THE STUDY FOR MATCHING ALGORITHMS AND MATCHING TACTICS ABOUT AREA VECTOR DATA BASED ON SPATIAL DIRECTIONAL SIMILARITY

Guo Li, Lv Zhiping, Zhang Bin, Wang Yaoge

Institution of Surveying and Mapping, Information Engineering University, Zhengzhou, China. Postcode: 450052
Email: gl_750312@163.com

KEY WORDS: matching algorithm, matching tactics, data matching, data fusion, spatial directional relation, spatial directional similarity

ABSTRACT:

The matching technique is the key to geospatial data integration and fusion. In this paper, we describe spatial directional relation through direction relational matrix, and discuss calculation methods of the spatial directional similarity. Combining with the area vector data matching algorithms, we introduce some matching tactics, which contribute to accomplish many-to-one, many-to-many matching. We describe matching process based on spatial directional similarity and matching tactics. Then we take a test, using building property area data and town planning area data at a scale of 1：500. In the test, we take the spatial directional relation matrix algorithm, and adopt two-steps matching tactics. Based on the test, we draw a conclusion: matching efficiency depends on not only algorithms, but also tactics. Proper tactics can help us accomplish more complex task.

1. INTRODUCTION

There are different expression forms for the same phenomenon. Data matching technique can recognize the same object from different expression forms. So data coherence matching technique is the key to spatial data integration and fusion. There are abundant semantic information, complex topological relations, different geometry shapes and location differences in the vector spatial data, so the automatic matching techniques is always the difficult point and hotsNW ı the corresponding study (Zhang Qiaoping, 2002).

Spatial vector data matching approaches consists of geometry matching, topological matching and semantic matching. In order to improve the matching efficiency and validity, we also have to make full use of matching tactics. We combine the spatial directional similarity algorithm and two step tactics to realize complex area objects matching.

2. SPATIAL DIRECTION CONCEPTION DESCRIPTION

Spatial direction can be described as an orientation from one spatial object pointing to another object in a direction reference system. There are two aspects, quantitative and qualitative attributes can be used to describe spatial directional relation. Quantitative directions model gives an accurate direction angle to depict the direction. Qualitative directions model have 8 kinds of models, including direction relation matrix model (Deng Min, 2006). We adopt directional relation matrix model to describe directional relation and to calculate similarity value.

2.1 Description for Spatial Directional Relation

Relation matrix model divide the space into 9 absolute direction pieces (Figure 1). The outer 8 pieces express 8 directions, and the centre piece is named of O. Generally, the reference object lies in the centre piece(Guo Qingsheng, 2004).

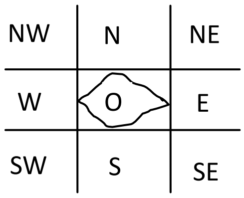

Figure 1. Spatial directions separations

2.2 Spatial Direction Relation Matrix

If A is a reference object, we divide the space according the minimum bounding rectangle of the reference object. A 3×3 direction relation matrix is the following(Ding Hong, 2004):

$$dir(A,B)=\begin{bmatrix} A_{NW}\cap B & A_N\cap B & A_{NE}\cap B \\ A_W\cap B & A_0\cap B & A_E\cap B \\ A_{SW}\cap B & A_S\cap B & A_{SE}\cap B \end{bmatrix} \tag{1}$$

Where A is a reference object and B is the target object.
A_{NW}, A_N, A_{NE}, A_W, A_O, A_E, A_{SW}, A_S, A_{SE} are 9 direction pieces.
$A_{NW}\cap B$ means the part of B, which is located in the direction of piece A_{NW}.

2.3 Spatial Directional Distance

Spatial directional distance represents the difference of spatial direction pieces; and its inverse presents spatial direction similarity. As the object changes its direction, the distance changes accordingly. The distance is defined in figure 2. For example, as the direction changes in one direction piece, the distance is defined 0; as the direction changes from NE to SW or from NW to SE, the distance is defined 4. Other instance refers to figure 2 (Roop K Goyal, 2000).

	N	NE	E	SE	S	SW	W	NW	O
N	0	1	2	3	2	3	2	1	1
NE	1	0	1	2	3	4	3	2	2
E	2	1	0	1	2	3	2	3	1
SE	3	2	1	0	1	2	3	4	2
S	2	3	2	1	0	1	2	3	1
SW	3	4	3	2	1	0	1	2	2
W	2	3	2	3	2	1	0	1	1
NW	1	2	3	4	3	2	1	0	2
O	1	2	1	2	1	2	1	2	0

Figure 2. The distance of four neighbouring directions pieces

2.4 Equation of Spatial Directional Similarity Value

The spatial directional similarity value between two directional is defined as the following equation (Roop K Goyal, 2000):

$$S_i(D^0, D^1) = 1.0 - \frac{d(D^0, D^1)}{D_{\max}} \tag{2}$$

Where $D_{max} = 4$

$d(D^0,D^1)$ = the least direction distance from matrix D^0 to matrix D^1.

