.

2 The Grid Index Algorithm of Node Matching

Every arc has tow end points, which are called nodes. The node matching is to merge the nodes whose coordinate difference in the bounds to one node, and build the topologic relations that one node connects with which arcs.

Suppose there are N arcs, so the have 2N nodes. In order to get the relation between the node and the arcs, we can compared the 2N nodes with each other. If we do this, the algorithm's time complexity will arrive (2N×(2N-1). The users can not consume so much time. So we have to adopt new optimized method. By the practice, I worked out a method called Grid Index Algorithm to improve the speed and effective of node matching.

2.1 The Idea of the Node Matching Grid Index Algorithm

First, we can establish the grid index for the nodes(as figure 1), and assign a number to the arc from one to N.

The index struct for each grid is:

```
struct LineSEpt{
    Carray<CGIS_Point,CGIS_Point>m_nodePointA; //node array
    CArray<int,int>m_lineIDA;  //the array of the node related arc numbers
                                   //if the node is the start point, the arc
                                   //number is positive, else is minus
};
```

After the nodes' grid index have been created, we first decide if there are node in some grid, if it is true. Then we judge whether the nodes that in the eight, five or three adjacent grid in the limitation or not. If there are some nodes, they must be merged into one new node, the related arcs' number also be added to the length varied ID array: m_lineIDA. At the same time , the node in the limitation must be deleted from the grid index. By the process, not only the nodes are merged, but also the relations between the nodes and the arcs are created.

In the node matching , how many adjacent grids must be judged is decided by which grid the node is positioned. In figure 1, if the node is in grid (1,1), (1,9), (6,1), (6,9), to find the node in the coordinate difference limitation , we must search in its three adjacent grid.。 however, the node in row 1, 6 and column 1, 9(not include the above four grid)，search in its five adjacent grid. On other condition，search in the eight adjacent grid.

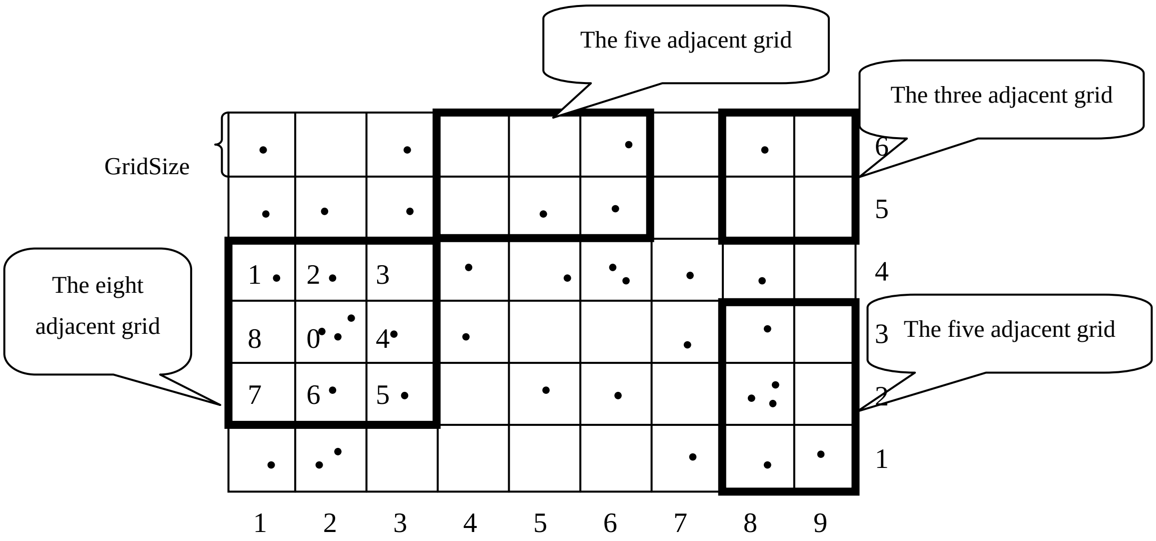

Figure 1 the nodes grid index

2.2 The Automatic Decision of Grid Size

The GridSize is decided by the number of the arc and the average length of the all arcs. Suppose there are N arcs distribute in the map bound equally., then:

GridSize = $(\text{Length}_{arc1} + \text{Length}_{arc2} + \cdots\cdots + \text{Length}_{arcN}) / N$

If GridSize>=NodeLimit(the node matching limitation)，the matched node affirm to be in the eight, five or three adjacent grids.

