

# Research on Multi-Level Association Rules Based on Geosciences Data

Dongmei Han

School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai 200433, P.R.China

Shanghai Financial Information Technology Key Research Laboratory, Shanghai 200433, P.R.China

Email: handongmei19610320@gmail.com

Yiyin Shi, Wen Wang and Yonghui Dai

School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai 200433, China

Email: shi\_yiyin@163.com, wangwen\_1010@163.com, dyh822@163.com

**Abstract**—This paper proposes a framework of multisource geo-knowledge discovery with association rules. Taking into account spatial data exist semantic fuzziness, and the conversion between qualitative concept and quantitative description is uncertain, in our study, conceptual partition algorithm and membership grade judgment algorithm based on cloud model was used. Meanwhile, there are many correlations among concepts in the field of geoscience, and the underlying correlation also need to be found with different membership grade functions, therefore, a method of multi-level association rules mining was proposed. In order to enhance frequent item set discovery efficiency, improved FP (Frequent Pattern)-Growth algorithm was presented. The algorithm was used in the empirical research on judgment of fault property at the south of Longmen Mountains in Chengdu city of China. The empirical result shows that the improved FP-Growth model acts better in frequent item-set mining.

**Index Terms**—geosciences data, multi-level association rules, cloud model

## I. INTRODUCTION

With the development of geophysical exploration, geochemical exploration, land surveying and remote sensing, a vast amount of associated data has been collected, which makes it possible for scientists to reveal the internal constitution of earth. Many scholars have contributed to the multi-source geospatial knowledge discovery include the research on geospatial data knowledge discovery and research on association rules algorithm. Fauvel, et al. (2012) researched on a spatial-spectral kernel based approach for the classification of remote-sensing images, the proposed method deals with the joint use of the spatial and the spectral information provided by the remote-sensing images [1]. Linear

This work was supported in part by National Natural Science Foundation of China (No.41174007) and Graduate Innovation Fund Program (No.CXJJ-2013-436, No.CXJJ-2013-445) of Shanghai University of Finance and Economics

Corresponding author: Yonghui Dai.

feature of remote sensing image was recognized with the Bayes classifier and genetic algorithm [2]. Quintero, et al (2012) presented FERD, a methodology aimed to automatically identify, extract and describe relevant spatial objects contained in raster spatial datasets [3]. Zheng and Zhou (2011) predicted the location information with Warrants theory [4]. Other research on geospatial data knowledge discovery such as gravity flow among gravity anomaly with the graph theory and shortest path theory, the visualization technologies used in spatial data mining as well as typical applications, seismicity of the Longmen Mountains fault zone and its vicinity before the 12 May 2008 Wenchuan Ms8.0 earthquakes, etc [5]-[7]. In addition, research on association rules algorithm main include three aspects, research on efficiency of association rules algorithm, algorithm adaptability and geospatial uncertainty, which can be described as following.

1. Research on efficiency of association rules algorithm. The Apriori algorithm was first proposed by Agrawal, he proposed the association rules based on candidate item set [8]. After this, many scholars proposed improvement based on the algorithm in order to improve the efficiency of operations. The classical algorithms are AprioriTid algorithm, AprioriHybrid algorithm, transaction reduction technique and so on [9]-[10]. In addition, the different data structures to further improve the efficiency of association rule discovery algorithm in many Chinese scholars [11]-[13].

2. Research on algorithm adaptability. Study on the applicability of the algorithm mainly included research on multi value attribute association rules, research on the multilevel association rules mining, association rules with constraints, positive and negative association rules, classification association rules, and sequence association rule. Among, association rules can be divided into two kinds of boolean type (such as Apriori algorithm) and the multi value type. Li, et al (2004) extended the original transactional database by the method of similar attribute set, making the association rules available to be applied

to the multi value attributes [14]. Fdez, et al (2009) presented a new fuzzy data-mining algorithm for extracting both fuzzy association rules and membership functions by means of a genetic learning of the membership functions and a basic method for mining fuzzy association rules [15]. Research on the multilevel association rules mining is mainly divided into two kinds of model, one is attribute-oriented induction method, the other is method based on concept tree. Wang, et al (2009) discussed the method of attribute induction threshold determination and multilayer multidimensional attribute generalization, proposed the Hastu concept climb method based concept lattice development to solve the problem that the granularity is too coarse or small to mine [16]-[17]. Manda, et al (2013) presented a data mining approach, which is called Multi-Ontology data mining at All Levels (MOAL) [18]. Hsu, et al (2004) focused on using correlation of multiple reference attribute threshold to improve the efficiency of generalization and accuracy of knowledge discovery, and gave the general path and approach of attribute induction. Knowledge discovery process with attribute oriented generalization method focus on how to control information thickness to meet the requirements of the rules discovery [19].

