International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 1 Issue 5, July - 2012

A Review: Data Mining with Fuzzy Association Rule Mining

Sandeep Kumar Singh, Mr.Ganesh Wayal, Mr.Niresh sharma

#CSE DEPARTMENT , RGPV UNIVERSITY
BHOPAL

Abstract—**Rapid growth of temporal data mining has increased the interest of researchers in this era. Temporal data mining generates temporal association rules. These rules encapsulate the transaction of entity with respect to time. The result generates a quantitative value. But it is easy manipulate the fussy set rather than quantitative values with the help of member function. The paper presents a review on fuzzy logic and temporal association rule. This paper gives an idea in order to apply fuzzy set on temporal data mining.**

Keywords— Temporal association rule, fuzzy set, fuzzy temporal association rule.

I. INTRODUCTION

The emerging growth of data mining raises the large range of complex applications. It leads the broad study of data mining frequent patterns. Mining frequent sets over data streams present attractive new challenges over traditional mining in static databases. Data mining is generally used for retrieving the desire information to make it into knowledge from the large size databases. Association rules discovery is one of the most important technologies which was given by Mr. Argawal in 1993 [1]. It gives the information like "if-then" statements. These rules are invoked from the dataset. It generates from calculation 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 dataset. Association Rule mining techniques looks for interesting associations and correlations among data set. An association rule [1,3,4,5] is a rule, which entails certain association relationships with objects or 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.
The paper is organized into eight sections. The first session gives an introduction of the paper. The second one shows data mining with frequent pattern with process. Third section explains temporal data mining. Fourth one show about fuzzy association rule and finally conclude the paper.

II. DATA MINING WITH FREQUENT PATTERN

Data Mining evolved from a simple taking out of raw data to an analytical process from large amount of data in order to collect knowledge [7]. It can be done in seven stages

1. **Data Integration:** The first step needs to collect the data from various resources.
2. **Data Selection:** As the data collected then select the data which is important for data mining will extract form this large data set.
3. **Data Cleaning:** The data we have collected are not clean and may contain errors, missing values, noisy or inconsistent data. So we need to apply different techniques to get rid of such anomalies.
4. **Data Transformation:** The data even after cleaning are not ready for mining as we need to convert them into appropriate manner.
5. **Data Mining:** Now we are ready to apply data mining techniques on the data to discover the interesting patterns. Techniques like clustering and association analysis are among the many different techniques used for data mining.
6. **Pattern Evaluation and Knowledge Presentation:** This step involves visualization, transformation, removing redundant patterns etc from the patterns we generated.
7. **Decisions / Use of Discovered Knowledge:** This step helps user to implement the knowledge to take better decisions.

There are many method has proposed like decision trees, association rules, and neural networks in order to implement data mining.

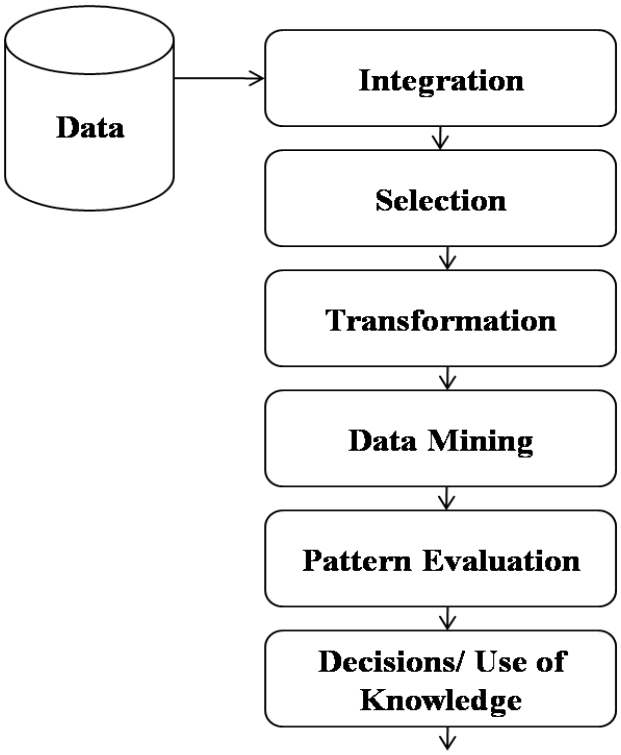

Figure 1: Steps of Data Mining

Temporal data mining become a large area of interest. There are many research has done in this era. It will discusses in the subsequently 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 [9]. 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 [9][11]. The association rules known as Temporal Association rules (TAR) are applicable during specific time periods [8][9][12]. The inclusion of constraints in the rule mining practical applications where utility mining [13][14][15] 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 [7]