If GridSize<NodeLimit，then use the NodeLimit as the GridSize to build nodeindex.

2.3 Dynamic RAM Assign During Node Matching

The computer RAM is very import in the process of creating topologic relationship fro the great amount of vector data. Thinking of the limitation of RAM, the two coordinate of the arc end point are saved in the ram before the node matching. When the matched node is found in the eight ,five or three adjacent grid, the topologic structure of the node is created. At the same time, the matched node is deleted from the grid index. By this means, the utility of the RAM is improved. One or more points may be deleted from the grid index, but only one node is added in the node list.

2.4 The Computing of the Node-Arc parameters and Sorting of the Arc Direction Angle

By the node matching, the node-arc topologic relation is built. But in order to create the polygon topology, the parameters of the node-arc must be computed first, which is the arc's MBR(Minimum Bounding Rectangle), the start and end direction angle. And the direction angle of the arc that connected to the node must be sorted first.

1. The computing of parameters

Among the parameters, the MBR can be gained by compare the coordinate of the arc end points.

However, the computation of direction angle is much complex. The computation of the direction angle is the main argument in the following. In order to get rid of the case as figure 2(the two end point of the arc are the same, and the direction angles equal with each other, it is difficult to distinguish them), and ensure the correctness of the computed direction angles, .the point which is the nearest to the node must be selected to compute the direction angle(as figure 2(b)).Two angle must be computed for every arc, which is the angle from the start node to the end node, and from the end node to the start node.

The formulas to compute the direction angle are as the follows（$dx=|x_I-x_0|\neq 0$）:

$$\alpha = arctg(| y_i - y_0 | / | x_i - x_0 |) = \mathrm{arctg}(dy/dx)$$

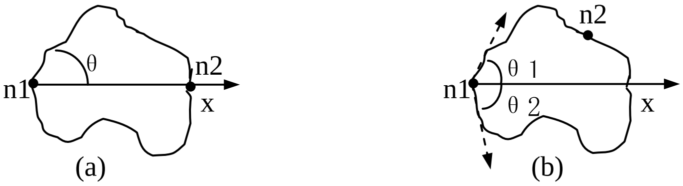

Figure 2 the computation of the arc direction angles

$$\alpha_i = \begin{cases} \alpha & (dx > 0, dy \geq 0) \\ \pi - \alpha & (dx < 0, dy > 0) \\ \pi + \alpha & (dx < 0, dy \leq 0) \\ 2\pi - \alpha & (dx > 0, dy < 0) \end{cases}$$

当 dx=0 时，$\alpha_i = \begin{cases} \dfrac{\pi}{2} & (dy > 0) \\ \dfrac{3\pi}{2} & (dy < 0) \end{cases}$

2. The Sorting of the Direction Angles

To sort the direction angle, using the node as the origin, the arc angles can be sorted from little to big. The sorted arcs can reduce the repeat computation of the arc direction angles, improve the time efficiency of polygon trace.

2.5 Sign of the Hanged Chain

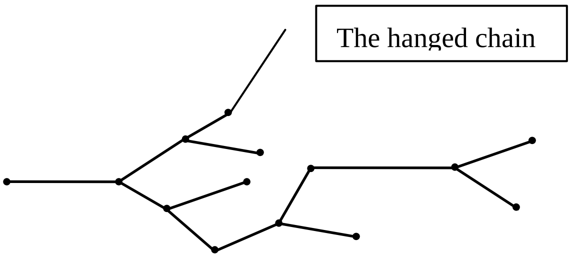

Figure 3 the sketch map of the hanged chain

The hanged chain(figure 3) is the case that whose one end point relate only to one arc. During the creating of the topologic relationship, the hanged chain must be signed to ensure the automatic tracing of the polygon. If one node has only one related arc, the arc affirm to be a hanged chain. The hanged chain can be judged by compute the node related arc number.