$S_i(D^0,D^1)$ = the spatial directional similarity value of D^0 to matrix D^1.

3. CALCULATION FOR SPATIAL DIRECTIONAL SIMILARITY VALUE

Equation 2 is tight, but is hard to calculate. We can realize the algorithm of spatial directional similarity calculation efficiently if we adopt raster data format. Let's take an example, the directional relation matrix between the target area object and the reference area object can be changed to calculate the matrix summation between all the raster cells of the target area object to the reference area object.

3.1 Calculation for Spatial Directional Similarity in Ideal Instance

We can get the average directional distance value by adding up each raster cell direction distance value and dividing by the sums of cells. The formula of calculating the area objects similarity value is the following (Ding Hong, 2004):

$$S(I_r, I_p) = 1 - \frac{1}{4n}\sum_{i=1}^{n} d_i \tag{3}$$

Where I_r = the reference object

I_p = the target object

n = the raster cells numbers of the target object

d_i = the moving distance of each cell

$S(I_r, I_p)$ = the similarity value of the two object.

3.2 Calculating for Spatial Directional Similarity in Ordinary Instance

Because of different data sets, the same object may present the different shape. The above-mentioned method does not adapt to the ordinary instances. How does it to calculate the similarity between the objects with different shape? We can solve it by this way.
Let's take Figure 3 for example.

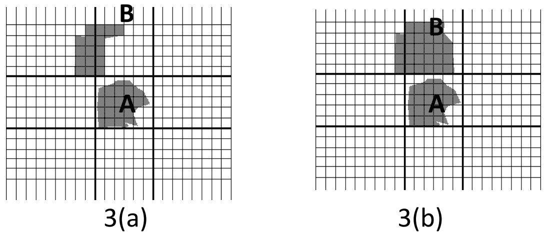

Figure 3. Similarity calculation for object with different shape

Firstly, we take A as the reference object. B is the object with different shape in 3(a) and 3(b).
We calculate the similarity value of B from figure 3(a) to figure 3(b). We take the raster cells of B in figure 3(a) as norm. The principle is:
① comparing to B in figure 3(b), if the cell does not exist in B of figure 3(a), but exist in B of figure 3(b), the moving distance is 0.
② comparing to B in figure 3(b), if the cell exist in B of figure 3(a), but does not exist in B of figure 3(b), the moving distance is 4.
③ the other instance, referring to figure 2.
The calculating course is as follows:
The number of raster cells in B of figure 3(a) is 16. There are 5 cells in B of figure 3(a) moving form NE to N comparing to B of figure 3(b). The spatial directional similarity value of B from figure 3(a) to figure 3(b) is *Sim*1(*a,b*).

$$Sim1(a,b) = 1 - \frac{1}{4\times 16}(1\times 5) = 0.92 \tag{4}$$

Secondly, we take the raster cells of B in figure 3(b) as norm. Then we calculate the similarity value of B form figure 3(b) to figure 3(a). The number of raster cells in B of figure 3(b) is 30. There are 5 cells in B of figure 3(b) moving from N to NE, comparing to B of figure 3(a). Furthermore, there are 13 cells exist in B of figure 3(b), but don't exist in B of figure 3(a), the moving distance is 4. The spatial directional similarity value of B from figure 3 (b) to figure 3 (a) is *Sim*2(*a ,b*)

$$\begin{aligned} &Sim2(a,b) \\ &= 1 - \frac{1}{4\times 30}(1\times 5 + 4\times 13) = 0.525 \end{aligned} \tag{5}$$

Finally, we set a threshold K. If the two similarity values are both bigger than K, we think that the object B in figure 3(a) and figure 3(b) are the same object. It is one to one match.

3.3 Matching Tactics for Complex Instances

The above-mentioned method can realize one-to-one match, but cannot solve matching problems of one-to-many, many-to-one, especially many-to-many. In order to solve complex matching question, we adopt some matching tactics. The main thought is the following:
There are two different data sets, A and B. The area objects in dataset A are $\{a_1,a_2,...,a_i,...,a_m\}$, and area objects in data B are $\{b_1,b_2,...,b_j,...,b_n\}$. The objects numbers of the two datasets may be not equal, that is to say, m may be not equal to n. The aim for match is to find out homonym entities which exist in different data sets.

