

## EVALUATION METHODS FOR ASSOCIATION RULES IN SPATIAL KNOWLEDGE BASE

Xiao NIU <sup>a,\*</sup>, Xiaole JI <sup>a</sup>

<sup>a</sup> Shandong Provincial Institute of Land Surveying and Mapping, Lishan East Road 9, Jinan 250013, China –  
(niuxiao710, jixiaole218)@gmail.com

**KEY WORDS:** Association rule, Evaluation criteria, Lift, Novelty, Evaluation process, Spatial data, Knowledge base

### ABSTRACT:

Association rule is an important model in data mining. It describes the relationship between predicates in transactions, makes the expression of knowledge hidden in data more specific and clear. While the developing and applying of remote sensing technology and automatic data collection tools in recent decades, tremendous amounts of spatial and non-spatial data have been collected and stored in large spatial database, so association rules mining from spatial database becomes a significant research area with extensive applications. How to find effective, reliable and interesting association rules from vast information for helping people analyze and make decision has become a significant issue. Evaluation methods measure spatial association rules with evaluation criteria. On the basis of analyzing the existing evaluation criteria, this paper improved the novelty evaluation method, built a spatial knowledge base, and proposed a new evaluation process based on the support-confidence evaluation system. Finally, the feasibility of the new evaluation process was validated by an experiment with real-world geographical spatial data.

### 1. INTRODUCTION

Association rules mining first proposed in 1993(Agrawal R, Imielinski T, Swami A., 1993.) is an important branch of data mining technology, which aims to find the associations between items in large databases. In some studies, association rules mining commonly relies on statistical methods, the generation of rules is entirely based on data analysis and the relationship of the rule itself are not taken into consideration. At the same time, the generation of association rules is also impacted by the quality and the threshold of support and confidence, which may result in that the discovered rules are useless, or may lose some useful rules. Therefore, the generation of association rules has some limitations.

The traditional data mining system uses the support-confidence evaluation system which uses the support and confidence threshold to exclude some useless rules. Then the rules left can be called strong association rules. However, the strong association rules are not necessarily interesting to users. Therefore, we need to improve traditional evaluation system and introduce new evaluation criteria and evaluation methods to dig out more effective and the user more satisfaction association rules.

### 2. EVALUATION METHODS

At present, there are already a lot of studies on the classification methods of evaluation criteria. The most common classification method is to divide into two types which are objective evaluation methods and subjective evaluation methods (Genlin JI, 2004). On basis of this method, some literature introduces comprehensive evaluation method (Zhandong SU, Fucheng YOU, Bingru YANG, 2004, Xiangguo Dou, 2005). In addition, there are three categories which are elementary evaluation criteria, quantitative evaluation criteria and qualitative evaluation criteria in some paper (Xudong SONG, Kun ZHAI, Weidong GAO, 2007).

In this paper the evaluation method is classified into objective evaluation methods, subjective evaluation methods and comprehensive evaluation methods.

### 2.1 Objective Evaluation Methods

Objective evaluation refers that the interest of the association rules depends on specific structure of rules and the data which is depended in the process of data mining. This kind of methods mainly uses statistical methods and a quantitative value to determine the interest of rules, thus avoiding artificial subjective opinions. So in this sense, objective evaluation of the rules is reliable and convincing. Objective evaluation is to measure the rules with the objective evaluation criteria.

Support which reflects the practicality of association rules and confidence which reflects the validity of association rules, are two commonly objective evaluation indicators. And support is usually expressed with S, and confidence is usually expressed with C. In this paper we mainly focus on the new objective evaluation criteria-lift. The following example can indicate the necessity of introducing lift.

Example 1. There is a meteorological station. We get average daily temperature and precipitation data of one thousand consecutive days of the station and divide the temperature and precipitation into high, medium, and low three categories through cluster analysis. The results are shown in Table 1.