3. Research on geospatial uncertainty. Geospatial data bears many kinds of uncertainties, such as uncertainties caused by approximation in data sampling process and model abstract, conversion between spatial concept and spatial data and so on. Therefore, in the face of the unavoidable problems, many scholars began to make research, which can be divided into the following two categories: the uncertainty of spatial relationship and spatial reasoning. Research on the uncertainty of spatial relations is mainly based on the following theories: the theory of probability, the rough set theory, evidence theory, and cloud theory [20]-[22]. In the geometric space, uncertainty reasoning between geometry is the main research direction of spatial reasoning. Viard and Lévy (2011) reviewed a typical example of decision making under uncertainty, where uncertainty visualization methods can actually make a difference [23]. Yao and Li (2008) first defined different geometric space features, then introduced spatial relations to define that between geometric bodies, established relevant inference mechanism [24], while Justice, et al (2011) combined spatial attribute and non-spatial attribute inference mechanism [25].

Up to now, many scholars have theoretically studied on spatial knowledge discovery, constraints of current research are as follows:

Firstly, the spatial data is often of multi-source and multi-level, which is often handled through the conversion of data into raster image, then visual interpretation. The disadvantages are as follows:

(1) The method is not conducive to the discovery of new knowledge, can only played great part in verifying prior knowledge and is more difficult to find new knowledge.

(2) The method is not conducive to deal with large-scale data, can't process large-scale data due to the

adoption of visual translation, affected the efficiency of data processing.

(3) Visual interpretation is subjective, results interpretation of different specialist is difficult to be quantified and popularized.

Secondly, the spatial data are fuzzy and uncertainty, we often need such expressions as " around the top,100 meters south of landslide displacement", and the traditional method of data-discretization is based on hard division method that divides concept into either this or that, the description of the uncertainty is not exact.

Thirdly, traditional association rule-mining is only applicable to single-level data. While geospatial data can often be hierarchically processed in accordance with elevation, it is necessary to discover knowledge at different information granularity .For example, mining at coarse granularity level so as to find the common meaning of knowledge, but the information granularity is too high and may leads to over-generalization of the rules discovered. Therefore, one of the problems of the traditional single level association rules is very difficult to carry out effective mining at appropriate level to find the useful knowledge.

Finally, apriori algorithm based on candidate item sets is widely used to generate frequent item sets, which greatly influences the efficiency of the algorithm due to the need to scan the database many times while generating frequent item sets. While the FP-Growth algorithm only needs to scan database two times to generate frequent item sets due to the unique data structure (FP-Tree structure), but it repeats traversal of parent node path while constructing the FP-Tree data structure, so the FP-Tree tree structure will show explosive growth in the face of massive data, which still affects the efficiency of data mining.

## II KNOWLEDGE DISCOVERY FRAMEWORK BASED ON MULTI-SOURCE GEOSPATIAL DATA

For some existing problems mentioned above in the field of multi-source geoscience knowledge discovery, this paper puts forward a complete set of geological knowledge discovery framework based on mining association rules which is as shown in Fig. 1.

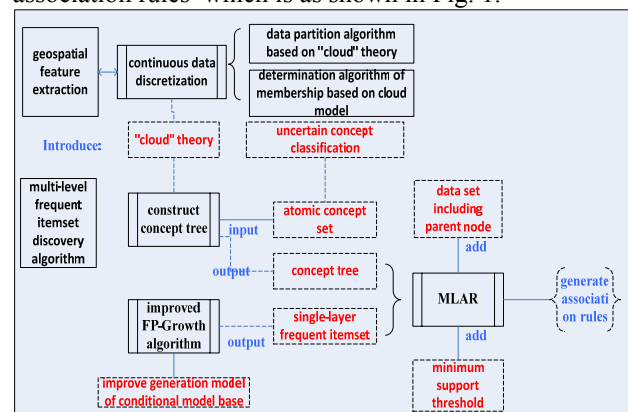


Figure 1. Geological knowledge discovery framework

First of all, characteristic data relevant to the research goal is found through feature extraction, then combined into a corresponding database, with which implicit knowledge is discovered with association rules automatically to avoid shortage visual solutions translation brings. Secondly, in the process of data preprocessing, considering the uncertainty of spatial relations, "cloud" model was introduced to partition continuous data of numerical type, and dynamically determines the membership degree in order to avoid shortage caused by hard partition problems. Finally, in order to mine in different information granularity level, this paper constructs multi-level association rules algorithm, improves the relevant frequent item sets and constructs improved FP-Growth algorithm.

