Research on Association Rules Reasoning and Application of Geosciences Data Based on Ameliorated Trapezoidal Cloud Transformation

Xu Jing

Research on Association Rules Reasoning and Application of Geosciences Data Based on Ameliorated Trapezoidal Cloud Transformation

Xu Jing

School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai, 200433, P.R.China
13491299@qq.com

Abstract. This paper proposes an association rules reasoning model based on ameliorated trapezoidal cloud transformation. It is aimed primarily at complexity and randomness geosciences data bears. The traditional trapezoidal cloud transformation is improved in order to avoid lack of data mutation information and to finish reasonable and sensitive exchange from qualification to quantification. A set of attributes for simulating faults extraction algorithm is designed, which breaks through limitations of traditional visual interpretation and ensures an effectiveness and completeness of test data. Multi-Level Association Rules (MLAR) model [1] is also adopted to reason and predict unknown faults and fault properties in Chengdu Office zone. The result shows that the MLAR algorithm enhanced an association mining between fault types with their classified attributes.

Keywords: Trapezoidal Cloud, MLAR algorithm, Chengdu Office

1 Introduction

As a series of systems deep geochemical exploration, space remote sensing and geodetic survey, were established and improved. Vast amounts of data have been gradually accumulated in the area of earth science, and these data need to be made use of by scientists. Association rules is a data mining method. It was first put forward by Agrawal [1] in 1993. Now it has become one of the most widely used algorithms in the field of data mining. Li [2] has proposed a new method for conversion-cloud information instead of "hard" division which divides a cloud environment into several quantitative concepts based on cloud model considering fully fuzziness and randomness of data. Subsequently, the cloud conversion has been used in many fields. Hu [3] proposes a new way of figuring out the weight of land evaluation factors by mapping qualitative linguistic words into a fine-changeable cloud drops and translating the uncertain factor conditions into quantitative values with the uncertain illation based on cloud model. Yu [4] attempts to evaluate the economics of wind power projects based on the cloud model. Han [5] imports the conceptual partition algorithm based on cloud model to finish the exchange from qualification to quantification. Expectation of some concept is not just a numeric value but in a sequence, so trapezoidal cloud makes description more accord with contiguous data. Wang [6] uses trapezoidal cloud model to advance the concepts of division, and transforms qualitative data in to quantitative conception which is proved to be effective. Complexity of geosciences data behaves as not only fuzziness and uncertainty but also large amount in quantities and global continuity.

In this paper, a traditional trapezoidal cloud transformation is described in order to avoid a lack of information about data mutation and to carry out a reasonable and sensitive exchange from qualification to quantification.

2 Construction of Association Rules Mining Model Based on Ameliorated Trapezoidal Cloud Transformation

2.1 Concept Partition Algorithm Based on Ameliorated Trapezoidal Cloud Transformation

According to the basic idea of data fitting and a certain rule [5], 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 in Eq. (1).

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

Geosciences data bears global continuity. As the concept is described, its expectation is not just a numeric value but in a sequence, so trapezoidal cloud makes description more accord with the features of geoscience data. In this paper, trapezoidal cloud is adopted to assist concept division, i.e., dividing a concept into atomic concepts in a number field. Fig. 1 shows the structure of trapezoidal cloud.

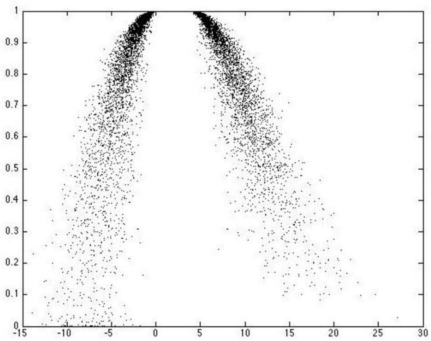

Fig. 1. Diagram of structure of trapezoidal cloud

The function Find_En is used to search for entropy and hyper entropy of cloud droplets. The idea of backward cloud generator [2] is applied during the search.

2.2 Association Rules Mining Algorithm Based on Multi-Level Association Rules Algorithm (MLAR)

Association rules is an important branch in the research area of data mining. It is aimed at mining correlations hidden behind massive data. As a large amount of data is collected and stored, it is increasingly demanded by scientists to discover knowledge from the data.