Name	Precipitation high	Precipitation medium	Precipitation low	Total
Temperature high	50	50	400	500
Temperature medium	20	20	190	230
Temperature low	20	30	220	270
Total	90	100	810	1000

Table 1. The classification results of average daily temperature and precipitation 1000 days (units: day)

The minimum support threshold is set to 0.2, and the minimum confidence threshold is set to 0.75. That can be got:  $S(\text{Temperature high}, \text{Precipitation low}) = 0.4$ ,  $C(\text{Temperature high} \Rightarrow \text{Precipitation low}) = 0.8$ . Here S means support, and C means confidence.

Obviously, the rule: Temperature high  $\Rightarrow$  Precipitation low (0.4, 0.8), satisfies the minimum support threshold and minimum confidence, namely strong association rules, but this rule is a wrong. Because the possibility of precipitation low already has 0.81 is higher than confidence 0.8. In spite of this rule is the strong association rule, but cannot explain the emergence of antecedent of the rule would increase the appearance of consequent of the rule. In fact, in this case the temperature high and precipitation low is negatively correlated.

From Example 1 it can be seen that the confidence of rule  $X \Rightarrow Y$ , which only gives out the estimation of conditional probability of set X and set Y, and cannot measure the actual strength and correlation between X and Y, has a certain deceptive. Therefore, the objective evaluation criteria lift is put forward and is defined as follows (Jiawei Han, Micheline Kanber, 2000).

Definition 1 IF  $P(X \cup Y) = P(X)P(Y)$ , the appearance of set X is independent of the emergence of set Y, otherwise they are relevant. The formulas of interest measures of the appearance of set X and the emergence of set Y are

$$lift(X \Rightarrow Y) = \frac{P(X \cup Y)}{P(X)P(Y)} \quad (1)$$

If the value of the formula (1) is greater than 1, then X and Y are positive correlation ,and it means that the appearance of one implicates the appearance of another. Then if the result is less than 1, then there appears that X and Y are negatively correlated. And if the result is equal to 1, then X and Y are independent, and there is no correlation between them.

Formula (1) is equivalent to  $P(Y|X)/P(Y)$ , which also is called the lift of rule  $X \Rightarrow Y$ . That is to say, it estimates the lift degree of one appearance to another appearance.

Table2 is the result after introducing lift.As it can be seen that the first rule and the fourth rule are wrong, and the antecedent and consequent of the third, fifth and sixth rule are independent of each other, and only the value of lift of the second rule is more than 1,that is to say the antecedent and consequent of which is relatively strong correlation. There is just the second rule which is interesting after filtering by lift. So we can get the conclusion that the objective evaluation criteria lift can filter and remove some wrong and meaningless rules.

Rule $X \Rightarrow Y$	S(XY)	S(Y)	C( $X \Rightarrow Y$ )	L( $X \Rightarrow Y$ )
Temperature high $\Rightarrow$ Precipitation low	0.40	0.81	0.80	0.99
Temperature medium $\Rightarrow$ Precipitation low	0.19	0.81	0.83	1.02
Temperature low $\Rightarrow$ Precipitation low	0.22	0.81	0.81	1.00
Precipitation low $\Rightarrow$ Temperature high	0.40	0.50	0.49	0.98
Precipitation low $\Rightarrow$ Temperature medium	0.19	0.23	0.23	1.00
Precipitation low $\Rightarrow$ Precipitation low	0.22	0.27	0.27	1.00

Table 2.the results of introducing lift of association rule mining(S is support, C is confidence, L is lift).

## 2.2 Subjective evaluation methods

Subjective evaluation refers that evaluation of rules and should reflect the subjective factors such as user participation and integration of domain knowledge. And subjective evaluation can be done using subjective evaluation criteria.

Subjective evaluation criteria mainly include several such as simplicity, user interest degree, novelty and so on. Simplicity is an indicator used to measure the ultimately understandable degree of association rules. And user interest degree reflects the potential usefulness. In this paper we pay more attention to the novelty criteria.

