EXTRACT SPATIAL ASSOCIATION RULES BY THE METHOD OF SPATIAL ANALYSIS

Liu Xiaosheng[a,b,*], Ren Haifeng[a]

[a] School of Environmental and Architectural Engineering, Jiangxi University of Science and Technology, 86 Hongqi Road, Ganzhou, Jiangxi, 341000, China – (lxs9103, renhf80)@163.com
[b] Tongji university, Shanghai, 200092, China – lxs9103@163.com

ABSTRACT:

When the ability of people's collecting data becomes strong, the quantities of data are increased rapidly and a mass of data are submerged. The present database system cannot discover the existed relations and rules behind data, also it cannot forecast the to-be development trend by the existing data. To solve this problem, the paper expatiates several related concepts about spatial association rules firstly, then describes the algorithm of extracting spatial association rules detailedly and an example is given. The mining algorithm discussed in this paper uses the supportive degree and the trustful degree to simplify spatial topological relation, but the algorithm does not transform the problem from mining the spatial association rules to mining property association rules. The algorithm uses the means of spatial analysis to describe spatial relations, then the spatial association rules can be got. By an example, we testified the feasibility of the algorithm.

Spatial data mining, which is an important research problem of KDD(Knowledge Discovery in Database), means extracting not-evident knowledge、spatial relations etc from spatial data. The aim of spatial data mining is to discover the relations among each items in the given database. With the development of database technology and the popularization of data application, kinds of different forms of data are coming forth. In order to extract useful data from excessive data, we must study on data mining.

Almost all the GIS used now utilizes spatial data only on query、spatial analysis and simple statistic, however it can only satisfy the requirement of some low levels. With the popularization of GIS, people need urgently to mine knowledge which is more directive to decision-making. To solve this problem, the paper discussed some related concepts firstly, then studied on the algorithm of extracting spatial association rules by the method of spatial analysis, and an example is given.

1. SEVERAL RELATED CONCEPTS

1.1 Spatial Association Rules

An item of spatial association rules can always be defined as $X \rightarrow Y(s\%,c\%)$,in which X and Y are collectives of spatial and not-spatial predications, s% and c% stands for the supportive degree and the trustful degree respectively. The supportive degree here means the probability of spatial or not-spatial predications X and Y happens at the same time. The trustful degree here means the probability of Y happens, when spatial or not-spatial predications X have happened. (He Jing.2003)

1.2 The Core Problems of Spatial Analysis

Spatial analysis, is a function system which describes objects' spatial relations and spatial behavior, offers reference to objects' spatial query and spatial correlation analysis and services for spatial decision-support, is based on the definition and classification of geographical spatial objects' configuration. The object of spatial analysis is geographical spatial object which has been classified and defined configuration (geometrical character). Different kind of object has different configuration, the analysis to configuration is called configuration analysis. For example, we can divide geographical spatial objects into point、line、plane and cube. A point has the configuration of location. A line has the configuration of length、 direction etc.

Spatial relation means all sorts of relations between spatial entities, which is formed because the entities have different location and figuration. The relations can be divided into three types: topological relation, scale relation (it includes distance relation and direction relation), order relation (it is decided by the location of the observer). The topological relation between two objects has the character that it is not changed by the topological change of the referenced object. Scale relation utilizes some spatial measurements to describe the relations between spatial objects, which mainly includes distance relation and direction relation. Direction relation means relative orientation relation between two spatial objects A and B, expressed by BRA, in which R means direction relation, A is the source object and B is the target object. The deciding of direction relation depends on the numbers of points that form the object. Here, we use representative point Exp(A) to stand for A and consider all points of B for B, then:

If B north A is right, to all points of B, we know their y coordinate is bigger than the y coordinate of Exp(A). Similarly, we can define south、west and east.

If B northeast A is right, to all points of B, we know their x coordinate is bigger than or equal to the x coordinate of Exp(A) and their y coordinate is bigger than or equal to the y coordinate of Exp(A) too. Similarly, we can define southeast、southwest and northwest。

To B any-direction A, we know it is always right.

