

An Apriori Based Algorithm Associated Point Line Pattern Applied in Seismic Spatial Data

Yuan Zhou and Lianxiong Gao*

Yunnan Land and resource vocational college in Kunming city Yunnan province China

*Corresponding author

Abstract—earthquake data in spatial database include two typical types point and line feature, this paper propose an algorithm based on apriori for analysis spatial association pattern during point and line feature, and description algorithm process. since the algorithm is a probability-based mining algorithm, it apply in spatial database of seismic spatial database of Yunnan province china for analysis correlation of point and line pattern. The algorithm have certain significance for analysis spatial probabilistic and spatial reasoning for seismic spatial data.

Keywords—spatial datamining ; seismic spatial data; probalistic

I. INTRODUCTION

Earthquake mechanisms and relevant principles suggest that the occurrence of the earthquake fault and geological activity showed great spatial correlation, which occurred mainly in the location of the earthquake fault zone adjacent geological activity or related areas. In spatial database, Geological active fault abstract as line feature, and earthquake source location abstract as point feature. *apriori* algorithm is association mining rule apply in database, and it extends to spatial database propose co-location algorithm. However, co-location algorithm is a common spatial data mining for point feature pattern and cannot apply in point and line features for seismic data[1][2].

Seismic data in spatial include two typical types, point and line, in this paper we propose an algorithm based on *apriori* algorithm for analysis spatial association rule between point and line, description the process of algorithm and experiment algorithm on seismic spatial database of Yunnan province. Use this algorithm reveal association rule between point and line of seismic data and give probability relation of them. It have certain significance for analysis spatial probabilistic and spatial reasoning for seismic spatial database.

II. SEISMIC SPATIAL DATA

Seismic spatial data has complex, multi-source, multi-dimensional and other characteristics, all seismic can abstract as point, line and polygon. Usually in spatial data, the activity fault zone is abstract as line feature and earthquake source location abstract as point feature[3][4].

A. Point Feature

The most common record of seismic data is earthquake catalog data, it recorded the earthquake source location and earthquake magnitude, earthquake happened time, earthquake depth. It recorded the earthquake source location and earthquake magnitude, earthquake happened time, and earthquake depth etc.

TABLE I. EARTHQUAKE CATALOG DATA.

date	Time	longitud e	latitu de	magnitu de	depth	count y
2005-08-13	12:58:43	23.37	104.04	5.3	3	Wenshang
2005-09-05	21:14:39	27.11	103.43	4.8	10	Zhaotong

In table 1, each record represent once a earthquake happen, which consist of date, time, longitude, magnitude, depth and county name.

B. Line Feature

Activity fault zone is a region created by crustal movement pressure and tension, it reveal the crustal movement and earthquake activity. In geographical scale, the activity fault zone abstract as line feature.

TABLE II. GEOLOGICAL FAULT ZONE LINE FEATURE.

name	property	length	Formative year	occurrence	landslide rate
xiaojiang	Walking landslide fault	440		N/70-80	1.7-3.5
jiangsha jiang	Dextral inverse go	330	Paleozoic	SN/60	5-7

In table 2, each record represent an activity fault zone consist of fault name, property, length, formative year, occurrence, landslide rate.

In seismic data, earthquake source location show great spatial correlation with fault zone. Then apply *apriori* algorithm conception in analysis point and line pattern.

III. POINT LINE ASSOCIATION PATTERN

In spatial, point feature is a discrete data, however, line feature is a continuous data.

apriori algorithm is based on transaction conception, definite point and line transaction conception, is the key step for point and line spatial data mining[5].

Suppose L represented line feature, and P represented point feature. Definitions: $R(L, \theta)$ is a buffer region in the center of L line and θ as the radius buffer region. if $P \in R$, means that the point and line feature have an association rule $L \Rightarrow$

$P(CS\%, CC\%)$, $CS\%$ as support rate and $CC\%$ as confidence rate.