Novelty is an important indicator of the interestingness of association rules and is mainly relative to the original knowledge. The novelty of association rules, which reflects in the difference between the rule found and the rule of knowledge base, can be calculated according to the difference between every item of antecedents and the difference between every item of consequents. In the knowledge base there are many rules which are from domain knowledge and some rules suggested by users. And the difference mentioned should be measured respectively from the angle of the language variables and different language value of the same language variables which have been defined(Bingru Yang,1998).

The calculating formula of novelty has been given in (Yanxia QI, 2004). After analyze these formulas it is clear that, in their article novelty just aims to find the rules which have large differences with the overall rules of knowledge base, but the defect is that novelty cannot evaluate the quality of the rules found. For that we improved the formulas of novelty and Figure 1 shows the calculation. Note that in Figure 1, E is a collection of the rules found, and |E| is the total number of rules. And K is a collection of the rules of knowledge base, and |K| is the total number of rules.  $W_{(i,j)}$  is the novelty of the rule  $E_i$  relative to  $K_j$  and then  $W_i$  is defined as the novelty of  $E_i$  relative to K.

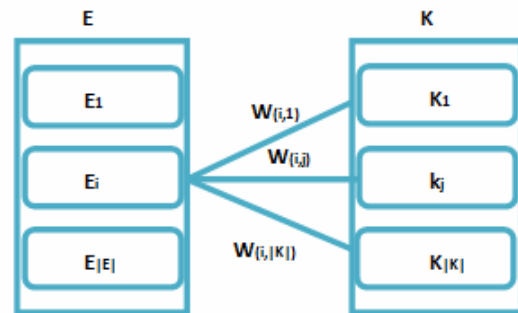


Figure 1.Novelty calculation

The process of calculation of  $W_i$  includes four steps as follows.

Firstly, the calculation of  $L_{(i,j)}$ .

$L_{(i,j)}$  is the novelty of antecedent of the rule  $E_i$  relative to antecedent of the rule  $K_j$  and is equal to the sum of the difference degree of each item in the antecedent. And the calculation formula is

$$L_{(i,j)} = \sum_{k=1}^{|I|} V_{(i,j)k} \quad (2)$$

Note that I is a collection of language variables of antecedent of the rule  $E_i$ , and |I| is the total number of language variables of antecedent of the rule  $E_i$ . And J is a collection of language variables of antecedent of the rule  $K_j$  in knowledge base.  $V_{(i,j)k}$  is the difference degree of  $l_k$  which is any item in I, and the rule  $K_j$ . There are two cases: when  $l_k \notin J, V_{(i,j)k} = 2$ , and when  $l_k \in J, V_{(i,j)k} = 1 + negk$ .

The negk is the difference degree of the language value of  $l_k$  and the language value of the same language variables in  $J$ . negk plus 1 is in order to avoid the value is equal to zero.

Secondly, the calculation of  $Z_{(i,j)}$ .

$Z_{(i,j)}$  is the novelty of consequent of the rule  $E_i$  relative to antecedent of the rule  $K_j$ . After the rules being simplified, the number of items of consequent of the rules in the knowledge is one, and the number of items of consequent of the rules found with data mining algorithm is also one. Therefore, there are two cases for  $Z_{(i,j)}$ : when the consequents of the two rules do not belong to the same language variables,  $Z_{(i,j)}$  equals to 2, when the consequents of the two rules belong to the same language variables,  $Z_{(i,j)}$  equals to 1 plus negk. Adding 1 to negk is to avoid the zero value.

Thirdly, the calculation of  $W_{(i,j)}$ .

The original formula is

$$W_{(i,j)} = \frac{L_{(i,j)}Z_{(i,j)}}{\max(|I|, |J|)} \quad (3)$$

And the improved formula is

$$W_{(i,j)} = \frac{L_{(i,j)}Z_{(i,j)}}{|I|} \quad (4)$$

Compare with the original formula, it can be seen that the calculation formula (4) emphasizes the differences between  $E_i$  and  $K_j$ , and can't reduce the differences because of the more language variables of  $K_j$  antecedent.

Finally, The calculation of  $W_i$ .