3 Prediction Model of Fault Property in Chengdu Office Area

In this paper, we will take Chengdu Office area (28°~32° E, 102°~108° N) as research zone to determine and predict nature, position and scale of simulated faults by the model of multilevel association rules algorithm based on improved trapezium-cloud model according the geophysical information 1:1,000,000 geologic map of this area conveys.

3.1 Data Preparation and Attribute Extraction Simulation

In this paper, we take fault in Chengdu Office area (28°~32° E, 102°~108° N) as the research subject in processing known fault data. Fig. 2 presents the spatial data this work used.

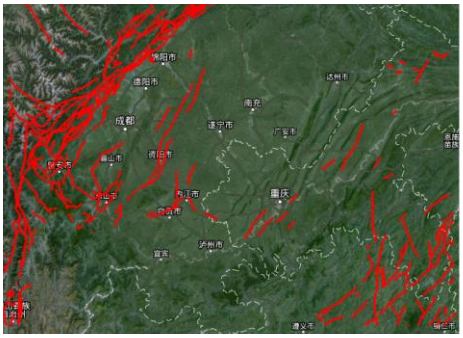

Fig. 2. Satellite image of known faults in Chengdu Office zone

Current findings in the field of earth science are built on the fusion of previous space data. According to prior knowledge of experts, subjective judgments on new knowledge are made by using visual interpretation, field measurements and other methods. Rare automated discoveries on new knowledge of earth science are achieved by the degree of inference through association rules, the methods previous adopted are difficult and inefficient.

3.2 Prediction of Faults Property in Chengdu Office Area Based on Ameliorated Trapezoidal Cloud Transformation

According to the data distribution curve of attributes of known and simulated faults (the blue curve in the Fig. 3), the threshold value of error is set at 0.5.It is defined as the input item, when the respective discretization of the number attribute of the 32 attributes in the table is implemented. After that, the trapezium-cloud concept of every attribute is built.

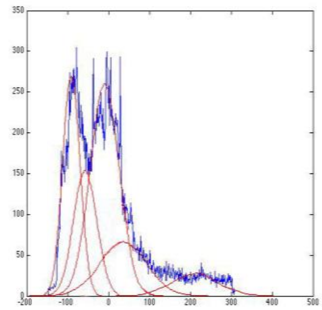

Fig. 3. Line graph of normal distribution fitting data distribution

3.3 Result Validation

In order to judge the quality of association rules, we introduced the scoring mechanism [5]: Score = 40%*support + 60%*confidence and the higher score the better quality on simulated faults. Table 1 presents the top 12 association rules used for the experiment on disposal of simulated faults.

Table 1. Association rules

No	Succeeding item	Antecedent item	Support (%)	Confidence (%)	Score
1	ATTR =F	STDMAG_3 =8 2 3 and AVGMAG0_2 = 5 3*	31.645	95.82	70.15
2	ATTR =F	AVGDEM_3=5 2 1 and AVGMAG_2=2 9 3 and AVGGRA_3=6 3*	27.34	97.89	69.67
3	ATTR = C	AVGGRA_3 = 10 1 3 and STD_GRA0_3 = 21 1 2	26.09	98.54	69.56
4	ATTR =C	AVGMAG0_2 = 5 1* and AVG_GRA45_2 = 12 3*	29.195	94.62	68.45
5	ATTR =D	AVGDEM_3 = 3 1 3 and AVGGRA45_2 = 21 3*	24.01	97.31	67.99
6	ATTR = F	STDDEM_3 = 4 1 3 and AVGGRA45_3 = 15 1 6	28.11	93.81	67.53
7	ATTR = A	AVRGRA_3=10 1 3 and AVG_MAG0_2=6 1*	35.87	88.12	67.22
8	ATTR = E	AVRGRA_3=10 1 3 and AVG_MAG=6 3* and	26.35	92.75	66.19

		var_gra451=16 2 1			
9	ATTR =E	AVGGRA_3 = 3 2 4 and AVGGRA4513_3 = 9 3 2	31.42	89.17	66.07
10	ATTR = F	AVRGRA_3=10 1 3 and AVG_MAG0_2=7 2* and STDDEM_3 = 5 2 3	35.46	85.66	65.58
11	ATTR = E	STDDEM_3 = 3 1 3	40.17	82.32	65.46
12	ATTR =C	AVGGRA_3 = 6 2 3 and STD_GRA0_3 = 20 2 1	58.705	69.73	65.32

Table 2 shows the simulation results. When the length of 40 steps (about 60 kilometers) is chose as a parameter, 42444 simulated faults in each angel were covered. After successive adjustments, it is inferred that 2305 faults are in a north-east direction（35-50°）and most of them are in 40°; 295 faults are in north-south direction; 354 faults are in a north-west direction (270°-360°).

