

An Assessment of Fuzzy Temporal Association Rule Mining

Sourabh Jain¹, Assistant Prof. Susheel Jain², Prof. Anurag Jain³

^{1,2,3} Computer Science Dept .Radharaman Institute of Technology & Science, Bhopal (M.P.) India

ABSTRACT

In the current scenario data is increasing day by day. To search the required result there is a need of fast searching techniques. But it is too difficult to make any fast searching algorithm. There are two way to search fast . First is to increase our processing speed. Second is, narrow the searching area. The second is more efficient and less costly then first one. Data mining reduces the searching space . Data mining tasks can be classified in to two categories . Descriptive mining and Predictive mining. Association Rule mining (ARM), Clustering and sequential pattern mining are some of the descriptive mining. The other technique works with respect to time are known as temporal data mining. The presented paper is a review in temporal data mining and the fuzzy association rule in order to get the required data fastly and efficiently as well.

Keywords: Data mining, fuzzy association rules, frequent patterns, temporal data mining.

1. INTRODUCTION

The growing world of temporal data mining plays a significant role in the world of mining data from the large database . On one hand Association rule encounters association relationship between the items of database and the attribute in a very simple manner . On other hand, the relations of temporal processes are complex up to certain extend , which can not be reflected fully through the some discrete temporal states. The temporal association rule mining is used to extract association relations among temporal processes.

As everyone know that temporal mining is a part of data mining . Mining frequent sets over data present gorgeous challenges over traditional mining technique in huge databases. Data mining is a generally used for discovering information and knowledge from huge databases , where 'association rules discovery , is one of the most popular technologies. It was introduced by Argawal in 1993 [1] . It provides information of the type of "if-then" statements. These rules are generated from the dataset and it draw from measurements of the support and confidence of each rule that can show the frequency of occurrence of a given rule . Association Analysis [1, 2, 4, 6] is the process of discovering hidden pattern or condition that occur frequently together in a given data . Association Rule mining techniques finds interesting associations and correlations among data set. An association rule [1,3,4,5] is a rule, which entails certain association relationships with items, for example the interrelationship of the data item as whether they occur simultaneously with other data item and how often. These rules are computed from the data and , association rules are calculated with help of probability. Mining frequent item-sets is a fundamental and essential problem in many data mining applications such as the discovery of association rules, strong rules , correlations, multi-dimensional patterns , and many other important discovery tasks . The problem is formulated as follows: Given a large database of set of items transactions, find all frequent item sets, where a frequent item set is one that occurs in at least a user- specified percentage of the database . This paper has divided into eight secessions where the first one is introduction of the area . The second one is data mining with frequent pattern . Third , one has the temporal data mining . Fourth secession explains about fuzzy association rule and finally the conclusion of the paper.

2. DATA MINING WITH FREQUENT PATTERN

Data mining acquired from a simple taking input raw data to an pre process with several steps then apply the data mining techniques in order to pattern reorganization . This pattern reorganization from large size of database will help to get knowledge from data base. It can be done in three stages.

1: Exploration - data training.

2: Architecture Construction and Verification – have the information about choosing the ones that are best appropriate method to be used.

3: Deployment – it will focus on using the selected data to be proceed with the generation of the output results.

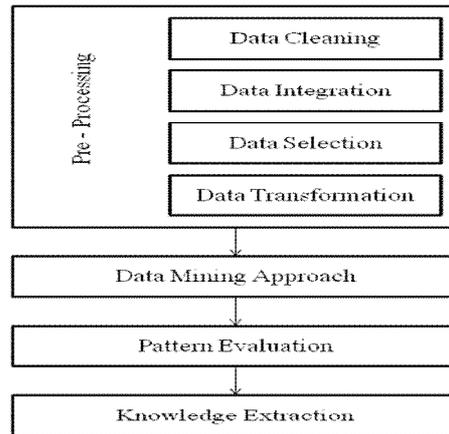


Figure 1: Steps use in data mining

A variety of data mining techniques such as, decision trees, association rules, and neural networks are already proposed and become the point of interest from numerous of years.

The result of rate of recurrence of event for all sorts of item sets known as Frequent pattern. A huge number of item sets are possible even for small number of attributes. To prune these large number of item sets, Currently temporal – based mining becomes point of interest, there are many research has been done in this area, and we will discussed them in the afterwards section. Along with that, idea of Temporal Association Rules (TAR) which includes time expressions into association rules to handle the time series for solving the problem is introduced in [7]. A standard temporal association rule is said to be frequent within its Maximum Common exhibition Period (MCP) if and only if its support and confidence are greater than the required minimum support threshold and minimum confidence respectively [7]. The association rules known as Temporal Association rules (TAR) are applicable during specific time periods [6][8]. The inclusion of constraints in the rule mining practical applications where utility mining is likely to be helpful. The objective of utility mining is to discover high utility item-sets which are attributable to a considerable portion of the total utility.