The original formula is

$$W_i = \frac{\sum_{j=1}^{|K|} W_{(i,j)}}{|I|} \quad (5)$$

By analyzing the above formula, it has been concluded that the calculation of  $W_i$  aims to find out the rules whose differences with the overall rules of knowledge is high, which can discover the rules whose novelty is high. And when the value of  $W_i$  is greater, it shows the novelty is higher. However, when the value of  $W_i$  is lower, it does not indicate any meaning. Therefore, this novelty cannot measure the quality of the discovered rules and it is needed to improve the original algorithm. The improved formula is as follows.

$$W_i = \min(W_{(i,1)}, W_{(i,2)}, \dots, W_{(i,j)}) \quad (6)$$

Formula (6) can find the minimum difference, which is also can be considered as the similarity, of the rules found relative to the overall rules of knowledge base. And the minimum difference can be used to measure the novelty of rules. When  $W_i$  is smaller, the similarity is higher, and when  $W_i$  is greater, the novelty is

higher. This formula limits the value of novelty between 1 and 4, which is important for the results analysis.

We generally believe that the rules in knowledge base are correct. By the improvement of original algorithm, novelty can measure the similarity between the rules found and the rules in knowledge base-in other words is the evaluation quality of the rules found. Besides, it can also find out reliable rules to enlarge the knowledge base. Novelty mentioned in this article refers to improved novelty.

As a conclusion, novelty can be used to evaluate the quality of association rules, can also perfect the knowledge base, and provide theoretical basis for the relevant research in the future.

### 2.3 Comprehensive Evaluation Methods

There are some defects for various existing evaluation indicators, and at present there is not a particularly effective evaluation method which can be applied to all of the data or to meet the needs of different users. Usually several methods are combined organically to measure the rules. So a comprehensive evaluation indicator is proposed in some literature, whose main idea is to combine with objective evaluation methods and subjective evaluation methods, then to produce an index containing a variety of evaluation indicators.

The comprehensive evaluation index is mainly the combination of objective evaluation criteria and subjective evaluation criteria. By the introduction of objective evaluation methods and subjective evaluation methods in front, it can be drawn that objective evaluation indicators have three, namely: support, confidence and lift, and subjective evaluation indicators also have three, namely: novelty, simplicity and user interest degree. Since the value of each index is not the same, so the comprehensive evaluation index is defined as the weighted average of these evaluation indicators(Zhandong SU, Fucheng YOU, Bingru YANG, 2004), and comprehensive evaluation index is denoted RI. The formula is

$$RI = S^{w_1} \times C^{w_2} \times Lift^{w_3} \times W_i^{w_4} \times USI^{w_5} \times \left(\frac{1}{CN}\right)^{w_6} \quad (7)$$

Where  $W_i$  = the value of novelty

$USI$  = the value of user interest degree

$CN$  = the value of simplicity

$$w_i \geq 0, (i = 1, 2, \dots, 6), \sum_{i=1}^6 w_i = 1$$

The main reason of using geometric weighted average is that the importance of the individual indicators is not equal, and for different areas and applications, their aims are diverse. Accordingly, it needs to determine its weight depending on the specific circumstances. At the same time the value of geometric weighted average could reflect the changes of the index system in general. However, the setting of specific weights is not easy to grasp.

Some person has tried using the method of the arithmetic average (Xiangguo Dou, 2005), and the formula is as follows.

$$RI = k_1 \times S + k_2 \times C + k_3 \times Lift + k_4 \times W_i + k_5 \times USI + k_6 \times CN \quad (8)$$

Where  $k_i \geq 0, (i = 1, 2, \dots, 6), \sum_{i=1}^6 k_i = 1$

However, not all indicators are proportional. And when the dimension is different, it will be affected using the arithmetic average. So this method is not without flaws.

It can be seen, in the two methods, setting the specific weight is not easy to grasp, thus comprehensive evaluation is not very practical. So the comprehensive criteria will not be as a part of the new evaluation process in section three.

