International Journal on Future Revolution in Computer Science & Communication Engineering
Volume: 4 Issue: 4
ISSN: 2454-4248
797 – 800

A Review on: Efficient Method for Mining Frequent Itemsets on Temporal Data

Miss. Dhanashree L. Patil
M.Tech Student: CST
Department Of Technology, SU
Kolhapur, India
dhanupatil19@gmail.com

Mr. Hemant. A. Tirmare
Asst. Professor: CST
Department Of Technology, SU
Kolhapur, India
hat_tech@unishivaji.ac.in

***Abstract*—** Temporal data can hold time-stamped information that affects the results of data mining. Customary strategies for finding frequent itemsets accept that datasets are static; also the instigated rules are relevant over the whole dataset. In any case, this is not the situation when data is temporal. The work is done to enhance the proficiency of mining frequent itemsets on temporal data. The patterns can hold in either all or, then again a portion of the intervals. It proposes another method with respect to time interval is called as frequent itemsets mining with time cubes. The concentration is building up an efficient algorithm for this mining issue by broadening the notable a priori algorithm. The thought of time cubes is proposed to handle different time hierarchies. This is the route by which the patterns that happen intermittently, amid a time interval or both, are perceived. Another thickness limit is likewise proposed to take care of the overestimating issue of time periods and furthermore ensure that found patterns are valid.

Keywords-*Temporal data, data mining, frequent itemsets;*

*****

I. INTRODUCTION

Data mining is the procedure for finding out correct patterns and knowledge from large data sets. The main aim of data mining process is to mine knowledge from current dataset and change it into the human understandable form to advance use. There are various data analysis techniques available for research studies. These analysis techniques include temporal data mining, association rule mining, frequent itemset mining. The motivation behind this paper is to explore data mining techniques, which is suitable for finding frequent itemsets using temporal data.

It also uses synthetic datasets for showing a quite effective experiment result of proposed algorithm. It discovers frequent itemsets with respect to an optimal time period. The aim of this paper is to propose and design system for frequent itemset mining on temporal data. The proposed solution uses an algorithm of frequent itemset mining with time cube. This work proposes an efficient flow to find out interested or frequently utilized itemsets from the database. But discovered patterns may not be valid [1], because there is no any proof which will show that a particular pattern hold at particular time interval. This will arises the problem of overestimation [1]. So to avoid this problem new threshold is used called as density. Also, it refers recommendation for itemsets having high occurrence value or support.

II. BRIEF LITURATURE SURVEY

A New Methodology for Mining Frequent Itemsets on Temporal Data [1], the main aim is to develop an efficient algorithm by expending a priori algorithm. The idea of time cubes is used to handle time hierarchies. The patterns will consider the way by which the patterns that happen periodically, during a time interval or both, are recognized [1]. A new value of threshold as a density is also proposed to solve the problem of overestimating with respect to time periods and also make sure that discovered patterns are valid[1].

A Novel Approach for mining frequent itemsets: AprioriMin [2], which presents a new algorithm named AprioriMin Algorithm which mines the frequent itemsets [2]. This algorithm aims at introducing a new strategy for the calculation of frequent itemsets to prune phase of frequent itemsets. This technique decreases the execution time, when the support threshold increases for AprioriMin Algorithm [2].

Enhanced Frequent Itemsets Based on Topic Modeling in Information Filtering [3], this article illustrates a new model which integrates topic modeling and enhanced frequent itemsets mining together to alleviate the excellent strength on enhancing the information filtering as well as the reducing time consuming of the overall process [3].

Evolutionary algorithms and fuzzy sets for discovering temporal rules [4], a proposed technique is given for mining fuzzy association rules that have a temporal pattern. The result of search was considered by examining the temporal association rules and composition of fuzzy, it also combined a multi objective evolutionary algorithm with iterative rule learning to mine many rules[4].

