

A Review: Data Mining over Multi-Relations

Deepak Meena
Research Scholar
Patel Institute of Technology
Bhopal

Hitesh Gupta
Head, CSE, Dept
Patel Institute of Technology
Bhopal

ABSTRACT

In this paper, Multi-relational data mining enables pattern mining from multiple tables. Multi-relational data mining algorithms can be used as practical proposal to overcome the deficiency of conventional algorithms. Multi-relational data mining algorithms directly extract frequent patterns from different registers in efficient manner without need of transfer the data in a single table will, on the other hand, used the available memory space is not enough to ensure the production of large amounts of data. For this reason, and the use of space, algorithms are an integral care for the prospection of large repositories. The paper provides the overview of multi relation data mining techniques and classification algorithms. It also defines the frequent pattern mining. The presented paper discussed the various architecture and issues related to multi table data mining. A lot of literature has been proposed in this area. Some of them has discussed in this paper.

Keywords

Data mining, multi relation classification, FP tree.

1. INTRODUCTION

The most of the available data mining technique are appropriate for a single relation of database. Now there may be some difficulty in order to use the multi relation of database. To perform join operation in database is a time consuming task. Ones the join operation applied then there is a possibility to lost some information or data may be repeated in database. So how can this complexity reduce or remove in order to get the meaningful patterns for multiple tables. Data mining is a rapid growth field in order to get useful information from database [1]. Data mining includes the foundation of probability; classical machine learning algorithms which belong to data mining. Association rule is also an important part of data mining. It has used in order to present the knowledge. If there is relation between itemsets and huge number of transactions then there is possibility of association rule mining. The researchers are working on it. It seems to be that there are two problems in traditional association rule mining. First, traditional methods suppose items in transaction database have same significance, thus the mining process is flooded in the combinatorial explosion of insignificant relationships. The mining result may comprise duplicate information and consume lots of space. Second, those methods ignore the difference between two items, which may lead to incorrect result biased with users' expectation in real applications. Frequent pattern mining [7], [13] plays a crucial role in many data mining tasks like data mining association rules, discovering patterns having useful correlations, constraint based patterns etc.

2. DATA MINING

A knowledge Discovery process is known as the data mining technique. Data mining is used to deal with huge data which are in the database, it is use in order to find the desired information and knowledge from the database. Data Mining is more oriented towards applications than the basic nature of the underlying phenomena. In other words, Data Mining is relatively less concerned with identifying the specific relations between the involved variables.

There are three basic steps in order to perform the data mining. (1) The initial exploration, (2) model building or pattern identification (3) deployment

There are many data mining techniques has proposed by the researchers such as, decision trees, association rules, and neural networks etc. The other most useful data mining technique appear in order to find patterns is multi-relational data mining (MRDM) approach. It includes multiple tables or relations from the given relational database [1,2]. Association, classification, clustering, prediction and sequential patterns are also some other techniques used in data mining. Data mining is not only using in business ambiance but also in other world such as weather forecast, healthcare, insurance, medicine, transportation etc. data mining can also use in banking, finance, retail and marketing. But some time there may be problem like privacy issues and security issues.

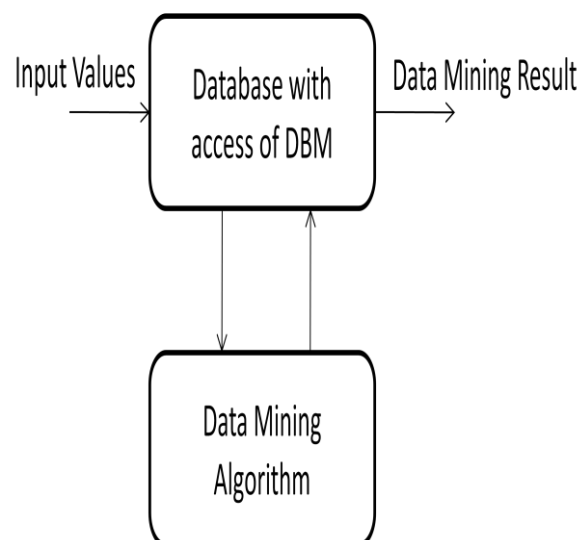


Figure 1 : Simple Working of Data mining Technique

The above figure shows the how the data mining method works on the raw data which is stored in database.