III. Temporal Data Mining

A mining technique with respect to time is known as temporal data mining. It involves application of data mining technique on temporal uses data in order to discover temporal pattern and it also deal with temporal behaviour of user or data. Temporal mining is a data mining that focus on time aspects and deal with the conclusion of temporal patterns from large data sets. The use of temporal data mining is continues to raise as huge amounts of temporal data of everyday's behaviour become present easily.

Temporal data mining is one step of knowledge discovery process from temporal data base. It calculates temporal patterns or models over the temporal data. The algorithm used in this process known as temporal mining algorithm.

Results of Temporal Association Patterns for temporal data mining can be effectively applied in number of fields like trading, marketing, social networking, medicos, earth quick detection, robotics and assisted design [16]. So that explorer's number of efficient algorithms for temporal data mining like symbolic time series, symbolic time sequences, symbolic interval series, numeric time series, item set sequences, etc have proposed.

IV. Temporal Associate Rules

Temporal data mining mostly pay attention to many real time disciplines like higher engineering mathematics, meteorology, telecommunication, finance, temporal pattern recognition, temporal database, and high speed computation & parallel processing. Earlier association rule mining did not focus on time constrain of activity, however, the application eras are always changing with respect to time. Temporal association rule overcome from this problem may recommend a number of narrations, such as

$(X) \rightarrow (Y)$, $(Z) \rightarrow X$ and Y, $Z \rightarrow c$ (P)

Where

X: - Preceding Event

Y: - Succeeding Event

Z: - Event

P: - Coincidental

Traditional association rules have no concept of order, while time implies an ordering. If we could find the associability of time with event, nothing will be hidden to us as the events are associated to each other in the form $X \rightarrow A \rightarrow Y$.

Where A: - Present event

In this way temporal association rule mining is to discover the valuable relationship among the items in the temporal database.

V. FUZZY SETS

Fuzzy sets are generalized sets which allow for a graded membership of their elements. Usually the real unit interval [0; 1] is chosen as the member-ship degree structure.

Let X be a space of points, with a generic element of X denoted by x. Thus X = {x}.

A fuzzy set A in X is characterized by a membership function fA(*x*)which associates with each point in X a real number in the interval [0,1], with the values of fA(*x*) at x representing the "grade of membership" of xin A. Thus, the nearer the value of fA(*x*) to unity, the higher the grade of membership of *x* in A.