3. TEMPORAL DATA MINING

A database developed with respect to time is called the temporal database. In particular way the temporal factor usually includes legitimate time and transaction -time. Both factors work together as a temporal data [7].

- Legitimate time defines the time period in which information is correct with respect to the real world.
- A time in which the information will store in the data base is called Transaction time.
- Bi temporal data is a combination of both legitimate and Transaction Time

It is possible to have timelines other than legitimate Time and Transaction Time, such as Decision Time, in the database [8].

The Main Goal of Temporal Database are :

- Identification of an appropriate data type for time.
- Prevent fragmentation of an object description.
- Provide query algebra to deal with temporal data.
- Compatible with old database without temporal data.

4. FUZZY ASSOCIATION RULES

Fuzzy logic is efficient technique for logic having more than one value. However its uses are comparatively different from their goal. So that, the fact is fuzzy logic work with estimated value rather than specific modes of reasoning implies that, in general, the chains of reasoning in fuzzy logic are short in length and rigor does not play as important a role as it does in conventional logical systems. In order to apply classical association rule mining there is different technique called fuzzy association rule mining. If there is a data which have any kind of range of values then it is possible to encounter the sharp boundary problem. There are many problems in association rule mining like boundary sharp. The fuzzy association mining gives the solution of this problem.

Suppose there are three ranges of grades.

F(x) is a function such that

$0 \leq x \leq 35$ then F(x) = D grade

$35 < x \leq 40$ then F(x) = C grade

$40 < x \leq 50$ then F(x) = C+ grade

$50 < x \leq 70$ then F(x) = B grade

$70 < x \leq 80$ then F(x) = B+ grade

$80 < x \leq 100$ then $F(x) = A$ grade

However, if we see the grade B then 50 to 70 marks are considered. Therefore, here is a problem of sharp boundary. There are three basic approaches can use in order to solve the sharp boundary problem. Quantitative approach, Fuzzy Taxonomic Structures and Approximate Item set Approach. There is an example enclosed here in order to explain the boundary sharp problem.

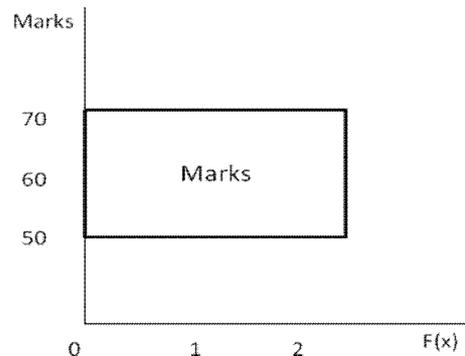


Figure 2: Example of sharp boundary problem

Fuzzy set theory is popular because of its reasoning power and its simplicity as well. Now days it is usually access for the intellectual system.

Fuzzy set theory has been used more and more habitually in intellectual systems because of its simplicity and similarity to human reasoning [5, 7,]. Numerous fuzzy data mining algorithms for inducing rules from given sets of data have been designed and used to good effect with specific domains [3, 6, 8]. As to fuzzy temporal data mining, since fuzzy calendar algebra could help users describe temporal requirements in fuzzy temporal calendars easily, Lee proposed two temporal patterns that were fuzzy temporal association rules and fuzzy periodic association rules based on fuzzy calendar algebra. Based on Lee's approach, Zhuo et al. introduced a relativity-based interest measure value for mining fuzzy calendar - based temporal association rules. However, those fuzzy data mining approaches didn't take item lifespan into consideration. Although Lee proposed two algorithms for discovering fuzzy temporal association rules and fuzzy periodic association rules by using fuzzy calendar algebra, lifespan of each item still didn't be considered.