Then abstract the transaction conception in continuous spatial. Definitions: $L(A)$ specific line feature A in spatial, $L(A)_i$ is object i of A feature set. for example, $L(A)$ is activity fault zone type feature, meanwhile, $L(A)_i$ is the i fault zone line in $L(A)$ set. set the threshold θ for generate a buffer region $R(L(A)_i, \theta)$. $P(A)$ and $P(B)$ is two different type sets of point feature, $P(A)_i \in P(A), P(B)_i \in P(B)$, $P(A)_i$ is once an earthquake happened in spatial and $P(B)_i$ is an settlement in the area.

Transaction $(L(A)_i, P(A)_i)$, type $P(A)$ include some subtype use low case latter represented $p(a), p(b), p(c)$ and $P(a) \cup P(b) \cup P(c) \cup \dots = P(A)$ and $P(a) \cap P(b) \cap P(c) = \emptyset$, it reveal that $P(a), P(b), P(c)$ are divisions of $P(A)$ set

Definitions of association rule:

$$L(A_i) \Rightarrow P(a)(CS\%CC\%) \quad (1)$$

$$CS\% = \frac{\sum_i^n \text{transaction}(L(A)_i, P(A)_i)}{\sum_i^n P(A)_i \cup \sum_i^n P(B)_i \cup \dots} \quad (2)$$

$$CC\% = \frac{\sum_i^n \text{transaction}(L(A)_i, P(a)_i)}{\sum_i^n \text{transaction}(L(A)_i, P(A)_i)} \quad (3)$$

Equation (1) is definition point and line association rule, equation (2) is definition support of association rule and equation (3) definite confidence of association rule.

Demonstrate of point and line association rule in diagram 1, show that hollow icon is point type B , and solid icon is point type A . in solid icon $P(\bullet) \cup P(\blacktriangle) \cup P(\star) = \sum_i^n P(A)_i$ $n=3$, $P(a)=P(\bullet), P(b)=P(\blacktriangle), P(c)=P(\star)$. Generate only one buffer region $R(L(A)_i, \theta)$, a point object intersection with R is consider as point and line have a **transaction** $(L(A)_i, P(A)_i)$, conception. In diagram 1 all transaction $(L(A)_i, P(A)_i)$ is be counted $\sum_i^n \text{transaction}(L(A)_i, P(A)_i) = 10$ means that the numbers of $P(A)$ point intersection with R is 10, $\sum_i^n P(A)_i \cup \sum_i^n P(B) = 42$ means that all numbers of A and B point Types is 42 objects.

$\sum_i^n \text{transaction}(L(A)_i, P(a)_i) = 4$ means that the subset a have four objects fall in buffer region of L . use equation (1)-(3), we computer the point and line association rule, $CC\%$ and $CS\%$ of Figure 1 in table 3

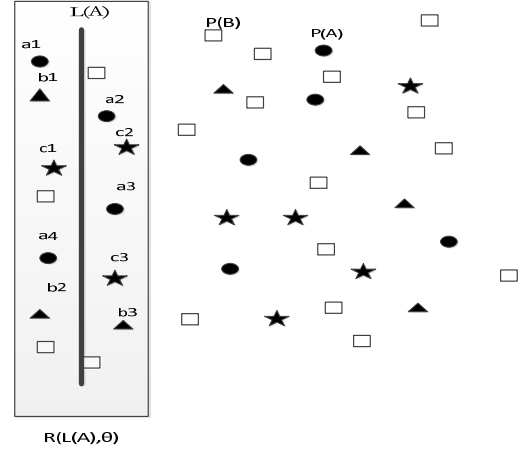


FIGURE 1. POINT AND LINE PATTERN RELATIONSHIP.

TABLE III. SUPPORT AND CONFIDENCE OF POINT AND LINE RELATIONSHIP.

Association rule	CS%	CC%
$L(A_i) \Rightarrow P(a)$	$CS\% = 10/42 = 23.8\%$	$CC\% = 4/10 = 40\%$
$L(A_i) \Rightarrow P(b)$	$CS\% = 10/42 = 23.8\%$	$CC\% = 3/10 = 30\%$
$L(A_i) \Rightarrow P(c)$	$CS\% = 10/42 = 23.8\%$	$CC\% = 3/10 = 30\%$

Above mentions transaction conception is based on *apriori* algorithm for database mining then extends to co-location algorithm for point feature, furthermore, for point and line pattern. In *apriori* algorithm all transactions must pruning by support and confidence in the purpose of find strong association rule in application. So it must set two threshold θ_s for support, and θ_c for confidence.