### 3. THE NEW EVALUATION PROCESS AND EXAMPLE

#### 3.1 The new evaluation process

In association rules mining, it is an important aspect to evaluate the mining results, for it directly affects the quantity and quality of these results. However, currently the researches on association rules mainly focus on improving the efficiency and performance of mining algorithm, only a few of them pay enough attention to evaluation method.

The traditional support - confidence evaluation framework itself is flawed. The strong association rules which meet the minimum support threshold and minimum confidence threshold are not necessarily interesting, and the framework has an insurmountable flaw of its own, that is we can't analysis on the rare items whose support value is less than the minimum support threshold. And if the lower support threshold or cancel the evaluation criteria of support will have a lot of rules in which there are may have many obvious or irrelevant rules. Therefore, the introduction of new evaluation criteria can dig out the rules more useful.

For the evaluation of the association rules use evaluation criteria to measure, but there is a certain order of evaluation. The evaluation process firstly use objective evaluation methods to measure and then use subjective evaluation methods to measure. In the paper, the process is as follows: In the first place, using basic evaluation criteria-support and confidence finds the strong rules. In the next place, using lift filters and removes those incorrect, redundant, and uninteresting rules. And then there is a correlation between the association rules. In the end, using novelty obtains the rules that the users will be interested. In simple terms, the new evaluation process is support-confidence-lift-novelty.

By the following example can validate the availability of the new evaluation process.

#### 3.2 A practical example

This paper uses temperatures, precipitation, and MODIS-NDVI data of the three northeast provinces of china from 2000 to 2006 as experimental data. Besides, temporal data such as the quarters and spatial data such as the region data are introduced in to the experiment. After data pre-processing using geographical spatio-temporal methods, the table required by association rule mining was created. There are 5638 transactions in the table, in this table all the fields are represented by Boolean type. Table 3 shows part of transactions in the transactional database table. From Table 3 it can be inferred that the data of temperature, temperatures, precipitation, and MODIS-NDVI are classified to three kinds such as high, medium and low, temporal data are classified according to quarter, and spatial data are classified according to the provincial administrative divisions

NO	Q1	Q2	Q3	Q4	HLJ	JL	LN	Precip high	Precip medium	Precip low	Temp high	Temp medium	Temp low	NDVI high	NDVI medium	NDVI low
3410	0	1	0	0	0	1	0	0	0	1	0	0	1	0	0	1
3411	0	1	0	0	0	1	0	0	0	1	0	0	1	0	0	1
3412	0	0	1	0	1	0	0	0	0	1	0	0	1	0	0	1
3413	0	0	1	0	0	1	0	0	0	1	0	1	0	0	0	1
3414	0	0	1	0	0	1	0	0	0	1	0	1	0	0	1	0
3415	0	0	1	0	0	0	1	0	0	1	0	1	0	0	1	0
3416	0	0	1	0	0	0	1	0	0	1	0	1	0	0	1	0
3417	0	0	0	1	0	1	0	0	0	1	0	1	0	0	1	0
3418	0	0	1	0	0	1	0	0	0	1	0	1	0	0	1	0
3419	0	0	1	0	0	1	0	0	0	1	0	0	1	0	0	1

Table 3.the transactional database table (Q means quarter, HLJ means the area of Heilongjiang province, JL means the area of Jilin province, LN means the area of Liaoning province, Precip means Precipitation, Temp means Temperature)

Association rules are mined using the Apriori algorithm, the minimum support threshold is 0.1, and the minimum confidence threshold is 0.75. There are 36 rules in mining result. After using the lift measure, it is learned that the lift values of some strong association rules are less than 1, which indicates the possibility of the emergence of the antecedent does not lift the possibility of the emergence of the consequent, and those rules are meaningless. There are 22 rules left (the results can be seen in Table 4) after deleting the rules whose values of lift less than 1. It can be seen that the lift can control the quantity of rules, delete meaningless rules, and reduce the follow up workload.