5. RELATED WORK

A framework can detect the popular relevant topics specific to a main keyword on certain periods [12]. Three different kinds of relevant topics could be selected by this work which are non - periodical topic, periodical topic and burst topic respectively. By considering the power of the time, it is possible to extract different relevant topics to specific keyword on different time. Guillame-Bert and James L. Crowley "New Approach on Temporal Data Mining for Symbolic Time Sequences: Temporal Tree Associate Rules" The author introduced a model based on temporal pattern. It is called Temporal Tree Associative Rule has also introduced as a data mining approach. This pattern model can be used to express both uncertainty and temporal inaccuracy of temporal events expressed as Symbolic Time Sequences [7]. Maragatham. G and Lakshmi. M "A Strategy for Mining Utility based Temporal Association Rules" The proposed algorithm is able to mine temporal association rules based on utilities by adapting the support with relevant to the time periods and utility. An approach of mining UTARM is designed and the efficiency of the method is discussed. [8] The problem of mining fuzzy temporal association rules [9] from a publication database, and propose an algorithm to get it. In the algorithm, the lifespan of an item is measured by its entire publication periods in a publication database. Also, an item set table structure is designed to effectively keep and efficiently obtain information of item sets for mining. Finally, experiments on a simulation dataset were made to show the comparison results of the mined fuzzy frequent item sets with and without consideration of life spans of items under different parameter settings.

6. CONCLUSION

The aim of this study is to present an overview of the techniques that deal specifically with temporal data mining. The presented paper shows the various types of data mining techniques. The study shows the data mining can apply on the specific time duration or it can apply with respect to time. The temporal mining can easily embed with fuzzy logic in order to mine the data. It seems to be that there are lots of problem found in data mining techniques like boundary sharp problem which is easy to solve by fuzzy logic. In future some related problem can also solved using temporal mining with the help of fuzzy logic.

ACKNOWLEDGEMENT

I would like to say thanks to my guide “ Mr. Susheel jain” who gives their knowledge and time in order to complete this paper. This paper will never complete without the support of HOD of CSE department Mr.Anurag jain and Guide Mr. Susheel jain RITS college Bhopal.

REFERENCES

- [1]. R. Agrawal, T. Imielinski , and A. Swami , “ Mining Association rules between sets of items in large databases “, In proceeding of the 1993 ACM SIGMOD International Conference on management of data, pages 207-216, Washington, DC, May 26-28 1993.
- [2]. R Srikant, Qouc Vu and R Agrawal. “ Mining Association Rules with Item Constrains ”. IBM Research Centre, San Jose, CA 95120, USA.
- [3]. Ashok Savasere, E. Omiecins ki and Shamkant Navathe “An Efficient Algorithm for Mining Association Rules in Large Databases”. Proceedings of the 21st VLDB conference Zurich, Swizerland, 1995.
- [4]. R. Agrawal and R. Shrikant, “Fast Algorithm for Mining Association Rules”. Proceedings Of VLDB conference, pp 487 – 449, Santiago, Chile, 1994.
- [5]. Arun K Pujai “Data Mining techniques”. University Press (India) Pvt. Ltd. 2001
- [6]. J. Han and M. Kamber, “Data Mining: Concepts and Techniques”, Morgan Kaufman, San Francisco, CA, 2001.
- [7]. Mathieu Guillame-Bert, James L. Crowley, "New Approach on Temporal Data Mining for Symbolic Time Sequences: Temporal Tree Associate Rules," ictai, pp.748-752, 2011 IEEE 23rd International Conference on Tools with Artificial Intelligence, 2011
- [8]. Maragatham. G and Lakshmi. M “A Strategy for Mining Utility based Temporal Association Rules”, IEEE 2010, pp 38-41.
- [9]. Guo-Cheng Lan, Chun-Hao Chen, Tzung-Pei Hong and Shih-Bin Lin, “A Fuzzy Approach for Mining General Temporal Association Rules in a Publication Database”, IEEE 2011, pp 611-615
- [10]. Guimarães, G.: The Induction of Temporal Grammatical Rules from Multivariate Time Series. ICGI (2000) 127-140.
- [11]. Keogh, E., Smyth, P.: A probabilistic Approach to Fast Pattern Matching in Time Series Databases. KDD (1997) 24-30.
- [12]. F. Z. Su, C. H. Zhou, V. Lyne, Y. Y. Du et al., "A data-mining approach to determine the spatio-temporal relationship between environmental factors and fish distribution," Ecological Modelling, vol. 174, pp. 421-431, Jun 2004.
- [13]. K. Koperski, J. Han, “Discovery of spatial association rules in geographic information databases,” Lecture Notes In Computer Science, 951, pp. 47-66, 1995.

AUTHOR



Sourabh Jain is a scholar of M.Tech, (Computer Science Engineering), at R.I.T.S. Bhopal, under R.G.T.U. Bhopal, India.

Susheel Jain, Assistant Professor in Computer science department of R.I.T.S., Bhopal, M.P. He has done his M.Tech. in Software Engineering From Gautam Buddh Technical University, Lucknow, India.

Anurag Jain, H.O.D. of Computer science department of R.I.T.S. Bhopal, M.P. He has done his M.Tech, in Computer Science and Engineering, From Barkatullah University, Bhopal, India.