In point and line association rule mining process, we conclude that the key step is abstract transaction conception, and set $\theta, \theta_s, \theta_c$ as thresholds for line buffer region, support and confidence. Next we will description the algorithm process.

IV. ALGORITHM DESCRIPTION

Input of algorithm are a line type set and two point types sets, however, output is a set of spatial association rule with $CS\%$ and $CC\%$. It need three parameters $\theta, \theta_s, \theta_c$, and the process of algorithm below.

Input:

- line feature set $L(A)$ is a set of line feature type;
- point feature set $P(A)$ and $P(B)$ are two sets of different types of point feature;
- $P(a), P(b)$ and $P(c)$ are a subsets of $P(A)$;
- threshold $\theta, \theta_s, \theta_c$, are buffer threshold, support threshold and confident threshold;

Output:

point and line association rule $L(A_i) \Rightarrow P(i) (CS\%, CC\%)$;

Variable:

$L(A)_k$ a set of line type contain k objects

$P(B)_m$ a set of point type B contain m objects;

$P(a), P(b), P(c)$ are three point type subsets of $P(A)$;

$R(A)_k$ a buffer region set of line with θ as threshold contain k regions

Rule ($L(A_i) \Rightarrow P(i), CS\%, CC\%$) a point and line association rule set with support and confidence.

Step:

$R(A)_k = \text{generate_Region}(L(A)_k, \theta)$;

Count=m+n; i=0;

while(i<k){ j=0; a=0; b=0; c=0;

while(j<n){

if(intersection_Region($(R(A)_i, P(A)_j)$)) && $P(A)_j \in P(a)$ {a++; j++;}

Else

if(intersection_Region($(R(A)_i, P(A)_j)$)) && $P(A)_j \in P(b)$ {b++; j++;}

Else if ((Intersection_Region($(R(A)_i, P(A)_j)$)) && $P(A)_j \in P(c)$) {c++; j++;}

Else j++;}

If(j/count>=0s&&a/j>=0c) Add_Rule(Rule($L(A_i) \Rightarrow P(a), j/\text{count}, a/j$);

If(j/count>=0s&&b/j>=0C) Add_Rule(Rule($L(A_i) \Rightarrow P(b), j/\text{count}, b/j$);

If(j/count>=0s&&c/j>=0C) Add_Rule(Rule($L(A_i) \Rightarrow P(c), j/\text{count}, c/j$);

i++;}

In the algorithm set $P(A)$ is divide into three subsets $P(a), P(b), P(c)$, furthermore it will divided into more than three subsets. The algorithms apply in seismic spatial data with point and line feature to find out association rule.

V. EXPERIMENT

Experiment collect the earthquake source catalog during 1965-2014 years of Yunnan province china then achieve it at ArcGIS platform. First, we establish a Yunnan earthquake spatial database in Geodatabase, which contain 122201 earthquake source catalog and 33 activity fault zone. Also it contains the topographic map, settlements, county, and river as the reference data for seismic spatial database. Second divided earthquake catalog as two type which are above 4 magnitude $P(A)$ and below 4 magnitude $P(B)$. Because when earthquake happened above 4 magnitudes is sense earthquake meanwhile below 4 magnitudes is no sense earthquake for people. Focus on sense earthquake of $P(A)$ set are 1342 items in it that occupy 1% of all earthquake data, it means that we must choose support less than 1%, so choose $\theta_s=0.4\%$ and $\theta_c=1\%$, and $\theta=10\text{km}$. final, divide $P(A)$ into three subsets, which are 4-5 magnitude, 5-6 magnitude, more than 6 magnitude earthquake

data. The experiment achieve in ArcGIS platform show in Figure 2, and the results of experiment show in table 4.