*A Data Preprocessing Based on "Cloud" Theory*

1. Concept partition algorithm based on "cloud" theory

Combined with the "cloud" theory, data fitting method will be applied to concept partition according to distribution characteristics of data in this paper. According to the basic idea of data fitting and a certain rule, spatial data of any irregular distribution are mathematically transformed so as to generate a set of atomic concepts and make the distributed spatial data become the superposition of several concept of different size. The basic idea is expressed as in (1).

$$g(x) = \sum_{i=1}^n (c_i * f_i(x)) + \varepsilon \text{ and } 0 < \text{Max}(|g(x) - \sum_{i=1}^n (c_i * f_i(x))|) < \varepsilon \quad (1)$$

Where  $g(x)$  denotes distribution functions for data,  $f_i(x)$  denotes a set of atomic concepts,  $n$  denotes the number of concept superposed,  $\varepsilon$  denotes the value of deviation. The number of concept  $N$  can be controlled by a given deviation value, the less  $\varepsilon$  is, the greater  $n$  is, the greater  $\varepsilon$  is, the less  $n$  is. Due to the uncertainty of spatial data, cloud model theory is introduced in this paper. According to the cloud model theory [30], the expected function of cloud model is defined. According to ideas above, flow chart of the concept partitioning algorithm based on the cloud model constructed in this paper is shown as Fig.2.

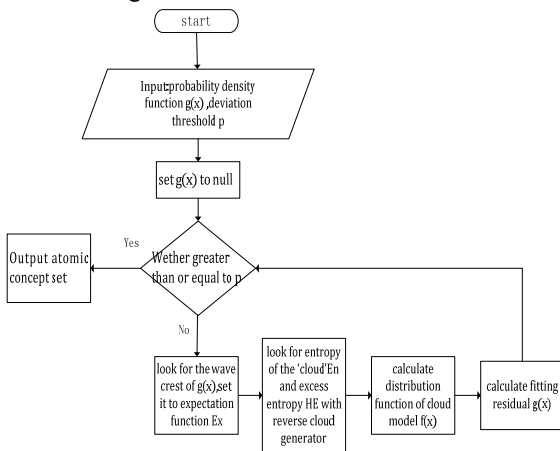


Figure 2. Flow chart of concept partition algorithm based on "cloud" theory

2. Determination algorithm of numerical data based on cloud model

After implementation of the above concepts partitioning, atomic qualitative concept set  $Ex$ ,  $En$  and  $He$  three features can be got, then which one of  $m$  concepts it belonged to can be determined, that is determination of membership degree of a numerical attribute value, with the use of  $X$  conditions and the introduction of cloud generator roulette thought, the membership degree is determined according to a certain distribution so as to realize uncertainty division of numerical data. Specific reference is as shown in Fig.3.

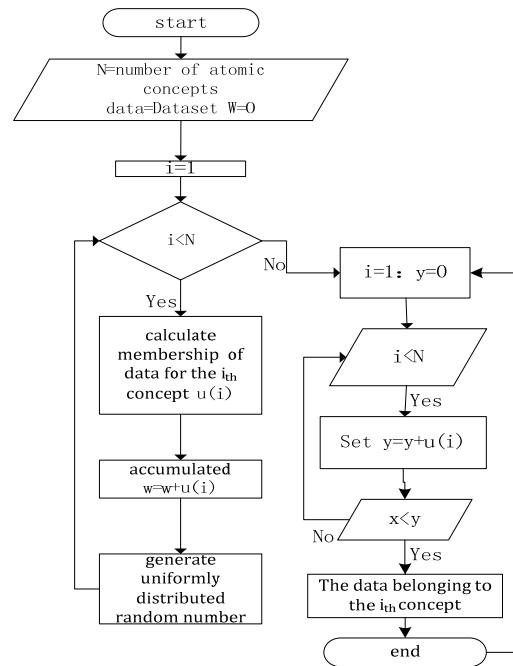


Figure 3. Program flow chart of determination algorithm of numerical data based on cloud model

*B. Frequent Item Discovery Algorithm Based on FP-Tree*

The traditional Apriori algorithm needs to produce a mass of candidate item sets in the implementation process, which will lead to huge I/O overhead. This paper argues that it can be solved by changing the structure of FP-Tree and search mode of conditional pattern base. This paper firstly added a new domain in the TP-Tree structure: proute, which is used for saving sub condition pattern base for each node, scanned database with above algorithm for generating FP-Tree to generate FP-Tree structure  $T$ , then started the implementation of FP-Growth algorithm and this paper has made the corresponding improvement. The detailed steps are as follows:

**Step 1:** Traversal frequent item header of each set in sequence from small to large, look for the position at which node first appears in the tree to find path to the root node, for every parent node, for, judge whether the Proute domain is null, if it is null, write the prefix subtree (the path from the node to the root node) into proute domain, or read the value, namely complete prefix path-the conditional pattern bases denoted as <CPB>. It can be

seen in this step that, Unlike FP-Growth algorithm, improved algorithm needs to record every node of the prefix subtree and if the prefix subtree of the node has been iterated before that it is enough to just read value without the need to traverse again so as to improve the efficiency.

