# MINING METALLOGENIC ASSOCIATION RULES COMBINING CLOUD MODEL WITH APRIORI ALGORITHM

*Cui Ying*[a], Binbin He[a]， Jianhua Chen[ab], Zhonghai He[a], Liu yue[b]*

a. Institute of Geo-Spatial Information Science and Technology, University of Electronic Science and Technology of China, Chengdu 611731, China
b. College of Information Engineering, Chengdu University of Technology, Chengdu 611731, China
c. College of Geology Science, Chengdu University of Technology, Chengdu 611731, China
cycherish@163.com, binbinhe@uestc.edu.cn, chjh3@163.com, hzh112@126.com,liuyx314@163.com

## 1. INTRODUCTION

Spatial data mining refers to extracting and "mining" the hidden, implicit, valid, novel and interesting spatial or non-spatial patterns or rules from large-amount, incomplete, noisy, fuzzy, random, and practical spatial databases [2]. However, spatial data mining research is currently focused on the efficiency of mining algorithm itself, while there are few studies in practical application. Association rules mining is an important part in spatial data mining. Therefore, mining metallogenic association rules will be worthwhile and provide some reference for evaluation of mineral resources. The study of spatial data uncertainty is particularly important [1, 2]. In the data preprocessing, use the cloud model to get uncertain transition between qualitative and quantitative in data so that it can reduces the data uncertainties impact. We developed the Apriori algorithm which can deal with relational table directly so that relational table need not be transformed to binary transaction. Finally, combining the cloud model based generation method with Apriori algorithm [4] for mining association rules using the spatial geological data of Eastern Kunlun orogenic belt shows its efficiency and effectiveness.

## 2. ATTRIBUTE GENERALIZATION BASED ON CLOUD MODEL

Cloud model is a model of the uncertain transition between a linguistic term of a qualitative concept and its numerical representation. In short, it is a model of the uncertain transition between qualitatives and quantitatives [3]. The unique definition of cloud is that only with the expectation Ex, entropy En and ultra-entropy He three values can be outlined by the tens of thousands of cloud droplets formed cloud, said the language of the qualitative value of the ambiguity and randomness fully integrated together to achieve a soft discretization of data. The normal compatibility clouds are most useful in representing linguisticatoms and the MEC [3, 7] (mathematical expected curve) of the normal compatibility cloud to a linguisticatom A is:

$$MEC_A(x) = \exp[\frac{(x-Ex)^2}{2En^2}]$$

Firstly, based on the above definition, this experiment chooses the quantitative property iron geochemical anomaly as discrete objects. Secondly concept definition and description, Iron geochemical anomaly domain is [0, 10], be divided into three concepts: Iron geochemical anomaly high, Iron geochemical anomaly middle, and Iron geochemical anomaly low atoms, each linguistic variable have some overlaps in Fig1. Experts give the Ex, En, He in the cloud model to define linguistic atoms. Thirdly, input Iron geochemical anomaly numerical data to achieve quantitative conversion and output results as text.

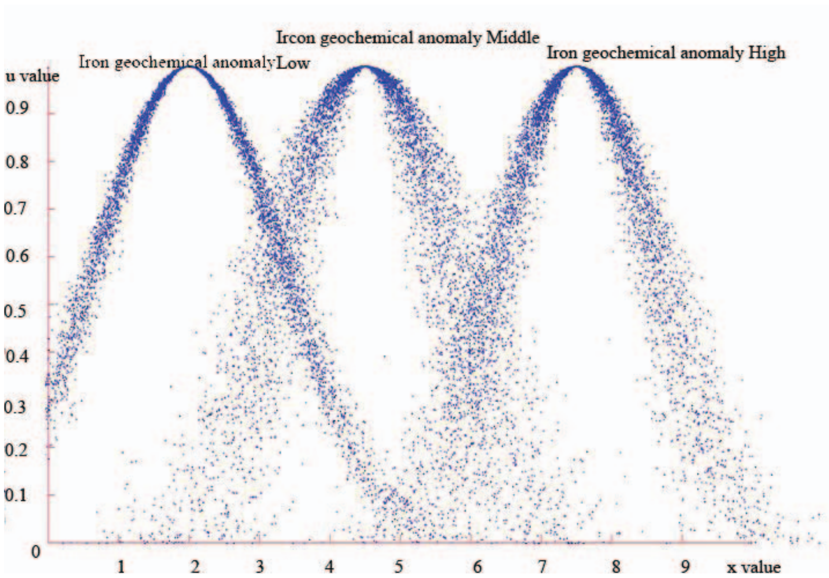


Fig.1 Linguistic atoms for the linguistic variable "Iron geochemical anomaly"