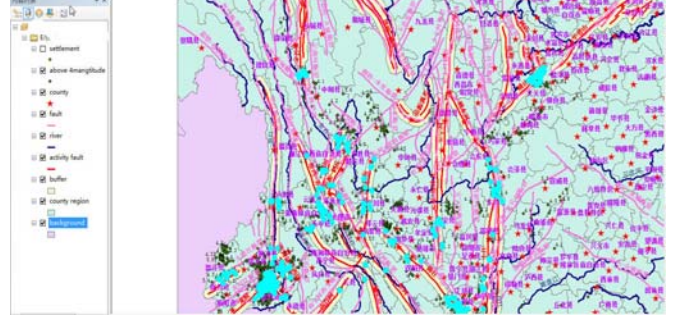


FIGURE II. POINT AND LINE PATTERN FOR MINING EXPERIMENT OF EARTHQUAKE OF YUNNAN PROVINCE

TABLE IV. POINT AND LINE PATTERN SUPPORT AND CONFIDENT IN YUNNAN EARTHQUAKE SPATIAL DATA.

Point-line association rule	support	confidence
$L(\text{fault}) \Rightarrow P(4-5)$	0.45%	80.9%
$L(\text{fault}) \Rightarrow P(5-6)$	0.45%	15.4%
$L(\text{fault}) \Rightarrow P(6\text{above})$	0.45%	3.6%

The experiment reveal probability correlation in spatial database between earthquake source and fault zone, furthermore, infer some conclusion about earthquake happened in Yunnan province:

Probability above 4 magnitude earthquake happen in the 10km buffer of activity fault zone is 0.45%.

If above 4 magnitude earthquake had happen, probability of 4-5 magnitude earthquake happened is 80.9% in the 10km buffer of activity fault zone.

If above 4 magnitude earthquake had happen, earth quake happen with 4-5 magnitude, 5-6 magnitude, above 6 magnitude probability are 80.9%, 15.4% and 3.6% ,in the area of the 10km buffer of activity fault zone.

VI. CONCLUSION

Yunnan Province is located in China's western Pacific plate and Indian tectonic plates, earthquake-prone region, a number of active fault zones in the region to study the relationship between earthquakes and fault avoiding loss of lives and property of some significance[6]. This paper proposal a algorithm for mining point and line spatial association rule base on the conception of transaction, and apply it at Yunnan seismic spatial database, get some probability of earthquake. however, the experiments some problems need improve or more thinking:

(1) Algorithm support and confident are probability and conditional probability of total probability that make it possibly to use some probability methods for further spatial reasoning such as Bayes[7].

(2) Seismic data are time sensibility, the algorithm only consider the spatial attribution of it, ignore time dimension at spatial data, next work, can analysis extends to time dimension.

(3) The presence of defects in the recording earthquake catalog data itself, the lack of historical statistics, resulting in the data is not comprehensive and so on.

REFERENCES

- [1] Yan Huang, Shashi Shekhar, hui Xiong Discovering co-location patterns from spatial Datasets A General Approach[C] .IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING 2004:1472–1485.
- [2] Martin Ester,Alexander Formmelt,Hans-Peter Kriegel Jorge Sander, Spatial Data Mining :Database primitives,Algorithms and Efficient DBMS Support,Data mining and konwledge Discovery,4193-216,2000,Kluwer Academic publisher,manufactures in the Netherlands.
- [3] Ding, W., Jiamthapthaksin, R., Parmar, R., Jiang, D., Stepinski, T. F., & Eick, C. F. (2008). Towards Region Discovery in Spatial Datasets.Advances in Knowledge Discovery and Data Mining. Springer Berlin Heidelberg.
- [4] Liu hongtao. Applied Research in Earthquake Prediction Data Mining . Liaoning University of Engineering and Technology A sissertaiton for Doctor's degree, 2006.
- [5] Ma ronghua,Ma xiaodong,Pu yingxia,Mining spatial association rules from the GIS database research. Remote Sensing, 2005, 9(6): 733-740.
- [6] Huangfu Gang. Research on the seismicity inYnnan china university of Science and Technology of china A dissertation for Doctor's degree 2009.
- [7] Guyuan zhu peiming ,rong hui ,zeng fang ,hai yang,Based on Bayesian Network seismic facies classification journal of Earth Science - China University of Geosciences, 2013, 38(5): 1143-1147.