3.3.1 One-to-one Matching Tactic

There are entities:
$a_i \in A, i=1,2,...m$;
$b_j \in B, j=1,2,...,n$;
A and B are two data sets.
If: $Sim(a_i,b_j) \geq$K, and $Sim(b_j,a_i) \geq$k;
K is the threshold, and $0<K<1$.
We can draw a conclusion: the object b_j in data set B is the homonym to a_i; meanwhile a_i in data set A is the homonym to b_j. So, a_i and b_j are homonym entities, and the matching relation is one-to-one matching. The chart is shown in the Figure 4.

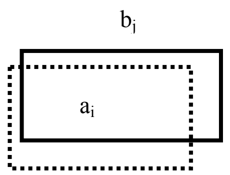

Figure 4. One-to-one matching

3.3.2 Many-to-one Matching Tactic

There are entities a_i,a_t,b_j,:
$a_i \in A$, $i=1,2,...m$;
$a_t \in A$; $t=1,2,...m$;
$i \neq t$;
$b_j \in B, j=1,2,...,n$;
A and B are two data sets.
If: $Sim(a_i,b_j) \geq$K and $Sim(a_t,b_j) \geq$K;
K is the threshold, and $0<K<1$.

We can draw a conclusion: the entity a_i and a_t are match to the entity b_j. So, the aggregation of $\{a_i, a_t\}$are match to entity b_j . The match relation is many-to-one. The chart is shown in the Figure 5.

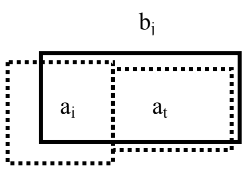

Figure 5. Many-to-one matching

3.3.3 Many-to-many Matching Tactic

There are entities a_i ,a_t ,b_j , b_k ,b_r;
$a_i \in A$, $i=1,2,...m$;
$a_t \in A$, $t=1,2,...m$;
$i \neq t$;
$b_j \in B, j=1,2,...,n$;
$b_k \in B, k=1,2,...,n$;
$b_r \in B, r=1,2,...,n$;
$j \neq k \neq r$;
A and B are two data sets.
If: $Sim(a_i ,b_j) \geq$K, *and* $Sim(a_i ,b_k) \geq$K,
and $Sim(a_t ,b_k) \geq$K, *and* $Sim(a_t ,b_r) \geq$K;
K is the threshold, and $0<K<1$.
We can draw a conclusion:
a_i is match to $\{b_j , b_k\}$,and a_t is match to $\{b_k ,b_r\}$.

So, the aggregation of $\{a_i ,a_t\}$are match to $\{b_j , b_k ,b_r\}$. The match relation is many-to-many. The chart is shown in the Figure 6.

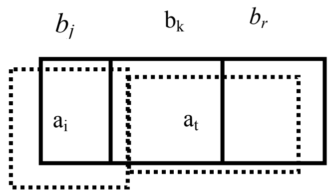

Figure 6. Many-to-many matching

4. DATA MATCHING PROCESS DESCRIPTION

If there are different data sets, we can adopt above algorithm and tactics to realize homonym entities matching. The matching process is shown in the Figure 7:

4.1 Vector Data Rasterization

In order to make use of above algorithm to realize data matching, we must change the data format from vector to raster.

4.2 Ascertain the Reference Object

In certain area, we should choose an appropriate object as reference object.

4.3 Similarity Value Calculation

We adopt above similarity value algorithm, calculate the similarity values of objects which are around the reference object.

4.4 Preliminary matching

According to experiment, we fix a range for threshold value. If the similarity values of some objects are within the scope of threshold, we can get a preliminary conclusion: these objects are probably homonym entities.

4.5 Matching Tactics

Based on the matching algorithms, the tactics can accomplish many-to-one, many-to-many matching.

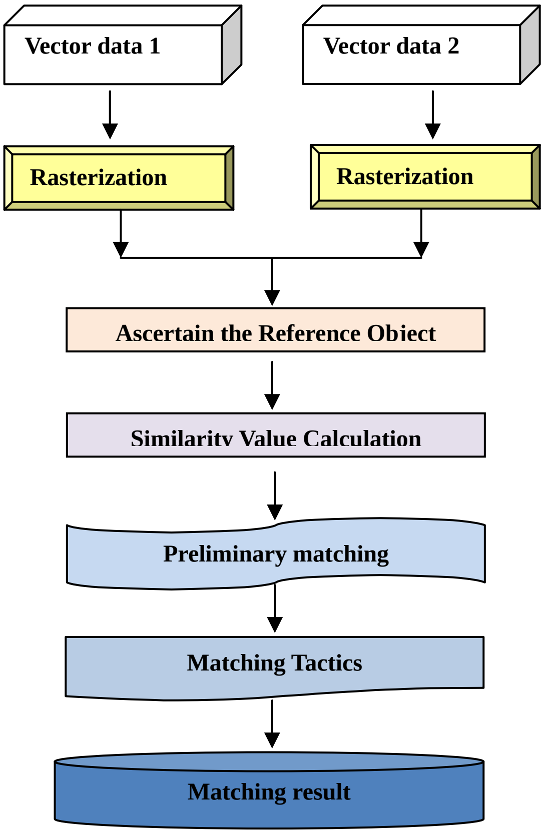

Figure 7. Matching Process

5. CONCLUSION

In the end, we take test and analysis for area vector objects data matching with building property area data and town planning area data at the scale of 1:500. The building property data is shown in the figure 8 and town planning data is shown in figure 9.