Order relation is decided by the location of the observer, so the former two relations are studied more. (Wu Xincai,2002)
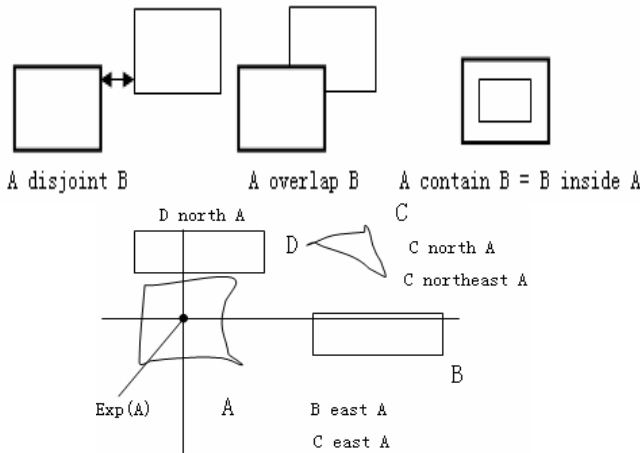
Figure 1. Several typical spatial relations

The analysis of spatial behavior means studying of the changes of object's configuration and structure in some conditions such as space and time etc. With the change of scale, the type of point、line and plane may be changed, a plane may be changed to a point. With the change of time and spatial location, some spatial configuration characters of the object may be changed.

1.3 The Extension of Spatial Analysis

Spatial query is defined as to find corresponding object in a group of spatial objects.

Spatial correlation analysis (also called subject analysis) analyses spatial relation of spatial objects under the integration of property and geometrical character, aims mainly at the property of spatial object. Spatial correlation analysis considers geometrical character and property together, will involve the problems of spatial objects' reclassification and spatial objects' spatial relation under some certain subject. For example, we can reclassify spatial objects according to a certain subject and show the objects by different subjects.

In a word, spatial analysis firstly classifies spatial objects; then describes spatial objects' configuration, the relation between objects and the behavior of objects, which are the core problems of spatial analysis; based on the concepts above, we extend the meaning of spatial analysis to spatial query and spatial correlation analysis, then the framework of spatial analysis is formed. (Chen Jiangping,2003)

2. EXTRACTING OF SPATIAL ASSOCIATION RULES

The mining of spatial association rules is extracting all kinds of spatial relations from masses of spatial objects, so its price is something high. Here, we will introduce a mining algorithm called progressive refinement mining algorithm, which can be used to analyse spatial association. The algorithm mines a relatively large data set quickly and glancingly first, then utilizes a costlier algorithm to perfect the quality of mining.

2.1 The Description Of Progressive Refinement Mining Algorithm

Progressive refinement mining algorithm can be described as: (1) carries out spatial analysis to find the objects needed to be mine in a certain range(analysis of buffer);(2) utilizes the

technology of MBR, R+ to query quickly. MBR means Minimal Bounding Rectangle is always used to stand for spatial object. An object's MBR is defined as the minimal rectangle that contains the object completely; (3) based on the topological relations above, calculates these predications' supportive degree, removes the items whose supportive degree is small; (4) utilizes the technology MRR(Multi-Rectangle Render) to check-up the predications' relations formed by step (3),gets rid of the spatial predications' relations that is not matched to the real situation, forms new topological relations datasheet, calculates the supportive degree of these predictions and gets rid of the items whose supportive degree is small; (5) adopts the concept level tree(figure 2 is a concept level tree of spatial association relation) to analyse the topological relations datasheet formed in step(4), forms news topological relations datasheet and extracts spatial association rules.(Zhu Xiaodong,2004)
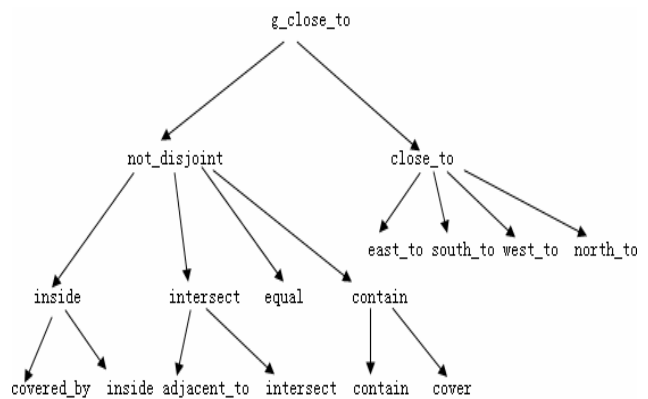