Table 2. Fracture property and direction proportion

Fracture direction	Direction proportion	Fracture property	Property proportion
NE trend （30°-50°）	78%	general fault	32.46%
		klippe	1.36%
		thrusting-nappe fault	42.16%
		transpressional fault	17%
		general reverse fault	7.48%
NS trend （0°）	10%	general fault	59.32%
		transpressional fault	40.68%
NW trend （270°-360°）	12%	thrusting – nappe fault	14.29%
		general fault	85.71%

It is important to note that unclassified faults mean that they do not conform to any one of the 12 association rules above. The superposition of the part verified by the simulated faults and the satellite image of this area is shown in Fig. 4.

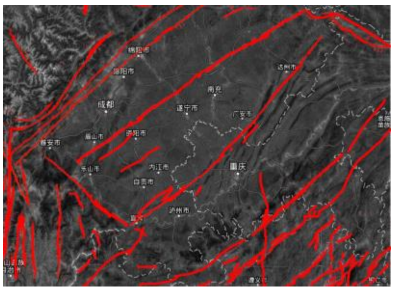

Fig. 4. Satellite image of simulated faults verified

4 Conclusion

This paper proposed the association rules reasoning model based on ameliorated trapezoidal cloud transformation, which is aimed primarily at complexity and randomness geosciences data bears. The traditional trapezoidal cloud transformation is improved in order to avoid lack of data mutation information and to finish reasonable and sensitive exchange from qualification to quantification. The attributes of simulated faults extraction algorithm was designed to overcome the limitations of traditional visual interpretation to ensure the effectiveness and completeness of the test data. MLAR model was adopted to reason and predict the unknown faults and fault property in Chengdu office zone. The result supports the judgements other academics made for the faults and their attributes in probability perspective, further explains it acts better in association mining between fault types and its attribute data automatically, through which the model's reliability has been testified.

Acknowledgments. Our appreciation goes to Qin Zheng, Wen Wang, Dongmei Han for their valuable comments to improve the quality of the paper.

References

1. Agarwal, R., Imielinski, T., Swami, A.: Mining Association Rules Between Sets of Items in Large Databases. In: Proceedings of The ACM SIGMOD Conference on Management of Data, pp. 207-2l6 (l993)
2. Li, D., Meng, H., Shi, X.: Cloud model and cloud model generator. Computer Research and Development 32 (6), 15–20 (1995)
3. Hu, S., Li, D., Liu, Y., Li, D.: Mining weights of land evaluation factors based on cloud model and correlation analysis. Geo-spatial Information Science 10(3), 218–222 (2007)
4. Fang, F., Yu, A.: The Economic Evaluation of the Wind Power Projects Based on the Cloud Model. Advances in Intelligent and Soft Computing 163, 443–448 (2013)
5. Han, D., Shi, Y., Wang, W., Dai, D.: Research on Multi-Level Association Rules Based on Geosciences Data. Journal of Software 8(12), 3269–3276 (2013)
6. Wang, Z.: Application of Cloud Theory in Association Rules. I.J. Information Technology and Computer Science 3, 36–42 (2011)
7. Tang, R., et a1.: Active faults and earthquakes in Sichuan Province. Beijing: Earthquake Press (1993)
8. Wu, W.: 1:1000000 Digital Geomor Phologic Mapping and Geo Morphology Database in China. Southwest University, pp. 23–26 (2006)
9. Huang, B., et a1.: Geotectonic and China Regional formations. Beijing: Geology Publishing House, pp. 119–254 (1996)
10. Du, F., et a1.: Interseismic deformation across the Longmenshan fault zone before the 2008 Ms8.0 W enchuan earthquake. Chinese Journal of Geophysics 52(11), 2729–2738 (2009)
11. Xu, Z., Liu, Y.: Preliminary research on visual interpreting of Chengdu Plain. Geological Review 28(5), 439–445 (1982)
12. Hu, J., Zhao, J.: Relation of deep geophysics fields and rupture construct in Yangtse Gorges Chongqing Reservoir Area 27(3), 49–54 (2005)
13. Cheng, J.: Digital Earth Introduction. Beijing: Science Press, pp. 102–145 (2000)