**Step 2:** Look for the next node of the same name according to the link domain of the node, calculate the conditional pattern bases and denote it as  $\langle \text{CPB}, 1 \rangle$  by the same method;. Then according to the node link domain of the node to find next node of the same name, follow such steps for traversing all the node of the same name.

**Step 3:** Merge all the sub condition pattern base and call FP-Tree algorithm to construct the sub condition pattern base FP-Tree.

**Step 4:** Repeat steps (1) - (3), leaving only the root node of the first constructed FP-Tree, and then the algorithm terminates.

*C. Multi-Level Association Rules Algorithm (MLAR)*

Combined with the above discretization results gained by "cloud model" and the improved FP-Growth algorithm, Multi-Level Association Rules Algorithm (MLAR algorithm) is constructed in order to mine useful information in different information granularity level. The single level association rule can't satisfy the need of user to view knowledge in different information granularity scope, while implicit premise of multi-level association rules discovery is to establish a concept hierarchy tree for each attribute, which is mainly reflected in the following aspects. The overall flow chart of the MLAR algorithm is shown in Fig.4.

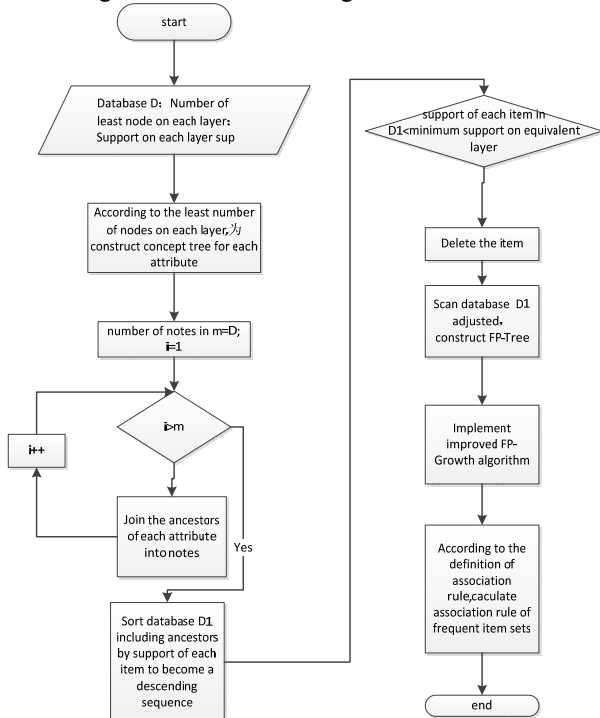


Figure 4. Flow chart of MLAR

III PREDICTION OF FAULT PROPERTY IN CHENGDU OFFICE AREA BASED ON MLAR

The fault is a kind of geological phenomenon, which makes original strata deform and fracture because of movement and interaction between plates, then lead to diastrophism of the earth's crust, Thus fault is created. Aggregation of two or more faults is called fault zone. As a kind of special geological phenomenon, fault should be made deep research on to provide important information for further understanding of plate tectonics in related areas and earthquake prediction.

In this paper, we take fault in Chengdu Office area ( $28^{\circ} \sim 32^{\circ} \text{ E}$ ,  $102^{\circ} \sim 108^{\circ} \text{ N}$ ) as research subjects (Fig.5) to study how to use other geological data to determine nature of fault by the method of multilevel association rules algorithm based on geo spatial data .

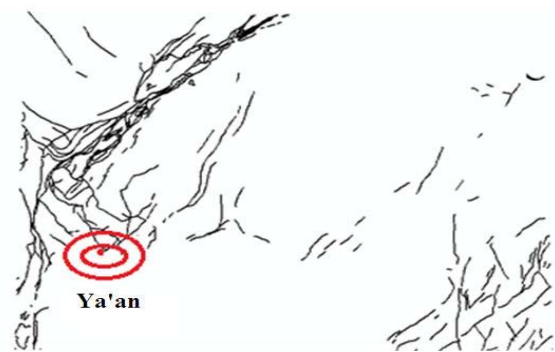


Figure 5. Fault in Chengdu Office area

*A Research Objectives and Data*

According to the known fault location and attribute information, we study the relationship between it with the other one, then predict the category the unknown fault belongs to with knowledge gained from association rules (such as thrust fault, reverse fault, compress shear faults). Thus, we collected geo spatial data in the Chengdu office area including fault data including gravity data, the aeromagnetic data, DEM elevation data. According to research objective and combined with expert knowledge, we extracted some feature possibly related to fault formation and properties based different data by different methods, which was mainly processed in Surfer 8 software and ArcGis 9.1 software. After data integration, spatial overlay operations, two-dimensional data table was finally formatted as shown in Table I.