Mining fuzzy periodic association rules [5], they develop techniques for discovering patterns with periodicity in this work. Patterns with periodicity are those that occur at regular time intervals, and therefore there are two aspects to the problem: finding the pattern, and determining the periodicity [5]. The difficulty of the task lies in the problem of

 ISSN: 2454-4248
797 – 800

discovering these regular time intervals, i.e., the periodicity [5].

Discovering frequent behaviors: Time is an essential element of the context [6], It relies on the extraction of frequent itemsets from usage databases. However, those databases are usually considered as a whole, and therefore, itemsets are extracted over the entire set of records. Our claim is that possible subsets, hidden within the structure of the data and containing relevant itemsets, may exist [6].

Discovery of Fuzzy Temporal Association Rules [7], they propose a data mining system for discovering interesting temporal patterns from large databases. The mined patterns are expressed in fuzzy temporal association rules which satisfy the temporal requirements specified by the user. Temporal requirements specified by human beings tend to be ill-defined or uncertain [7].

Twain: Two-End Association Miner with Precise Frequent Exhibition Periods [8], the investigation is on general model of mining associations in a temporal database, where the exhibition periods of items are allowed to be different from one to another [8]. The database is divided into small candidates or partitions according to the time granularity imposed. Prior work may omit some temporal association rules and thus have limited practicability [8].

On Mining General Temporal Association Rules in a Publication Database [9], it explores a new issue of mining general temporal association rules in publication databases. A publication database is a set of transactions, where Transaction T has a set of items with respect to their individual exhibition period [9]. The current model of association rule mining is not able to handle the publication database due to the fundamental problems as lack of consideration of the exhibition period of each individual item and lack of an equitable support counting basis for each item [9].

Mining general temporal association rules for items with different exhibition periods [10], it explored an new model of mining general temporal association rules from large databases where the exhibition periods of the items are allowed to be different from one to another [10].

Cyclic association rules [11], they have studied the issue of discovering association rules that show regular cyclic variation over time. Data regarding such variations can permit marketers to determine trends in association rules and facilitate better prediction [11]. By exploiting the connection between cycles and enormous itemsets, they identified improvement techniques that permit minimizing the number of wasted work performed throughout the data mining method [11].

Progressive partition miner: An efficient algorithm for mining general temporal association rules [12], they proposed an inventive algorithm as a Progressive Partition Miner (PPM) used to find out general temporal association rules in a publication database [12]. The basic notion of PPM is it initial partition the publication information with reference to exhibition periods of items then calculates the frequency of occurrence count of every candidate [12].

On the discovery of interesting patterns in association rules [13], they have studied the issue of discovering useful patterns within the variation of association rules over time. Data regarding such variations can enable analysts to determine trends in association rules and facilitate better prediction [13]. By learning the interaction between huge itemset detection and calendars, they devised a series of improvement techniques that considerably speed up the invention of calendar association rules [13].

Frequent Itemsets Mining Using Random Walks for Record Insertion and Deletion [14], they have proposed an incremental maintenance that can handle case of record insertion and deletion simultaneously. The proposed algorithm predicts prospective frequent itemsets by using random walk process and update frequent itemsets when transactions are added or deleted from database [14].

A novel approach for discovering retail knowledge with price information from transaction databases [15], traditional research on mining retail knowledge focuses on assortment planning, demand correlation analysis, and customers shopping behavior analysis. It does not take into account the prices of products, and how price setting can affect potential demands [15].

III. Choice Of The Topic With Reasoning

To find out frequent itemset from large database is time consuming. However, main database originate from transaction in a bank, supermarket or department stores that all related to time. This work is not created on temporal data in a perspective of data mining. So introduced new algorithm is frequent itemset with time cube. It is helpful to find out frequent itemsets on time-stamp data. The result of algorithm will helpful to make better decision.

