Exceptional patterns in Multi database mining

R.Suganthi*[1], Dr.P.Kamalakannan [2]

[1]Department of Computer Applications, Valluvar College of Science and Management, Karur, Tamilnadu, India
[2]Department of Computer Science, Government Arts College, Namakkal, Tamilnadu, India
[1]mcasuganthi2011@gmail.com, [2]kamal_karthi96@yahoo.co.in

Abstract

Nowadays, A multi database mining is a important role between head quarters company and their corresponding branch offices in various places among the world. It allows to forward their branch office local patterns to the head quarters which will be synthesized for taking decision in multi databases. we propose a new approach to finding exceptional patterns and compared with the previous methods in multiple databases. the result is compared with the various methodologies yields high performance result than the other methods. it perform well with multiple dataset and it is simple and effective. it yields high outstanding accuracy and the resulting data could be used for further mining.

Keywords: Exception patterns, machine learning, multi database, local patterns

1. Introduction

The development in communication technologies and different kinds of data in business activities demand various new perceptions in industrial growth. Data mining can be viewed as a result of the natural evolution of information technology. The discovery of useful patterns in multi database environment consists of a group databases distributed in a wide area network. For effective decision making, global organizations are required to mine multiple databases that are distributed throughout their local branches which may be physically located even in different places in the world. Zhang defined multi data base mining can be defined as the process of mining data from multiple databases ,which may be heterogeneous and finding out novel and useful patterns of significance. The most familiar research journals on multi database mining have been presented by Shichao Zhang and Adhikari and others [1-7] (Thirunavukarasu Ramkumar and srinivasan[8] in their work discuss the synthesizing high frequency rules while mining multiple databases, which synthesis the global support based on transaction population and it is deduced directly through the following 3 algorithms, namely Rule selection, Rule synthesizing and global support rate(ri).

Zhang et al have denounce moving data to a central location and then mined which they call mono database mining (figure 1). The initial step of Multi database mining is to taking large data to head quarters company for giving the better decision in all aspects. To avoid the cost of moving huge data to the head quarters, the strategy will effectively obliterate the local patterns in each local databases. Here we have to consider the two levels of decisions of multiple databases in business organization. The first important decision is global or head quarters level and local or branch level decision . the following is categorize the patterns in multi database systems as local patterns, high vote patterns, exceptional patterns and suggested patterns. Machine learning is a technique that can learn automatically and predict the results based on the previous observation. Machine learning based exceptions provides an accuracy and also improves the performance of the result.

2. Previous work

The literature surveys on synthesizing global exceptional patterns which is related to our work are discussed below. The Multi database Mining is the significant strategy which is used to take assessment in head quarters level which leads to growth of the company. Particularly, various patterns are used by the researchers to give better solution according to taken the necessary steps to improve the company profit. Adhikari [9] discuss the exceptional frequent item sets which is a pre requirement step for avoiding the

transportation cost and human efforts and expiry of the products in the business organization. For example, Let X be extracted from D1,D2,...D$_k$.1<k<A.Let the actual support of 'X' in Di be supp a(X, Di),i=1,2,...k. Then the average of theses supports is obtained by the following formula.

Avg (Support(X),D1,D2,...DK)=($\sum_{i=1}^{k}$ Suppa(X, D1))/k

The Exceptional patterns is in 2 ways.

(1) The pattern that has been extracted from the most of the databases, but has a low support in D(type I exceptional pattern)

(2) A pattern that has been extracted from a few databases. but has a high support in D (Type II exceptional pattern).Due to the consideration of all branch data bases. These two types of exceptional patterns are most important role in all ways of the company. Zhang et al   describe the process of Multi database mining and patterns that exist in multi data bases. They have developed a model for identifying exceptional patterns from multiple databases which would be considered as a post processing work after mining the relevant multiple databases. Zhong have proposed the method of peculiarity patterns from multiple statistical and transaction databases. It is slightly different from exceptional patterns in the proposed method.