3. MULTI RELATION CLASSIFICATION

Classification [1] is an important technique, has a broad collection of applications in the world of data mining and machine learning. Classification has goal to determine a set of Association mining rules in the database that assure some minimum support and minimum confidence constraints and forms an accurate classifier [2]. Associative classification using association rules is a method that builds classifier with the help of association rules. Generally it has two steps: first one is to find all the class association rules whose right-hand side is a class label, after that select the strongest rules from the CARs to build the classifier. In this approach, associative classification can generate rules with higher confidence and better support with conventional approaches.

Let D is the dataset. Let I be the set of all items in D and C be the set of class labels. We say that a data case $d_i \in D$ contains $X \subseteq I$, a subset of items, if $X \subseteq d_i$. A class association rule (CAR) is an implication of the form $X \rightarrow c$, where $X \subseteq I$, and $c \in C$. Bing Liu et al. [9] first proposed the AC approach, named classification based on association algorithm (CBA), for building a classifier based on the set of discovered class association rules. The difference between rule discovery in AC and conventional frequent item set mining is that the former task may carry out multiple frequent item set mining processed for mining rules of different classes simultaneously [10]. Data mining in associative classification (AC) framework usually consists of two steps

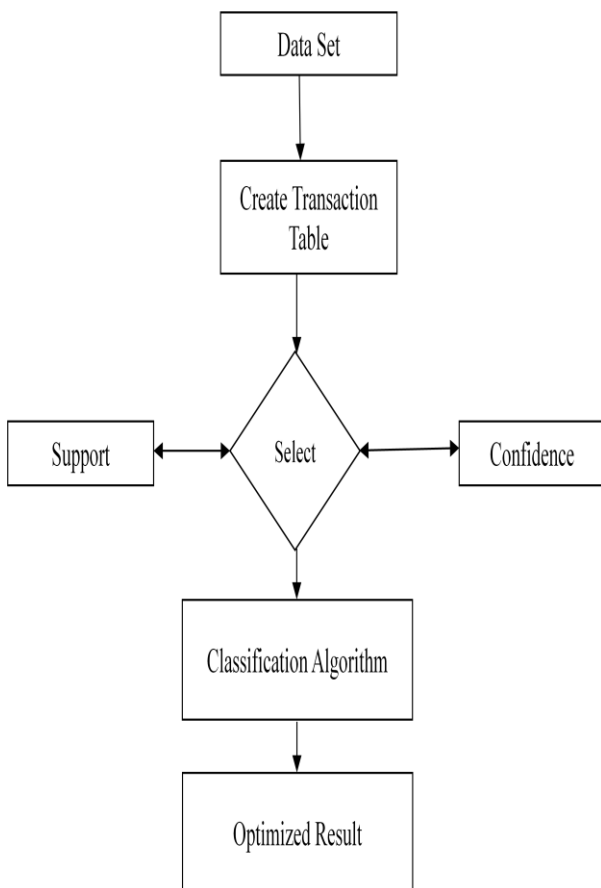


Figure 2 Simple architecture of Multi relation classification

Multi-relational data mining exploits for pattern mining from more than one table in database. The present multi-relational mining techniques with association rules mining are unable to process large volumes of data. The reason seems to be that the higher use of memory but there are many algorithms like MR-Radix and many more which supports the optimized memory use.

The figure below shows the example in order to explain the multi relation concept in database.

The classification needed to better mining from large database. Classification is a data analysis task where a model or classifier is constructed to calculate categorical labels. There are various issue need to concern regarding classification. These issue shows in figure 4. Classification represents data into predefine groups or classes. Pattern recognition is a type of classification where an input pattern is classified into one of several classes based on its similarity to these predefined classes