TABLE I. PART OF FAULT DATA AFTER PROCESSED

Attribute	Instruction	Resource
ATTR	Fault type	Vector data of fault
Direction	Wether consistent with terrain trends or not	Vector data of fault
STDDE <sub>M</sub>	Standard deviation of DEM	DEM
AVGDE <sub>M</sub>	Mean value of DEM	DEM
STDMA <sub>G</sub>	Standard deviation of magnetism	Aeromagnetic data after preprocessed
AVGMA <sub>G</sub>	Mean value of magnetism	Aeromagnetic data after preprocessed

AVGGR A	Mean value of gravity	Gravity data after preprocessed
STDGR A	Standard deviation of gravity	Gravity data after preprocessed
AVGM AG0	Mean first derivative of aeromagnetic, 0 degree	First derivative data of aeromagnetic, 0 degree
STDMA G0	Standard deviation of the first derivative of aeromagnetic, 0 degree	First derivative data of aeromagnetic, 0 degree
AVGM AG90	Mean first derivative of aeromagnetic, 90 degree	First derivative data of aeromagnetic, 90 degree
STDMA G90	Standard deviation of the first derivative of aeromagnetic, 90 degree	First derivative data of aeromagnetic, 90 degree
STDMA G45	Mean first derivative of aeromagnetic, 45 degree	First derivative data of aeromagnetic, 45 degree
...	...	...

**B Implement of Multi-level Association Rules Algorithm**

MLAR (multi-level association rules) algorithm is the key algorithm of multi-level association rules mining. According to the definition of association rules: the form of association rules is  $A \Rightarrow B$ , implication,

$A \subset I, B \subset I$ , and,  $A \cap B = \phi$ , any items in B are not ancestors of A. Among, A and B are respectively called premise and conclusion of association rules  $A \Rightarrow B$ . its support is denoted as support  $support(A \Rightarrow B) = support(A \cup B)$  and  $confidence(A \Rightarrow B) = \frac{support(A \cup B)}{support(A)} \times 100\%$  as

confidence, where A and B denote both frequent item sets.

We define minimum support degree generating association rules as 20% and the minimum confidence as 60%, 73 association rules meeting the conditions were generated according to the frequent item sets. in order to judge the quality of association rules, we introduced the scoring mechanism:

**score = 40% \* support + 60% \* confidence**

Where score denotes the final score (the higher the better), so we picked the top 8 association rules, which are shown in Table II.

TABLE II. ASSOCIATION RULES FOUND

No	Succedent item	Antecedent item	Support (%)	Confidence (%)	Score
1	ATTR = general fault	AVGGRA_3 = 5 1 2 and STDGRA_2 = 6 2*	34.43	90.79	68.246
2	ATTR = reverse fault	STDDEM_3 = 1 1 2	33.22	86.18	64.996
3	ATTR = thrusting napped fault	AVGMAG0_2 = 7 2* and AVG_GRA45_2 = 19 2*	22.84	92.82	64.828
4	ATTR = reverse	AVGGRA_3 = 5 1 2 and	21.09	92.75	64.086

	fault	AVGGRA4513_3 = 11 2 3			
5	ATTR = transpressional fault	AVGDEM_3 = 2 1 2 and AVGGRA45_2 = 18 2*	12.13	97.3	63.232
6	ATTR = general fault	STDMA_3 = 3 1 2 and AVGMAG0_2 = 7 2*	14.27	93.81	62.194
7	ATTR = general fault	STDDEM_3 = 3 1 2 and AVGGRA45_3 = 18 2 3	10.27	96.81	62.194
8	ATTR = thrusting napped fault	AVGGRA_3 = 5 1 2 and STD_GRA0_3 = 26 1 1	7.54	98.55	62.146

After analysis above, it can be seen that association rules No. 1, 3, 5, 6 etc, contained frequent item sets of different levels, for example antecedent of No. 1 association rule can be explained that if the average gravity anomaly value of both sides of the fault belongs to the second concept at the third layer of attribute concept tree and the standard deviation belongs to the second concept at the second layer of attribute concept tree, the fault can be introduced as a general one, the support degree of the rule is 34.43% and confidence is 90.79%. The higher level is (i.e. the level number smaller), the more general concept represented is, that is the coarser information granularity is, so it can be seen that there are both the concept of third layer and that of second layer in the two antecedent conditions of the rule No. 1, so the algorithm successfully realized discovery of the multi level association rules at various levels of concept tree. Compared to the original single association rules that can find only relationship among leaf nodes, multilayer association rules can discover knowledge more effectively.