Figure 2. A concept level tree of spatial association relation

MBR is always used to stand for spatial object approximately, but MBR has an inherent defect that the relation of MBR may not be consistent with the real topological relation of the objects, for example, if two objects' MBR is overlapping, but the real situation is that maybe they are disjoint (figure 3 is an example). To reduce the abnormity of topological relations decided by MBR, increase precision expressed by topological relations and efficiency of relation calculation and query, we introduce a variation of MBR---MRR (Multi-Rectangle Render), which can express object's spatial relation comparatively factually. MRR can be divided into three types:

① Consistent MRR. Consistent MRR plots out an object's MBR into four same parts in any level, and the part that is not intersectant to any portion of the object will be ignored. Different object has different number of grids, but to an object's MBR, the rectangles been plotted out have the same size. Figure 3 is a simple example of consistent MRR.

②Not consistent but regular MRR. The application of not consistent but regular MRR can be understood as four-fork tree for spatial index. Any MRR can be divided into four same parts in any degree until it satisfies the requirement. Any rectangle that contains nothing of the object will be got rid of. Why do we say the rectangles are regular? Because although some rectangles are not equal to each other, the big one is always 4X times bigger than the small one. Compared with MBR and consistent MRR, the verge of this kind is more accurate, at the same time the influenced range of inconsistent topological relation is reduced, and more accurate deduction can be carried

out for detailed levels. Figure 4 is an example of not consistent but regular MRR.
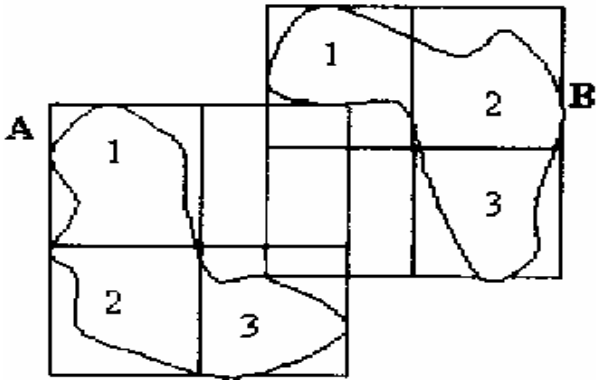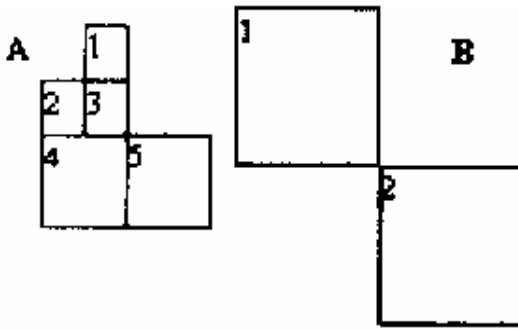

Figure 3. An example of consistent MRR

Figure 4. An example of not consistent but regular MRR

③ Not regular MRR. Not regular MRR requires special methods to plot out spatial objects. The former two methods divide MBR into the same size, but not regular MRR may not divide MBR into the same size, whose aim is to express the object with the least grids and reduce difficulty of calculation. (Chen Lin,2002)

2.2 The Application Example Of Extracting Spatial Association Rules

To testify the feasibility of the algorithm, we utilized the algorithm to analyse the distributing situation of our country's freeway both in east area and central and west area. By analysis, we got the topological relation datasheet of table 1 and table 2. (the data in the both two tables have been disposed by step(1) and step(2) of the algorithm, the entrance distance in the tables means the nearest distance from the entrance to city)

City name	Population(ten thousand)	Rank of city	entrance distance
east area city 1	>80	Capital of a province	<3km
east area city 2	>80	Capital of a province	<3km
east area city 3	>80	Regional city	<3km

City name	Population(ten thousand)	Rank of city	entrance distance
east area city 4	>40	Regional city	<3km
east area city 5	>80	Regional city	<3km
east area city 6	>40	Regional city	<5km
east area city 7	>80	Regional city	<3km
east area city 8	>40	Regional city	<5km
east area city 9	>40	county	<3km
east area city 10	>10	county	<5km
east area city 11	>10	county	<5km
east area city 12	>10	county	<5km
east area city 13	<10	county	<5km
east area city 14	<10	county	<5km
…	…	…	…