Wu and Zhang [10] advocated an approach for identifying patterns in multiple database by weight. Ribeiro et al described a way of extending the INLEN system for Multi database mining by incorporating primary and foreign keys as well as developing processing knowledge segments. Wrobel extended the concept of foreign keys to include foreign links since Multiple database mining also involves accessing non key attributes.

Arionei et al   introduced a system called world that uses spreading activation to enable inductive learning from multiple tables in multiple databases spread across the network. Existing parallel data mining techniques can also be used to deal with multi database

The above mentioned papers provide a good insight into multi database mining. However there are still some limitations in traditional multi database mining.

3. Methodology

Multi database mining has been recognized as an important research topic among KDD Community. A method for Multi data base mining technique is depicted in figure 1.

Figure 1. Traditional Multi database mining
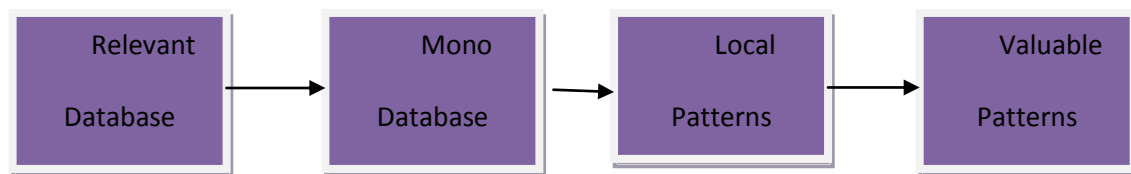
It is important to classify all the databases before mining data in multi databases. After classifying, it is important to apply the techniques for mining mono database to multi databases and then find all the

- classify the relevant databases
- mining mono database to multi database
- find local patterns
- analyze patterns to find valuable information

after that, every database local patterns will find easily. At last, we can analyse the needed patterns in the form of local patterns, exceptional patterns, high voting patterns and synthesized patterns. these type of patterns will be evaluated in order to make valuable information in head quarters level.

Exceptional patterns

few local branches only highly supported for these kind of patterns, also other branches will be zero support. It reflect the individuality of each branch. Here the head company can make some special policies and adjust measure to the local conditions.

Process

the data extracted from the multiple databases is pre processed as follows according to the user perspective

- generates the set of patterns from each database
- counts each patterns votes and the average votes of patterns average votes
- generate the candidates exceptional patterns
- compute the CEP $supp_g(p)$ values.
- based on $supp_g(p)$ values, rank the CEP.
- find the exceptional patterns

Generation of the patterns

The initial step of multiple database is to analyse the local patterns in the local branches which is controlled by head quarters company. In some situation, head company want to know the value of exceptional patterns for further prospects. so that they can avoid the non profitable things to circulate to all branches.

Example :

Consider 3 databases D1,D2,D3 and the corresponding patterns of the 3 databases are {P,Q,R,S,T,U,V}and the minimum support of database is,$mins_1=0.49$,$mins_{2=0.48}$,$mins_{3=0.82}$,$mins_{4=0.20}$,$mins_{5=0.13}$ are 5 databases minimal support respectively.

LP1={{P:0.69};{R:0.68};{U:0.52}}

LP2={{P:0.52};{Q:0.62};{K:0.91};{T:0.82};{U:0.76};{V:0.86}}

LP3={{P:0.87};{R:0.85};{S:0.86};{T:0.86};{F:0.95}}

LP4={{Q:0.36};{R:0.31};{T:0.28}}

LP5={{U:0.22}}

proposed algorithm(exceptional patterns in local patterns)is used to search all the exceptional patterns from the given local patterns. according to the design of the algorithm, we can get the global patterns is{ p,q,r,st,u,v} and the average votes18/7=2.57 because {q,s,v}have less votes than the average votes. here q,s,v are called the candidates exceptional patterns. the support $_g(p)$ value of each pattern in candidate exceptional patterns are follows,

supp (q)=0.215 pattern q comes from {d1,d4}

supp (s)=0.2v01 pattern r comes from {d3}

supp (v)=0.69 pattern v comes from {d2}

after analyzing of the all patterns, the rank will be{s,q,v}.here supp{v}is the highest support and it is supported by only a database. so it can be called as an exceptional patterns after getting the such kind of exceptional patterns, then head company   is able to use this type of patterns to make important decision among the branches and it is the main role of head company