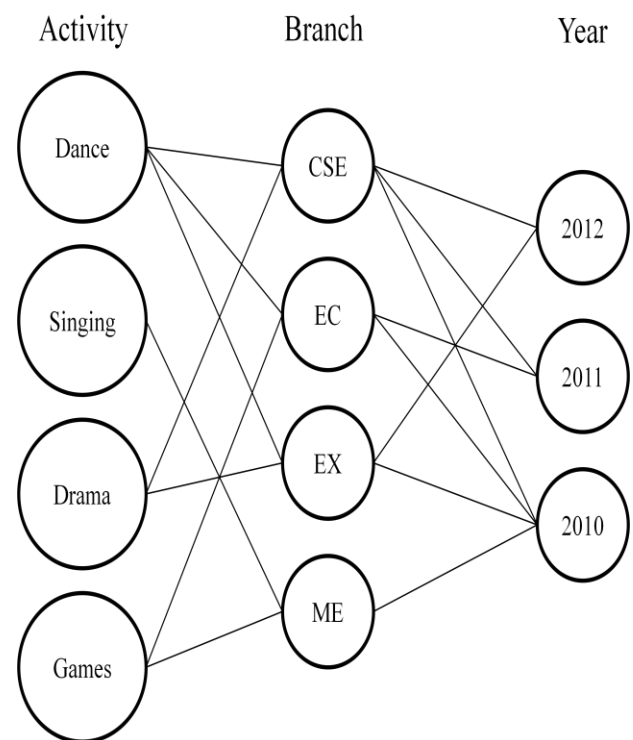


Figure 3 Multi Relation Classification Examples

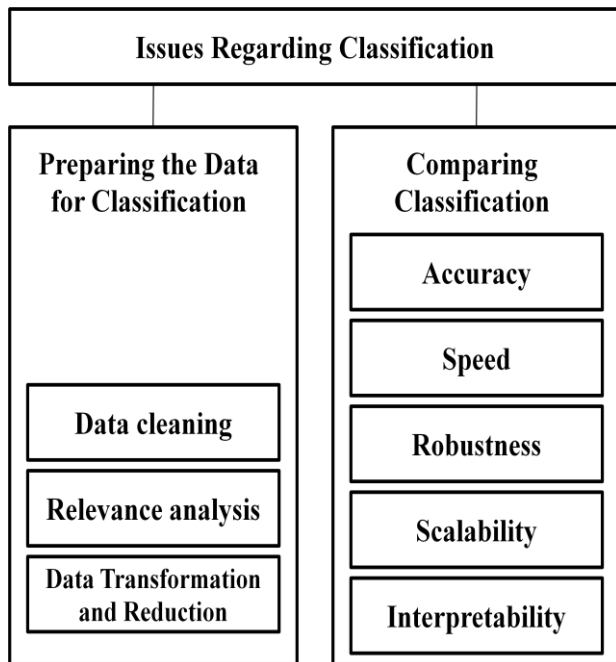


Figure 4 Issues in Classification technique

4. FREQUENT PATTERN

Frequent-pattern mining plays a crucial and important role in many data mining tasks, including association rules mining and patterns analysis. The concepts of support, confidence and frequent item sets based on multiple tables are similar to the definitions based on a single table. How to generate frequent patterns efficiently has always been the main and active topic in this field. The FP-Growth Algorithm is a secondary algorithm used to find frequent item sets. It is different from the Apriori algorithm. Size of FP-tree depends on how items are ordered.

In many cases, the Apriori algorithm significantly reduces the size of candidate sets using the Apriori principle. However, it can suffer from two-nontrivial costs [2]:

- A. Generating a huge number of candidate sets, and
- B. Repeatedly scanning the database and checking the candidates by pattern matching.

FP-growth method that mines the complete set of frequent item sets without candidate generation. FP-growth follows the divide-and-conquer approach. The first scan of the database gets a list of frequent items. This list has the items ordered by frequency-descending order. According to the frequency-descending list, the database is compressed into a frequent-pattern tree (FP-tree), which retains the item set association information. The FP-tree is mined by starting from each frequent length pattern, constructing conditional pattern base, then constructing its conditional FP-tree, and performing mining recursively on such a tree [2,5,10].

Frequent patterns mining can be applied in order to a many application domains, for example indexing and similarity search of complex structured data, multimedia mining, and web mining etc.

5. LITERATURE REVIEW

There is lots of work has been done in the world of multi relation classification and FP growth tree. We studied lots of research articles, conclusion is presented here.

A new frequent pattern mining algorithm [8], called LPS-Miner, which bases the pattern growth principle and uses two new data structures, LPS-FP-Tree (Light Partial-Support FP-Tree) and LPS-Forest (Light Partial-Support FP-Tree Forest) to present the database. LPS-FP-Tree is a variation of FP-Tree with lighter unidirectional nodes and the mining process depends on the partial-support of the patterns.