## 3. MINING METALLOGENIC ASSOCIATION RULES

Apriori algorithm was first proposed by Agrawal et al in 1993[4]. Association rule mining finds interesting association or correlation relationships among a large set of data items [4, 6]. However, Apriori algorithm is only used for mining association rules among one-dimensional binary data. We developed the Apriori algorithm and it can deal with relational table directly so that relational table need not be transformed to binary transaction. The problem of discovering all association rules can be decomposed into two subproblems. Firstly, find all sets of items that have transaction support above minimum support. The support for an itemsets is the number of transactions that contain the itemset. Itemsets with minimum support are called large itemsets, and all others small itemsets. Secondly, use the large itemsets to generate the desired rules. For every such subset a, output a rule of the form a => (l-a) if the ratio of support (l) to support (a) is at least minconf. We use support, confidence and lift on quality evaluation of association rules.

$$\text{Support}(A \rightarrow B) = P(A \cup B)$$

$$\text{Confidence}(A \rightarrow B) = P(B|A)$$

$$\text{Lift}(A \rightarrow B) = P(B)/P(AUB)$$

In order to verify the feasibility and effectiveness of our uncertain spatial data mining model, we conducted experiment using geospatial data of eastern Kunlun orogenic belt in China, which contains geoscientific data and geochemical data. After extracting, the task relevant data are 481 records. It has seven attributes, the regional fault trend, ore-bearing rocks, wall rock alteration, and ore body status, the size of deposits, iron anomaly, and genetic type of deposit. The task relevant data are organized as a relational table.

Since iron geochemical anomalies are numerical, it is impossible to discover strong association rules at the primary level directly. Therefore, we define linguistic atoms for attribute generalization which has been done in the previous step. The generalized database is mined with the minimum support 7% and minimum confidence 80%. We will "contact metasomatism hydrothermal iron ore" as the rule consequent and the other properties as the rule antecedent. There are many rules in the data mining processing, and they can be described in the form of production rules, such as table 1:

Table.1 Metallogenic association rules

| Rule No. | Rules | Confidence | Support | Lift |
|---|---|---|---|---|
| 1 | Iron geochemical anomaly middle ∧Mine has always been mainly oblique structure; faults, north-south; geo-mining areas have a large-scale thrust faults, the north section of the dumping, dip 50 ° -70 °∧(ore body status)lentoid, size of deposits(middle)==> contact metasomatism hydrothermal iron ore. | 0.8 | 0.0748 | 1.619 |
| 2 | Ore production in the schist, marble and monzogranite and granite porphyry contact zone ∧ East-west fault ∧ size of deposits(small) ∧ Iron geochemical anomaly middle ==> hydrothermal iron ore. | 0.875 | 0.090 | 30.24 |
| 3 | Ore production in the schist, marble and monzogranite zone ∧ East-west fault ∧ size of deposits(small) ∧ Iron anomaly middle ==> hydrothermal iron ore. | 1 | 0.103 | 34.56 |
| 4 | Deposits produced in the rock mass outside the contact zone of chlorite schist ∧ size of deposits(small) ∧ NW-trending fault ∧ Iron anomaly middle==> contact metasomatism hydrothermal iron ore. | 1 | 0.090 | 3.646 |

## 4. CONCLUSION

The experiments show that the model of uncertainty-based metallogenic association rules mining is efficient, the whole process is efficient. At the same time, combining cloud model and Aprirori algorithm mining association rules in the geological mineralization can provide some reference for evaluation of mineral resources. However, in the experiments spatial relationships were not discussed, this is where the future experiments need to be improved.

## REFERENCES


[1] Binbin He, Cuihua Chen, "Spatial Data Mining with Uncertainty," *Proceedings of SPIE-The International Society for Optical Engineering*, v 6787, 2007.

[2] Binbin He, Fang Tao, and Zhida Guo, "Uncertainty in spatial data miningm," *Proceedings of 2004 International Conference on Machine Learning and Cybernetics*, v2, pp.1152-1156, 2004.

[3] Deyi Li, Kaichang Di, Deren Li, and Xuemei Shi, "Mining Association Rules with Linguistic Cloud Model's," *Journal of Software*, pp. 143-158, 200,11(2).

[4] Rakesh Agrawal, Ramakrishnan Srikant, "Fast algorithms for mining association rules in large databases," *In: Proceedings of the 20th International Conference on Very large Data Bases,* pp.478-499, 1994.

[5]Michael Chau, Reynold Cheng, Ben Kao, and Jackey Ng, "Uncertain Data Mining: An Example in Clustering Location Data," *In Proc. Workshop on the Sciences of the Artificial*, Hualien, Taiwan, 2005.

[6] Han J, "Data mining techniques," *In: Proceedings of the ACM 2S IGMOD' 96 Conference on Tutorial,* June 1996.

[7] Kaichang Di, DerenLi, KDD and its application and extension in GIS, *In: Proceedings of the 2nd Annul Meeting of Chinese Association of GIS,* Beijing, 1996.