3 The Right-Turn Algorithm for Polygon Tracing

The basic idea of the right-turn algorithm(Huang Xinyuan, 1992) is as follows:

From one arc that composed the polygon, if the angle between the x axis and the arc is the maximum, the arc among which have relation to the same arc node , and that has the minimum direction angle , is the behind continuous arc of the polygon. If the arc direction angle is among the angles of the other arcs that related to the same arc node, then the arc that has the minimum warp angle is the polygon’s next arc.

To the regular graph, by the right-turn algorithm, the polygon can be got. But to the graph as figure 4, we must use the recursive method to get the polygon.

In figure 4(a), suppose we trace the polygon from node 1, when we trace to node 2. There are four arcs related to the node. According to the right-turn algorithm, the next arc of the polygon is b, the we return to node 2. So, we can regulate that if the next node is not the start node, but a traced node, a new polygon must be created, and assign the polygon number to the arc’s related left or right polygon number. After we form a new polygon, recur to node 2, and continue to trace polygon along arc c. When we return to node 1, end the polygon tracing.

In figure 4(b), suppose we trace from node 2 and arc a. When trace to node 3, we will continue to trace along arc c, and trace to node 9. Node 9 is a hanged node, we recur to node 7, continue the next arc 3 of node 7. But the node 8 is also a hanged node, so we recur to node 3, trace the other arc along arc c. At the end , we return to node 2, and get a new polygon.

On the polygon tracing, the end condition is very important, otherwise, there will be a dead cycle. By a great deal of practice, I conclude the end conditions as the follows.

1) If the first arc is a hanged chain, or the right and left direction of the arc has been

traced, end the trace.

2) If the start point equal to the end point of the arc, form a new polygon, and end the trace. For example, as the arc 1 in figure 4(c).

3) If there are no arcs and nodes to trace, end the trace. In figure 4(d), from node 1, trace the polygon along arc b. At the end, we will recur to node 5, and return node 2. There are node arcs and nodes to continue tracing, so end the trace.

4) If the traced node number equal to the start node number, form a new polygon, end the trace.

5) If the traced node number does not equal to the start node number, but it equals to one node in the traced node list, a new polygon formed by the middle nodes and the arcs. Then continue tracing, and return to the start node, end the trace.

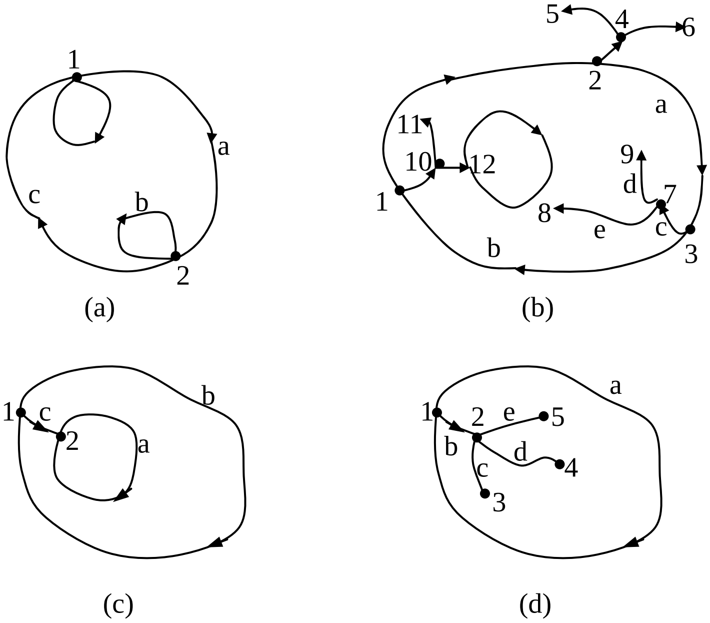

Figure 4 the unconventionality process of polygon tracing

4 Seeking the Polygon Inside point Based on the MBR

MBR is the minimum rectangle that contain the spatial the spatial object. To the vector graphic，the MBR is got by computing the minimum and maximum x and y coordinate，and represented by the lower left and upper right coordinate，that is $MBR_i=\{(X_{min}, Y_{min}), (X_{max}, Y_{max})\}$。