Another method [9] proposed called Incremental Share-Frequent Pattern Tree which is a enhance tree of previous methods. Here the author used the previous tree structure as well as mining results at time of database updated or minimum threshold got change. It required only two resultant share-frequent patterns in incremental databases. Its performance is slow in compare to other method but still it is efficient.

The multi relation mining can apply on network intrusion detection [10]. Without analysis of huge data given by network safety terminals it is not possible to detect the network intrusion which might be complex attack. So multi relational mining algorithm using association rule mining as well as probability function has used to find the desired results.

The author proposed an efficient method [11] for mining strong negative association rules in multi-database. The method produces some infrequent item sets in relational patterns by pruning and scanning constructed Multi-Database Frequent Pattern tree, and extracts strong negative association rules according to the proposed correlation model.

A framework work on Constraint base mining has also proposed [12].

(1) It allows getting patterns not only under anti-monotonic constraints, but also under monotonic constraints and closeness constraints, among others, expressed over complex aggregates over multiple relations;

(2) It builds on a declarative graphical representation of constraints that links closely to data models of multi-relational Databases and constraint networks in constraint programming;

(3) It maps multi-relational pattern mining tasks into constraint programs.

6. CONCLUSION

This paper throws some light on multi relation classification and frequent pattern. It is also shows the previous work in this field. We studied many research papers and found that frequent pattern mining gives the batter results in multi relation classification. But having higher degree of false -ve rate, lower classification rate and higher time complex. In future we will try to negotiate all of these problems. .

7. ACKNOWLEDGEMENT

I would like to say thanks to my guide “Prof. Hitesh Gupta” who gives their knowledge and time in order to complete this paper. This paper will never complete without the support faculty member of CSE department of Patel Institute of Technology, Bhopal.

8. REFERENCES

- [1]. Han J. and M. Kamber (2000), *Data Mining: Concepts and Techniques*, Academic Press, San Diego, CA.
- [2]. Xun Zhu, Hongtao Deng and Zheng Chen, “A Brief Review On Frequent Pattern Mining”, *IEEE* 2011, PP-4-11.
- [3]. Yingqin Gu^{1,2}, Hongyan Liu³, Jun He^{1,2}, Bo Hu^{1,2} and Xiaoyong Du^{1,2} “A Multi-relational Classification Algorithm based on Association Rules” pp.4-9 2009 *IEEE*.
- [4]. S.P.Syed Ibrahim¹ K. R. Chandran² M. S. Abinaya³ “Compact Weighted Associative Classification” *IEEE* pp.8-11, 2011.
- [5]. Pei-yi hao, yu-de Chen “a novel associative classification algorithm: a combination of LAC and CMAR with new measure of Weighted effect of each rule group” *IEEE* pp.9-11, 2011.
- [6]. Xun Zhu¹, Hongtao Deng², Zheng Chen³ “A Brief Review On Frequent Pattern Mining” PP-4-11 2011 *IEEE*.
- [7]. Thair Nu Phyu “Survey of Classification Techniques in Data Mining” Vol I *Imeccs* 2009, March 18 - 20, 2009, Hong Kong.
- [8]. Xiaoyun Chen, Huiling Liu, Pengfei Chen, Longjie Li, “A High Performance Algorithm for Mining Frequent Patterns: LPS-Miner”, *IEEE* 2008, pp 7-11.
- [9]. Chowdhury Farhan Ahmed, Syed Khairuzzaman Tanbeer, Byeong-Soo Jeong, “An Efficient Method for Incremental Mining of Share-Frequent Patterns”, *IEEE* 2010, pp 147-153
- [10].Zhang Youdong, “Multi Relational Mining in Network Intrusion Detection”, *IEEE* 2008, pp 445-448.
- [11].Hong Li and Xuegang Hu , “Efficient Mining of Strong Negative Association Rules in Multi-Database”, *IEEE* 2009.
- [12].Siegfried Nijssen, Aida Jimenez and Tias Guns, "Constraint-based Pattern Mining in Multi-Relational Databases", *IEEE* 2011, pp 1120-1127.
- [13].J. Han, J. Pei, Y. Yin, and R. Mao, “Mining frequent patterns without candidate generation: a frequent pattern tree approach,” *Data Mining and Knowledge Discovery*, vol.8, *IEEE* 2004, pp. 53-87.