The final step of the new evaluation process is to use novelty to measure the discovered rules. In the calculation of novelty value, the establishment of spatial knowledge base is the most important step which is the basis for the quality of evaluation rules and mining the rules of users interested. In this experiment

the rules of the spatial knowledge base have been summarized out according to the geographical knowledge of the three northeast provinces and some other studies. Table 5 is a part of rules in the spatial knowledge base.

Using the formulas shown in 2.2, we can get the novelty of every rule in Table 4 and we can see the distribution of novelty value in Figure 2.

Novelty is a value between 1 and 4. When the value is 1, it indicates that the antecedent and the consequent of discovery rule are contained in the rule of the corresponding front piece and back piece in the spatial knowledge base, and it means that the discovered rule and the rule in the spatial knowledge base are similar in a large extent; when the value is 4, it indicates that the antecedent and the consequent of discovery rule can't be found in the rule of the corresponding front piece and back piece in the spatial knowledge base, and in this case novelty

reached its upper limit; and when the value is closer to 1, the difference is smaller, and the value is closer to 4, the difference is greater.

It can be inferred from Figure 4 that the values of novelty of the association rules mostly range between 1 and 2, which indicates that the discovered rules and the rules in spatial knowledge base are very similar-in other words, the quality of these rules is high. Then, the rules whose values of novelty are relatively large can be added to the spatial knowledge base which can improve spatial knowledge base and also can provide the basis for future association rules evaluation or other study.

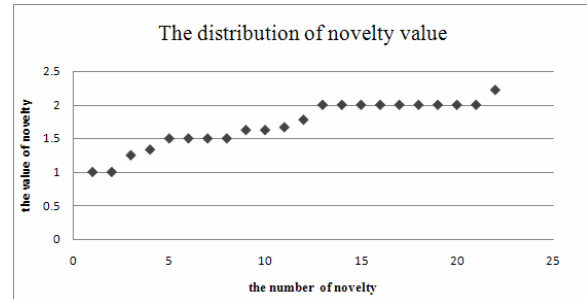


Figure 2. The distribution of novelty value

No	antecedent	consequent	support	confidence	lift
1	Q1	Precip low	0.22899972	0.9488281	1.0016142
2	Q3	Precip low	0.2607712	0.94855964	1.0013309
3	JL	Precip low	0.28839445	0.9511816	1.0040987
4	LN	Precip low	0.295371	0.9496817	1.0025153
5	Temp medium	Precip low	0.31724334	0.961154	1.0146258
6	Temp medium	NDVI medium	0.25087208	0.76006854	1.5332916
7	Temp medium	Precip low	0.22815122	0.9591756	1.0125374
8	NDVI high	Temp high	0.25520882	0.9396043	2.1746416
9	NDVI medium	Precip low	0.4723296	0.9528338	1.0058428
10	NDVI low	Precip low	0.22334307	0.95988655	1.0132879
11	Q1, NDVI medium	Precip low	0.11548977	0.951826	1.0047789
12	Q2, NDVI medium	Precip low	0.12265485	0.9517191	1.0046661
13	Q3, NDVI medium	Precip low	0.13330819	0.96125084	1.0147281
14	HLJ, Temp medium	Precip low	0.12171207	0.9562963	1.0094979
15	HLJ, NDVI medium	Precip low	0.17969266	0.9487307	1.0015115
16	JL, NDVI medium	Precip low	0.14735553	0.95363027	1.0066836
17	LN, Temp medium	Precip low	0.10153872	0.96591926	1.0196562
18	LN, NDVI medium	Precip low	0.14528142	0.95714283	1.0103915
19	Precip low, Temp medium	NDVI medium	0.24125578	0.76047546	1.5341125
20	Precip low, NDVI high	Temp high	0.23635335	0.9393031	2.1739447
21	Temp medium, NDVI medium	Precip low	0.24125578	0.96166855	1.015169
22	Temp low, NDVI low	Precip low	0.1541435	0.9583822	1.0116998