3. Results and Discussion

We test this proposed method by using the dataset which is available in www.ics.uci.edu/mlearn/MLSummary .html the dataset D1,D2,D3 be 3 databases of 5000,3500 and 10,000 transactions respectively. Here DU be the union of D1,D2,D3.Assume that α=0.2 , α=0.5. The set of patterns extracted from these databases are given as follows.

CEP(D1={Q(0.11),PQ(0.09)}

CEP(D2={Q(0.10),U(0.12),V(0.15)}

CEP(D3={S(0.10),T(0.71)}

and average length of the transactions, number of transaction , average frequent items denote ALT,NT,AFI which is the needed step to analyses local and exceptional patterns

Algorithm 1 describes the finding local patterns from the data set as well as Candidate Exceptional patterns(CEP).

Table1 show the performance of the various algorithms. with the proposed methodology yields high performance when compare to the mentioned algorithm. It achieves 84% accuracy and has an error rate of 15.27% which is better than other algorithms and It acquired high level exceptional patterns when compared to the other algorithms. It obtained better result

Table 1. Density and Distance based methods performance

Measure	Density based approaches	Distance based approaches	Proposed result
Experiment	0.175	0.173	0.186
Error rate	14.7894	19.9781	12.1184
Accuracy	71.2727	78.1246	82.2105

4. Conclusion

This paper has proposed a Exceptional pattern algorithm on the datasets to classify the local item sets and exceptional patterns. From the given data sets the result is compared with the outlier detection algorithms. The proposed methodology gives exceptional accuracy better than the other algorithms accuracy and performs well with the large dataset. It is simple and effective .The result and data are helpful for reducing transportation cost of the manufacturing company and multinational company for the mining information.

5. References

1. Adhikari, Synthesizing global exceptional patterns in different data sources. *Journal of Intelligent Systems*. 2012; 21 (3), 293-323.
2. Adhikari, P. Ramachandrarao, W. Pedrycz, Study of select items in different data sources by grouping, *Knowledge and Information Systems*. 2011; 27 (1) 23-43.
3. Adhikari, P. Ramachandrarao, W. Pedrycz, Developing Multi-Database Mining Applications, Springer-Verlag, London, 2010. 266-279.
4. Adhikari, P.R. Rao, B. Prasad, J. Adhikari, Mining multiple large data sources. *International Arab Journal of Information Technology.* 2010; 7 (2), 243-251.
5. Adhikari, P.R. Rao, Capturing association among items in a database. *Data & Knowledge Engineering.* 2008; 67 (3), 430-443.
6. Adhikari, P.R. Rao, Mining conditional patterns in a database. *Pattern Recognition Letters. 2008;* 29(10), 1515-1523.
7. Adhikari, P.R. Rao, Efficient clustering of databases induced by local patterns. *Decision Support Systems. 2008;* 44(4), 925-943.
8. T.R Ramkumar, S. Hariharan, S. Selvamuthukumaran. A survey on mining multiple data sources. *WIREs Data Mining and Knowledge Discovery.* 2013; 3, 1-11.
9. Adhikari A, Synthesizing global exceptional patterns in different data sources. *Journal of Intelligent Systems.* 2012; 21(3), 293-323.
10. Zhang S, Zaki JM. Mining multiple data sources: lo-cal pattern analysis. *Data Mining and Knowledge Discovery.* 2006; 12(3), 121-125.