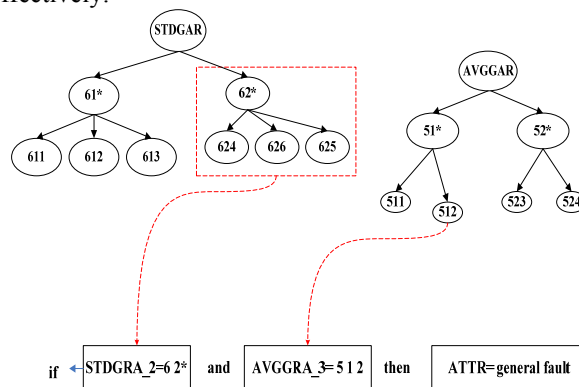


Figure 6. Diagram of result of MLAR

IV RESULT VALIDATION

At the junction region of Chengdu palace' southwest and Sichuan basin, after a series of geological survey during 90's of 20th century, a series of faults was found in the region such as the Peace River line fault mentioned in literature, as well as Blackwater River fault zone. The fault zone along Anning River refers to the fault north TianWan in Shimian County, south through Mianning, Xichang, Dechang, Miyi to Jinsha River in Panzhuhua, north-south trending faults, and the Blackwater River



fault zone refers to the fault from Haining Puig to Yuexi line, with total length of 130km, which also presents North South direction. The approximate position of fault zone along Anning River and that along the Blackwater river is as shown in A area as shown in Fig.7. The projection has a geographic coordinate system associated with it.

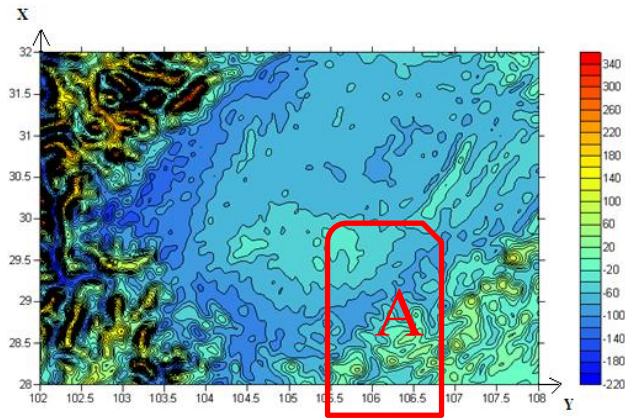


Figure 7. Diagram of fault zone along Anning River and Blackwater river

According to the related fault evidences, micro topography, mutation belt or seismic activity and other related evidence. It is indicated that the regional faults have features own by reverse faults, Further, the fracture zone presents certain extrusion characteristics. In this section, this paper will verify the conclusions mentioned above based on the results of association rule mining. For faults in the area mostly showed a north-south orientation, I first use the related function of ArcGis to draw 50 North-South fault line in Region A, the 50, for simplicity, I assume that the fault lines are parallel. As is shown in the black box in Fig.8

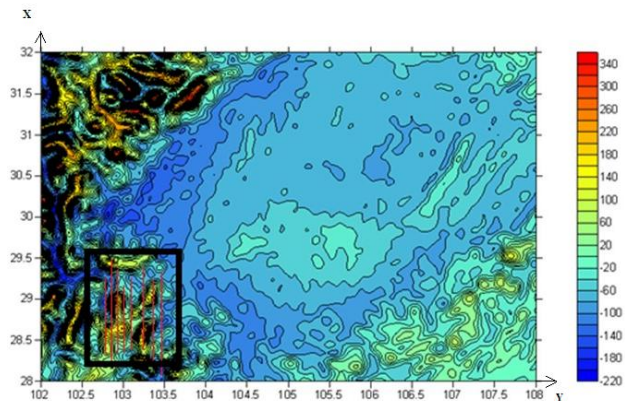


Figure 8. 50 simulated fault

According to using feature extraction steps mentioned above, we extract 31 relevant attributes of the simulated 50 fracture, determine that the objective of the study was to determine the nature of the fault. Organize relational tables used for data mining according to 50 simulated faults. Using 8 association rules as shown as Table III to process the data, finally obtain the results as shown in Table III:

TABLE III. FAULT TYPE JUDGED ACCORING ASSOCIATION RULE

No.	Fault type	25	Reverse fault
1	NULL	26	Reverse fault
2	Reverse fault	27	Reverse fault
3	Reverse fault	28	Reverse fault
4	NULL	29	Transpressional fault
5	Reverse fault	30	Transpressional fault
6	Reverse fault	31	Transpressional fault
7	Reverse fault	32	Transpressional fault
8	Reverse fault	33	Transpressional fault
9	Reverse fault	34	Reverse fault
10	Reverse fault	35	Reverse fault
11	Transpressional fault	36	Reverse fault
12	Transpressional fault	37	Reverse fault
13	Reverse fault	38	Reverse fault
14	Reverse fault	39	Reverse fault
15	NULL	40	Reverse fault
16	NULL	41	Transpressional fault
17	General fault	42	Reverse fault
18	General fault	43	Reverse fault
19	Reverse fault	44	Reverse fault
20	Reverse fault	45	Reverse fault
21	Reverse fault	46	NULL
22	Reverse fault	47	NULL
23	Reverse fault	48	Transpressional fault
24	reverse fault	49	general fault
		50	general fault

Among 50 fault used for prediction, the number of thrust is 31, accounting for 62%, transpressional faults are 9, accounted for 18%; the general fault 5, accounting for 10%; fault unable to define 5, accounting for more than 10%.Fault unable to predict is fault that does not meet any one of the 8 association rules above. On specific information, please refer to Fig.9.

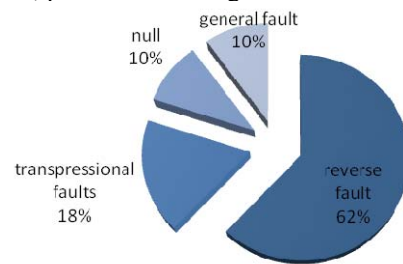


Figure 9. Prediction result of fault property

It can be seen from the results that, in Region A, reverse fault is in the majority, followed by transpressional faults that accounted for 18%, which is consistent with the description of nature of fault in existed literature .While general faults and unpredictable faults appear due to the emergence of a certain deviation .After all, the simulated fault is different from that in reality in location, direction and so on. But the sum of the two fault accounted for only 20% within the

acceptable range. The results confirm the prediction of other scholars on the nature of regional fault judgment from the viewpoint of probability, which confirms the reliability based association rule conclusion.

#### V. CONCLUSION

This paper proposes a framework of multi-source geospatial knowledge-discovery with focus on association rules, which is aimed primarily at multi-source geospatial knowledge discovery problems. The empirical research shows that the improved FP-Growth model acts better in frequent item-set mining. But for time, we didn't research into topology relations and location relations, which are focus of attention. A further improvement is expected.

#### ACKNOWLEDGEMENT

This work was supported in part by National Natural Science Foundation of China (No.41174007) and Graduate Innovation Fund Program (No.CXJJ-2013-436, No.CXJJ-2013-445) of Shanghai University of Finance and Economics.