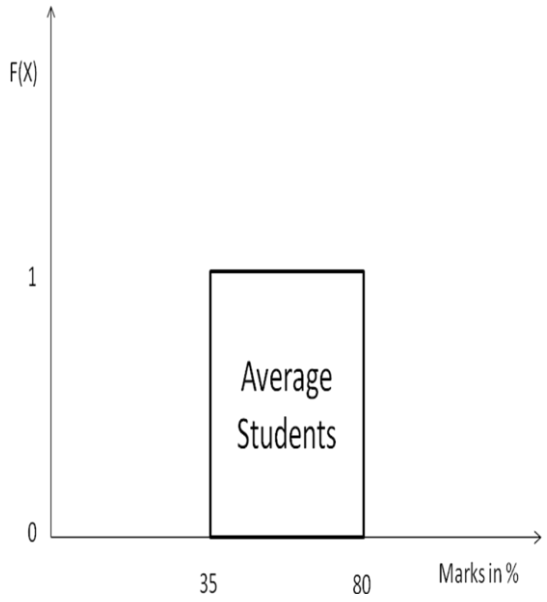

Fig. 1 Example of sharp boundary problem

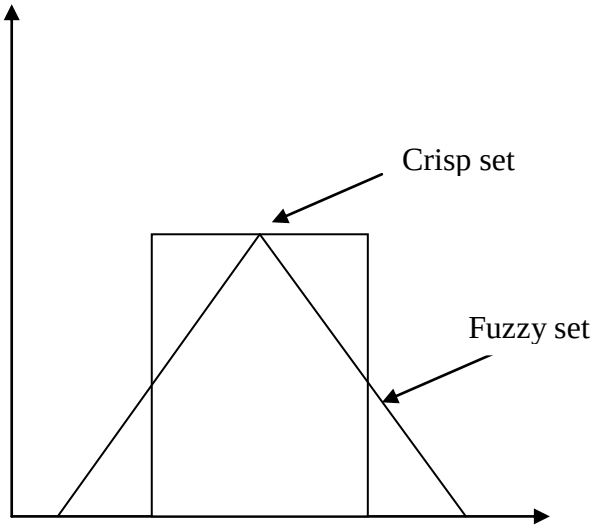

Fig. 2 Boundary of fuzzy set

VI. FUZZY LOGIC

Fuzzy logic is an enhance version of multivalve logic. However its uses are relatively different from their goals. Thus, the fact that fuzzy logic deals with approximate rather than precise 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 classical logical systems.

VII. FUZZY ASSOCIATION RULES

A nascent methodology based on classical association rule mining is Fuzzy Association Rules mining. Whenever data set having a certain range of values then it might possible to face the sharp boundary problem.

Suppose we have three range of marks of any examination.

F(x) is a function such that

0<X<=35 then f(X)	=	fail students
35<X<=80 then f(X)	=	average students
X>80 then f(X)	=	intelligent students

From the above scenario now suppose a student got 79.8% of marks then he is a average student. But if he got 80.1 marks then the tag will intelligent. But the student who got 79.8 % of marks is also an intelligent student, which is not the huge difference between both the conditions. Figure 1 show the scenario for boundary shape problem, which happened above.

There are some basic approaches to solve the sharp boundary problem.

1. Quantitative approach
2. Fuzzy Taxonomic Structures
3. Approximate Item set Approach

To resolve the sharp boundary problem by using Quantitative approach divide the variable marks into three fuzzy sets. The fuzzy sets and their membership functions will have to be defined by a domain expert. For easy demonstration, we will just define the borders of the sets and split the overlapping part equally between the so generated fuzzy sets. For an example, we will use the following borders for the fuzzy sets of the variable marks: Fail={0−35}, Average students={33−70}, intelligent students={70−∞}. The generated fuzzy sets is shown in Figure 1 . For all areas having no overlap of the sets,the support will simply be 1 for the actual itemset. If there is an overlap, the membership can be computed by using the borders of the overlapping fuzzy sets. The added support will here always sum up to 1.

Fuzzy set theory has been used more and more habitually in intellectual systems because of its simplicity

and similarity to human reasoning [18, 20]. Numerous fuzzy data mining algorithms for inducing rules from given sets of data have been designed and used to good effect with specific domains [17, 19, 21, 22]. 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 [22]. 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.

VIII. CONCLUSIONS

Temporal data mining have temporal association rule. These association rules have time terminologies in order to handle the problem of time sequence. The solution might be time specific or written for the window of the time. The objective of temporal mining is to discover temporal item-sets which are attributable to a considerable portion of the demand. Today intelligent systems used fuzzy set because of its simplicity and similarity to human reasoning. Where data mining with fuzzy set inducing rules from given sets of data have been designed and used to good effect with specific domains. We will present a frame work for fuzzy temporal data mining that generate effective fuzzy temporal association rule.

ACKNOWLEDGMENT

I would like to say thanks to my guide " Mr.Ganesh wayal" who gives their knowledge and time in order to complete this paper. This paper will never complete without the support facility member CSE department of RKDF institute of science & technology 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 26-281993.