Table 1.The topological relation datasheet of east area

City name	Population(ten thousand)	Rank of city	entrance distance
central area city 1	>80	Capital of a province	<3km
central area city 2	>80	Capital of a province	<3km
west area city3	>80	Capital of a province	<3km
west area city 4	>80	Capital of a province	<3km
central area city 5	>80	Regional city	<3km
central area city 6	>40	Regional city	<5km
west area city 7	>80	Regional city	<3km
west area city 8	>40	Regional city	<5km
central area city 9	>10	county	<5km
central area city 10	>10	county	>5km
central area city 11	<10	county	>5km
west area city 12	>10	county	<5km
west area city 13	>10	county	>5km
west area city 14	<10	county	>5km
…	…	…	…

Table 2. The topological relation datasheet of central and west area

We got the topological relations datasheet table 3 after disposal

of pruning

City name	Population (ten thousand)	Rank of city	entrance distance	Supportive degree
east area city 1*	>80	Capital of a province	<3km	89
east area city 2*	>80	Regional city	<3km	83
east area city 3*	>40	Regional city	<3km	72
east area city 4*	>40	Regional city	<5km	85
east area city 5*	>10	county	<5km	71
east area city 6*	<10	county	<5km	50
central area city 1*	>80	Capital of a province	<3km	81
central area city 2*	>80	Regional city	<3km	75
central area city 3*	>40	Regional city	<3km	56
central area city 4*	>40	Regional city	<5km	63
central area city 5*	>10	county	<5km	33
central area city 6*	<10	county	<5km	13
west area city 1*	>80	Capital of a province	<3km	73
west area city 2*	>80	Regional city	<3km	69
west area city 3*	>40	Regional city	<3km	42
west area city 4*	>40	Regional city	<5km	51
west area city 5*	>10	county	<5km	23
west area city 6*	<10	county	<5km	11
…	…	…	…	

Table 3. The topological relations datasheet
after disposal of pruning

we can get association rules from table 3 as follows:

is_a city∧Rank of city= Capital of a province→there is an entrance to freeway in the range of 3kms∧Population>80 ten thousand
is_a city∧Population>80 ten thousand→ there is an entrance to freeway in the range of 3kms
is_a city∧City name= east area city *∧Population>10 ten thousand→ there is an entrance to freeway in the range of 5kms

From the last rule, we know that there is always an entrance to freeway for an east area city in the range of 5kms, while it is not always true for a west one, this is consistent with the real situation, so the algorithm is feasible.
Of course, we can get other association rules by superposing according to other methods.

3 .Epilogue

This process of extracting spatial association rules can be divided into three steps: decide the spatial objects, simplify the spatial object's relation and extract spatial association rules. The progressive refinement mining algorithm discussed in this paper is different to other mining algorithm of spatial association rules : although the supportive degree and the trustful degree is used in the step(2) and step(4) to simplify spatial topological relation, but the algorithm does not transform the problem from mining spatial association rules to mining property association rules. The algorithm uses the means of spatial analysis to describe spatial relations, then the spatial association rules are got. By practice, we know the algorithm is easy in the aspects of both time and space.

References
Chen Lin, Du Youfu, Wang Yuanzhen. 2002. Spatial Relationship Model Based on MBR. *Computer Engineering and Applications*,05,pp.76-78

Chen Jiangping, Fu Zhongliang, Bian Fuling, Sha Zhongyao. 2003. Mining Spatial Association Rules with Spatial Analysis. *Computer Engineering* ,07,pp.29-31

He Jing, Wang Lizhen, Zou Likun. 2003. Discovery of Spatial Association Rules Based on the Meteorological Data of Yunnan Province. *Computer Engineering and Applications*. 34.pp.187-190

Wu Xincai, Liu Shaoxiong. 2002. Spatial Data mining Based on Neighborhood Relation. *Computer Engineering*. 07.pp.89-91.

Zhu Xiaodong, Ni kai. 2004. The algorithm and realization of mining association rules based on spatial data. *Journal of Ningbo Polytechnic*. 06.pp.90-92.