#### REFERENCES

- [1] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "A spatial-spectral kernel-based approach for the classification of remote-sensing images," *Pattern Recognition*, vol. 45 (1), pp.381-392, 2012.
- [2] X. Wang, and R. Niu, "Spatial forecast of landslides in Three Gorges based on spatial data mining," *Sensors*, vol. 9 (3), pp.2035-2061, 2009.
- [3] R. Quintero, et al, "An ontology-driven approach for the extraction and description of geographic objects contained in raster spatial data," *Expert Systems with Applications*, vol. 39 (10), pp.9008-9020, 2012.
- [4] Y. Zheng, and X. Zhou, *Computing with spatial trajectories*, New York: Springer, 2011.
- [5] Q. Wang, and V. Megalooikonomou, "A performance evaluation framework for association mining in spatial data," *Journal of intelligent information systems*, vol. 35 (3), pp.465-494, 2010.
- [6] Z. Jia, and Y. Liu, "Survey of Visual Spatial Data Mining Research," *Geospatial Information*, vol. 7, pp.10-15, 2009.
- [7] Z. W. Zhang, et al, "Seismicity and tectonic stress of the Longmenshan fault zone before 2008 Wechuan Ms8.0 earthquake," *Acta Seismologica sinica*, vol. 31, pp.117-127, 235, 2009.
- [8] R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," *ACM SIGMOD Record*, vol. 22, pp.207-216, 1993.
- [9] R. Agrawal, and R. Srikant, "Fast Algorithms for Mining Association Rules," *Proceeding of the 20th International Conference on Very large databases*, vol. 1215, pp.487-499, 1994.
- [10] A. M. Bica, "Apriori error estimation in terms of the third derivative for the method of successive approximations applied to ODE'S," *Journal of Applied Mathematics and Computing*, vol. 22, pp.199-212, 2006.
- [11] J. Han, et al, "Mining frequent patterns without candidate generation: A frequent-pattern tree approach," *Data Mining and Knowledge Discovery*, vol. 8, pp.53-87, 2004.
- [12] H. Yu, et al, "An Improved Apriori Algorithm Based On the Boolean Matrix and Hadoop," *Procedia Engineering*, vol. 15, pp.1827-1831, 2011
- [13] Q. Liu, "Research on Distributed association rules mining algorithm and its applications," Zhejiang Normal University, 2011.
- [14] P. Li, J. Chen, and F. Bian, "A developed algorithm of apriori based on association analysis," *Geo-Spatial Information Science*, vol. 7, pp.108-102, 2004.
- [15] J. Alcalá-Fdez, et al, "Learning the membership function contexts for mining fuzzy association rules by using genetic algorithms," *Fuzzy Sets and Systems*, vol. 160, pp. 905-921, 2009.
- [16] D. X. Wang, X. G. Hu, X. P. Liu, and D. M. Huang, "Attribute-Oriented Algorithm Based on Extended Concept Lattice," *Journal of Shanghai Jiaotong University*, vol. 43, pp.476-479. 2009.
- [17] D. X. Wang, X. G. Hu, and X. P. Liu, "Research on attribute-oriented induction based on quantitative extended concept lattice and multi-granule association rule mining," *Journal of Systems Engineering*, vol. 24, pp.54-61, 2009.
- [18] P. Manda, F. McCarthy, and S. M. Bridges,, "Interestingness measures and strategies for mining multi-ontology multi-level association rules from gene ontology annotations for the discovery of new GO relationships," *Journal of biomedical informatics*, vol. 46, pp.849-856, 2013.
- [19] C. C. Hsu, "Extending attribute-oriented induction algorithm for major values and numeric values," *Expert Systems with Applications*, vol. 27, pp.187-202, 2004.
- [20] X. Pan, et al, "A variable precision rough set approach to the remote sensing land use/cover classification," *Computers & Geosciences*, vol. 36, pp.1466-1473, 2010.
- [21] R. M. Durão, et al, "Assessing spatial uncertainty of the Portuguese fire risk through direct sequential simulation," *Ecological Modelling*, vol. 221, pp.27-33, 2010.
- [22] J. Mennis, and D. Guo, "Spatial data mining and geographic knowledge discovery—An introduction," *Computers, Environment and Urban Systems*, vol. 33, pp 403-408, 2009.
- [23] T. Viard, G. Caumon, and B. Lévy, "Adjacent versus coincident representations of geospatial uncertainty: Which promote better decisions?," *Computers & Geosciences*, vol 37, pp. 511-520, 2011.
- [24] J. Yao, and S. Li, "A Study of DEM Fractal Simulation and Spatial Uncertainty Analysis," *Journal of Henan University (Natural Science)*, vol. 6, pp.589-592, 2008.
- [25] C. O. Justice, B. Ramachandran, and M. J. Abrams, "Land remote sensing and global environmental change:NASA's Earth Observing System and the science of ASTER and MODIS". New York, NY: Springer-Verlag, pp.873.2011.
- [26] C. Dan, and Marinescu, "Cloud Computing: Applications and Paradigms," *Cloud Computing*, pp.99-130, 2013.
- [27] H. Z. Hu, W. H. Dai, Y. H. Dai, H. S. P. R. Zhang, and X. F. Liu, "Flexible Business Support System for Stock Index Futures Transaction," *Journal of Software*, vol. 7, pp.2630-2639, 2012.
- [28] D. M. Han, Y. H. Dai, and Z. Y. Zhang, "Early Warning and Monitoring System of the Economic Situation in Real Estate Market," *Journal of Software*, vol. 8, pp.2064-2071, 2013.

**Dongmei Han** is full Professor and Doctoral Advisor at the School of Information Management and Engineering, Shanghai University of Finance and Economics, China. She is also a researcher at Shanghai Financial information Technology Key Laboratory. She received her Ph.D. in 2002 from Quantitative Economics in Jilin University, China. Currently, her research interests include finance engineering, data mining and decision support system.

**Yiyin Shi** received his master degree in management science and technology from Shanghai University of Finance and Economics, Shanghai, China. His research interests include data mining and decision support system.

**Wen Wang** received her master degree in management science and technology from Shandong University of Science and Technology, Qingdao, China. She is current a Ph.D. student at the School of Information Management and Engineering, Shanghai University of Finance and Economics, China. Her research interests include data mining and decision support system.

**Yonghui Dai** is current a Ph.D. student at the School of Information Management and Engineering, Shanghai University of Finance and Economics, China. He received his Master degree in Software Engineering in 2006 from Fudan University, China. His research interests include data mining, behavioral finance and mobile information technology integration.