Table 4. The rules filtered by lift

NO	Association rule
1	Q2, Temp high $\Rightarrow$ NDVI high
2	Q3, Temp high $\Rightarrow$ Precip high
3	Q3, Precip high $\Rightarrow$ Temp high
4	Q3, NDVI high $\Rightarrow$ Temp high
5	HLJ, NDVI high $\Rightarrow$ Temp high
6	JL, NDVI high $\Rightarrow$ Temp high
7	LN, NDVI high $\Rightarrow$ Temp high
8	Q2, NDVI $\Rightarrow$ Temp high
9	Temp high $\Rightarrow$ Precip high
10	NDVI high $\Rightarrow$ Precip high
11	HLJ, Temp high $\Rightarrow$ Precip high
12	HLJ, Temp high $\Rightarrow$ NDVI high
13	LN, Temp high $\Rightarrow$ NDVI high
14	NDVI high $\Rightarrow$ Temp high

Table 5. The rules in spatial knowledge base

#### 4. CONCLUSION

In this paper, evaluation methods include objective evaluation methods, subjective evaluation methods and comprehensive evaluation methods are analysed and discussed. And evaluation methods evaluate the rules with the evaluation criteria.

There are mainly three objective measures: support, confidence and lift. Meanwhile, the common subjective measures are simplicity, user interest degree and novelty. The comprehensive measures are mainly the combination of objective evaluation criteria and subjective evaluation criteria. However, setting the specific weight is not easy to grasp, thus comprehensive evaluation is not very practical.

On the basis of the support-confidence evaluation system, it introduced lift and novelty evaluation indicators and formed the new evaluation process. In the paper, there were three steps in the new process. Firstly, strong association rules are extracted using basic objective evaluation criteria such as support and confidence. Secondly, incorrect, redundant, and uninteresting rules are filtered and removed using the objective evaluation criteria lift. Finally, the rules that the users would be interested are obtained using the subjective evaluation criteria novelty. In simple terms, the new evaluation process is support-confidence-lift-novelty.

The lift measure could filter out meaningless and uninteresting rules and save the time of follow-up subjective evaluation. And the improved novelty is an important indicator in the subjective evaluation methods of the association rules, which also can better measure the quality of the rules. However, the spatial knowledge base, of which the establishment is the foundation of novelty evaluation, will be influenced by the people's subjective factors. So its perfection cannot be guaranteed, which may affect the quality of the evaluation of association rules mining. Therefore, it is an issue that the improvement of the spatial knowledge base requires continuous research.

#### REFERENCE

- [1]Agrawal R, Imielinski T, Swami A, 1993. Mining association rules between sets of items in large databases. *In Proceedings of ACM SIGMOD*, pp. 207-216.
- [2]Genlin JI, 2004. Mining optimized support and interesting quantitative association rules. *Mini-Micro Systems*, 25(2), pp. 225-227.
- [3]Zhandong SU, Fucheng YOU, Bingru YANG, 2004. Comprehensive evaluation method of association rules and an practical example. *Computer Applications*, 24(10). pp. 17-20.

[4]Xiangguo Dou, 2005. Study on methods of evaluation for the association rules. Hefei University of Technology, Hefei, China. pp. 34-35.

[5] Xudong SONG, Kun ZHAI, Weidong GAO, 2007. The research of evaluation index in association rules. *Microcomputer information*, 23(4-3), pp. 174-176.

[6]Bingru Yang, 1988. A type of language field integrated algorithm used for analysis and control of complicated system. *Journal of System, Engineering and Electronics*, 9(1), pp. 66-76.

[7]Jiawei Han, Micheline Kamber, 2000. *Data Mining: Concept and Techniques*. Academic.[8]Yanxia QI, 2004. Measurement of novelty: factor of evaluation for the association Rules. *Application Research of Computers*, 21(1). pp. 17-19.

#### ACKNOWLEDGEMENTS

This research is funded jointly by National High Technology Research and Development Program of China and The Project of Basic Geographic Information Database Management System of Shandong Province. It is sincerely grateful my graduate tutor and the members of research group in geographical spatio-temporal data mining in Wuhan University and the leadership of the work unit for the help and support of this research.