We take the spatial directional relation matrix algorithm, and adopt two-steps matching tactics. Firstly, we set up a big threshold to accomplish one-to-one matching. Secondly, we reduce the threshold to match the rest objects, realizing many-to-one, many-to-many matching.

Based on the test, we get some experimental data and make two tables. From table 1, we can see, when the thresholds are big, from 0.6 to 0.9, the accuracy of one-to-one match is from 53% to 85%. When the threshold is 0.8, the accuracy is the highest.

From table 2, we also can see that when we reduce the threshold, the accuracy of many-to-one, many-to-many matching are increased. When the threshold is 0.3, the accuracy of many-to-one, many-to-many matching is highest.

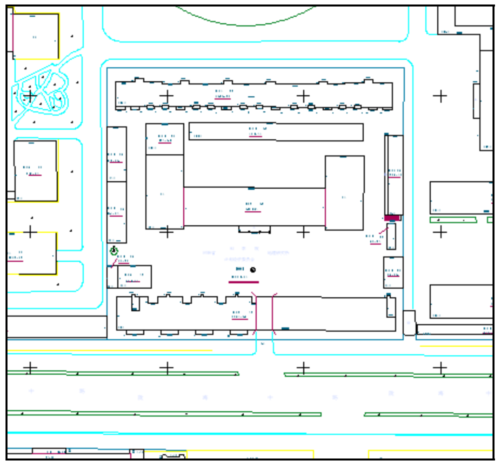

Figure 8. Building property Data

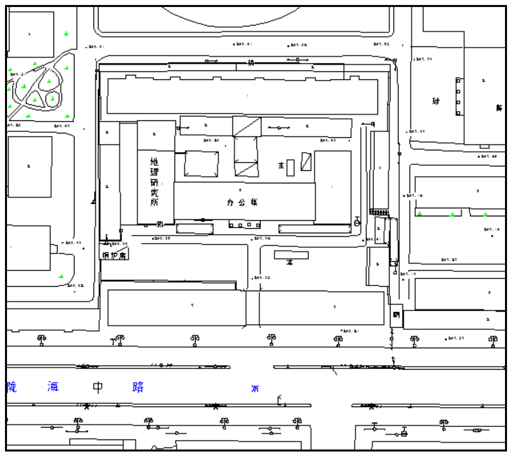

Figure 9. Town planning data

Threshold	To be matching number	Correct matching number	Accuracy (%)
0.6	72	38	53
0.7	72	58	81
0.8	72	61	85
0.9	72	51	71

Table 1 Big threshold match (one-to-one match)

Threshold	To be matching number	Correct matching number	Accuracy (%)
0.3	14	11	78
0.4	14	9	64
0.5	14	7	50
0.6	14	3	21

Table 2 Small Threshold Matching (many-to-many match)

Based on the test figures, we can draw a conclusion. The matching efficiency is not only lies on algorithm, but also depends on tactics. Proper tactics can help us accomplish more complex task.

6. REFERENCES

Deng Min, 2006， Liu Wenbao, Li Junjie, Sun Dian . Computational Model of Spatial Direction Relations in Vector GIS. *Journal of Remote Sensing*, 10(6):821-828

Ding Hong, 2004. *A Study on Spatial Similarity Theory and Calculation Model.* Doctoral Thesis, Wuhan University.

Guo Qingsheng, 2004, Ding Hong. Similarity for Spatial Directions Between Areal Objects in Raster Data. *Editorial Board of Geomatics and Information Science of Wuhan University*, 29(5):447-450

Goyal P K，2000．*Similarity Assessment for Cardinal Directions Between Extended Spatial Objects*：[Ph.D.Thesis]．Orono：The University of Maine.

Roop K Goyal, 2000. *Similarity assessment for Cardinal Directions between Extended Spatial Objects*: [Ph.D.Thesis]. The University of maine.http://www.spatial. maine.edu/ Publications /phd_thesis/ Goyal2000.pdf

Zhang Qiaoping, 2002. *The Research on Areal Feature Matching among the Conflation of Urban Geographic Databases.* Doctoral Thesis, Wuhan University.

7. ACKNOWLEDGEMENTS

The paper presented here is supported by *National High Technology Research and Development Program (863) Project（2009AA12Z305）*.

The author likes to thank the anonymous reviewers for their valuable comments.