IV. Outline of Proposed System

The system used temporal data transaction so first time is main attribute to be defined. The structure for time hierarchy helps to easily merge neighboring frequent itemsets to find out different temporal patterns [1]. To know time hierarchy and its domain, an equal-length partition or candidate generation is employed according to the minimum interval defined by the user. Each candidate represent time interval with different time hierarchy. The main parts of system are described as follows:

A. Data preprocessor

 ISSN: 2454-4248
797 – 800

Input data may contain noise or any irrelevant records. This module helps to reduce this noise. Also it requires converting or modifying input data into specific predefined format which is useful for implementation.

B. Support calculator

It takes preprocessed data as input and count occurrences to calculate support for each item. Also it convert item into itemset and generate support for that itemsets.

C. Candidate generator

It takes preprocessed data and its support as input to generate candidates. It uses candidate generation algorithm for generating candidates.

D. Frequent itemset mining

It takes input from support calculator and candidate generator to discover frequent itemsets. It uses frequent itemset mining with time cube algorithm. So it will find out frequent itemsets with respect to time interval.

V. Different work in Temporal Data Mining

Data mining is used for extracting important features from huge amount of data. The potential aim of data mining is nothing but it can be divided into different categories. It is also used for temporal data mining [18]. The categories are as prediction, classification, clustering, search & retrieval and pattern discovery. The work of those categories is considered with temporal data.

A. Prediction

The task of time-series prediction should do with prediction future values of the statistic supported its past samples. One has to build a prophetical model for the information. The most likely probably the earliest example of such a model is due to Yule way back in 1927 (Yule 1927).The auto-regressive family of models, for instance, is wont to predict a future worth as a linear combination of earlier sample values, provided the statistic is assumed to be stationary (Box et al 1994; Chatfield 1996; Hastie et al 2001)[18]. Linear non- stationary models like ARIMA models have conjointly been found helpful in several economic and industrial applications wherever some appropriate variant of the method is assumed to be stationary. Another well-liked work-around for non stationary is to assume that the statistic is piece-wise stationary. The series is then attenuated into smaller "frames" at intervals every of that, the stationary condition is assumed to carry then separate models square measure learnt for every frame. Additionally to the present customary ARMA family of models, there square measure several nonlinear models for statistic prediction [18].

B. Classification

Classification is the supervised learning technique. It is the task of generalizing known structure to apply to new data. This technique is uses the database to extract the classification rules. The behavior pattern of the data in the database is analyzed to develop the rules. Once the rule is developed, then the rules are used to predict the behavior of the upcoming transactions.

Classification techniques are designed for classifying samples using information provided by a set of samples. This set is usually referred to as a training set. The Training set may contain known or unknown samples. If given test data is wrong then it will create wrong rules. So it uses data cleaning. Data cleaning involves removing the noise and treatment of missing values.

C. Clustering

Clustering of sequences or time series is concerned with grouping a collection of time series based on their similarity. Clustering is of explicit interest in temporal data mining since it provides a main mechanism to automatically notice some structure in massive data sets that might be otherwise tough to summarize [18]. There are many applications where a time series clustering activity is relevant. For example in web activity logs, clusters can indicate navigation patterns of different user groups. In financial data, it would be of interest to group stocks that exhibit similar trends in price movements [18].

D. Search and retrieval

Searching for sequences in large amount of databases is another important category in temporal data mining. Sequence search and retrieval techniques play a crucial role in interactive explorations in huge successive databases [18]. The main issue is considered with expeditiously locating subsequences in large number of sequences. The best example of query-based searches is shown in study of automata theory and language [18].

E. Pattern discovery

This part will consider the temporal data mining category of task as a pattern discovery. The main origin or start of pattern discovery has in data mining only [18]. In pattern discovery, the main aim is to discover pattern by considering particular condition of that data attribute or feature.

VI. Conclusion

In this paper, we studied temporal data mining and frequent itemset mining using temporal data. The system will discover frequent itemsets along with their temporal information. Some itemsets occur during some time period while others may not occur to the particular time interval. The main feature of the proposed algorithm is that a new

 ISSN: 2454-4248
797 – 800

concept of time cube is presented to consider time hierarchies in data mining process. It will be used to discover different itemsets with their time of arrival. The density will be calculated to show that particular itemset is a valid for the specific time interval.