Based on the MBR, the principle to compute the polygon inside point is as figure 5. First, we can compare the length in the x direction and width in the y direction of the MBR. In figure 5, DX>DY，draw a line on DX/2 that upright to the X axis, so the inter-points' y coordinate with the polygon can be got. Then sort the y coordinate, the result is the y series $\{Y_1, Y_2, \cdots\cdots Y_n\}$， among them, Y_1 is the maximum，and Y_n is the minimum。And then, we get odd and even partnership of them. For example Y_1 and Y_2，Y_3 and Y_4 etc. Among them get the one that has the max distance. Suppose they are Y_i and Y_{i+1}，the polygon inside point is $((X_{min}+X_{max})/2$, $(Y_i+Y_{i+1})/2)$。

If one polygon contain island polygon, especially contain multi-islands, the inside point maybe in the island. So we must shift the covered inside point from the island. In the convention methods, the inside point must be shifted by manual work. If the method is based on the MBR, it is very convenience, and is automatic.

The method is as figure 6. First, we judge whether one polygon contains some islands or not, if it is true, then compute the MBR that contain all the island polygons, as the MBR2 in figure 6. By comparing Dx1, Dx2, Dy1, Dy2, get the maximum one. In figure 6, the maximum is Dx1. So we draw a uprightness line on the Dx1/2, and get the y series of the inter-points, and sort them. By comparing the distance between the odd and the even y coordinate, get the maximum, suppose it is Y_i and Y_{i+1}, then the polygon inside point is $((MBR1(X_{min})+MBR2(X_{min})/2$, $(Y_i+Y_{i+1})/2)$.

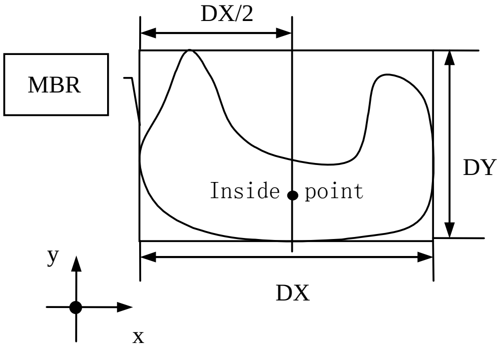

Figure 5 automatic seeking the polygon inside point

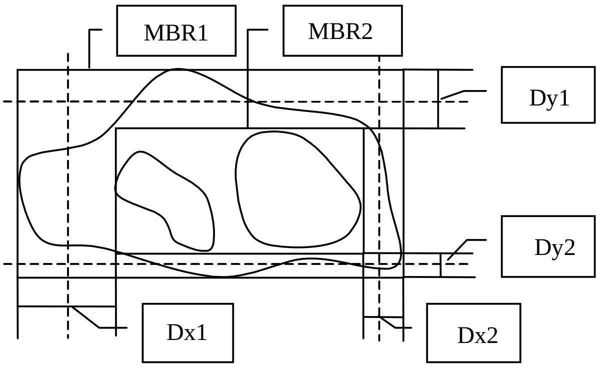

Figure 6 automatic get the nested polygon's inside point

5 Automatic Identification of the Nested Polygon

During the automatic creating of topologic relationship, if the area of the polygon is minus，it is possible a island polygon，and nested by the other polygon .After the polygon whose area is minus is determined, the next important work is to affirm the nested relation among the polygons.

The principle to identify the nested polygon is as the follows.

First, we can select one polygon whose area is minus, find the polygons whose absolute value of the area are greater than it, and their MBR contain it. Then, we judge the inside point of the selected polygon contained in which polygons. If some polygon contain the inside point, the endothecium polygon is the one that called its upper polygon, and it called the nested polygon. If there are no polygon contain its inside point, it has no geometry meaning, we must delete it.

6 Conclusion

The topologic relationship is very important to the Geographic Information System.. It ensures the spatial analysis and is the base of the assistant decision-making. On the creating of the topologic relationship, I successfully use the above algorithms in the node matching, polygon tracing, polygon inside point getting, and nested polygon identification., etc.

Reference

1. Cai shaohua．The Researching and Practicing of GIS Graphics Spatial Relationship．The PhD paper of the PLA institute of surveying and mapping, 1999
2. Huang xinyuan, Tang qin． The Generality of Geographic Information system． The higher education press of China，1992
3. Wang Jiayao． Map Digitizing Techniques and Digitized Map Production． Xi an map press，China, 1998