[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. Agrawaland R, Shrikanth, "Fast Algorithm for Mining Association Rules". Proceedings Of VLDB conference, pp 487 – 449, Santigo, 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] Yao H., Hamilton, H. J. and Butz, C. J., "A Foundational Approach to Mining Itemset Utilities from Databases", In Proceedings of SIAM International Conference on Data Mining, pp. 482-486, 2004.

[8] Yingjiu Li, Peng Ning, Xiaoyang Sean Wang, Sushil Jajodia, "Discovering Calendar-Based Temporal Association Rules", In Proceedings of the 8th International Symposium on Temporal Representation and Reasoning, pp. 111–118, Cividale del Friuli, Italy,June 2001.

[9] Tarek F. Gharib , Hamed Nassar , Mohamed Taha , Ajith Abraham, "An efficient algorithm for incremental mining of temporal association rules", Data & Knowledge Engineering, Vol. 69, No. 8, pp. 800-815, 2010

[10] Cheng-Yue Chang, Ming-Syan Chen, Chang-Hung Lee, "Mining General Temporal Association Rules for Items with Different Exhibition Periods", In Proceedings of the 2002 IEEE International Conference on Data Mining (ICDM'02), pp. 59–66, Maebashi City, Japan, December

[11] Jen-Wei Huang, Bi.-Ru. Dai, Ming-Syan Chen, "Twain: Two-End Association Miner with Precise Frequent Exhibition Periods", ACM Transactions on Knowledge Discovery from Data, Vol. 1, No. 2, August 2007.

[12] Chang-Hung Lee, Ming-Syan Chen, Cheng-Ru Lin, "Progressive partition miner: an efficient algorithm for mining general temporal association rules", IEEE Transaction on Knowledge and Data Engineering, Vol.15, No.4, pp. 1004–1017, 2003.

[13] Ying Liu, W. K. Liao, and A. Choudhary, "A Fast High Utility Itemsets Mining Algorithm", In Proceedings of the 1st international workshop on Utility-based data mining, pp. 90 – 99, Chicago, Illinois, 2005

[14] Yao H., Hamilton H. J., Geng L., "A Unified framework for Utility based Measures for Mining Itemsets", In Proceedings of the Second International Workshop on Utility-Based Data Mining, Philadelphia, Pennsylvania, 2006.

[15] S.Shankar, T.Purusothaman, "Utility Sentient Frequent Itemset Mining and Association Rule Mining: A Literature Survey and Comparative Study", International Journal of Soft Computing Applications, No. 4, pp.81-95, 2009.

[16] Mathieu Guillame-Bert, James L. Crowley , "New Approach on Temporal Data Mining for Symbolic Time Sequences Temporal Tree Associate Rules" IEEE International Conference on Tools with Artificial Intelligence , 2011

[17] C. C. Chan and W. H. Au, "Mining fuzzy association rules," The Conference on Information and Knowledge Management, Las Vegas, pp. 209-215, 1997.

[18] P. A. Heng, T. T. Wong, Y. Rong, Y. P. Chui, Y. M. Xie, K. S. Leung and P. C. Leung, "Intelligent inferencing and haptic simulation for Chinese acupuncture learning and training," IEEE Transactions on Information Technology in Biomedicine, Vol. 10, No. 1, pp. 28-41, 2006.

[19] T. P. Hong, C. S. Kuo and S. C. Chi, "Trade-off between time complexity and number of rules for fuzzy mining from quantitative data," International Journal of Uncertainty, Fuzziness and Knowledge-based Systems, Vol. 9, No. 5, pp. 587-604, 2001.

[20] H. Ishibuchi and T. Yamamoto, "Rule weight specification in fuzzy rule-based classification systems," IEEE Transactions on Fuzzy Systems, Vol. 13, No. 4, pp. 428-435, 2005.

[21] C. Kuok, A. Fu and M. Wong, "Mining fuzzy association rules in databases," SIGMOD Record, Vol. 27, No. 1, pp. 41-46, 1998.

[22] Y. C. Lee, T. P. Hong and W. Y. Lin, "Mining fuzzy association rules with multiple minimum supports using maximum constraints", Lecture Notes in Computer Science, Vol. 3214, pp. 1283-1290, 2004.