ACKNOWLEDGMENT

This work is done under the guidance and motivation of Mr. H. A. Tirmare. I would like to give special thanks of gratitude to my guide who gave me an opportunity to do work on the topic. Furthermore, I might want to thanks my parents and friends who supported me, helped me a ton in finishing this work

REFERENCES

[1] Mazaher Ghorbani and Masoud Abessi**,** "A New Methodology for Mining Frequent Itemsets on Temporal Data"IEEE Transactions on Engineering Management, Volume: PP, Issue: 99, Year-June 2017.

[2] Houda Essalmi, Mohamed El Far, Mohammed El Mohajir, Mohamed Chahhou, "A Novel Approach for mining frequent itemsets: AprioriMin" 2016 4th IEEE International Colloquium on Information Science and Technology (CiSt) Pages: 286- 289, Year: 2016.

[3] Than Than Wai, Sint Sint Aung, "Enhanced Frequent Itemsets Based on Topic Modeling in Information Filtering", 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS) Year-2017.

[4] S. G. Matthews, M. A. Gongora, and A. A. Hopgood, "Evolutionary algorithms and fuzzy sets for discovering temporal rules," Int. J. Appl. Math. Comput. Sci., vol. 23, no. 4, pp. 855–868, Year-2013.

[5] W.-J. Lee, J.-Y. Jiang, and S.-J. Lee, "Mining fuzzy periodic association rules," Data Knowl. Eng., vol. 65, no. 3, pp. 442–462, Year- 2008.

[6] B. Saleh and F. Masseglia, "Discovering frequent behaviors: Time is an essential element of the context," Knowl. Inf. Syst., vol. 28, no. 2, pp. 311– 331, Year- 2011.

[7] W.-J. Lee and S.-J. Lee, "Discovery of fuzzy temporal association rules," IEEE Trans. Syst., Man, Cybern., Part B, Cybern., vol. 34, no. 6, pp. 2330–2342, Year-2004.

[8] J.-W. Huang, B.-R. Dai, and M.-S. Chen, "Twain: Two-end association miner with precise frequent exhibition periods," ACM Trans. Knowl. Discovery Data, vol. 1, no. 2, 2007, Art. no. 8.

[9] C.-H. Lee, C.-R. Lin, and M.-S. Chen, "On mining general temporal association rules in a publication database," in Proc. IEEE Int. Conf. Data Mining, 2001, pp. 337–344.

[10] C.-Y. Chang, M.-S. Chen, and C.-H. Lee, "Mining general temporal association rules for items with different exhibition periods," in Proc. IEEE Int. Conf. Data Mining, 2002, pp. 59–66.

[11] B. Ozden, S. Ramaswamy, and A. Silberschatz, "Cyclic association rules," ¨ in Proc. IEEE 14th Int. Conf. Data Eng., 1998, pp. 412–421.

[12] C.-H. Lee, M.-S. Chen, and C.-R. Lin, "Progressive partition miner: An efficient algorithm for mining general temporal association rules," IEEE Trans. Knowl. Data Eng., vol. 15, no. 4, pp. 1004–1017, Jul./Aug. 2003.

[13] S. Ramaswamy, S. Mahajan, and A. Silberschatz, "On the discovery of interesting patterns in association rules," in Proc. 24th Int. Conf. Very Large Data Bases, 1998, pp. 368–379.

[14] Panita Thusaranon, Worapoj Kreesuradej, "Frequent Itemsets Mining Using Random Walks for Record Insertion and Deletion" 2016 8th International Conference on Information Technology and Electrical Engineering (ICITEE), Year: 2016.

[15] Y.-L. Chen, T. C.-K. Huang, and S.-K. Chang, "A novel approach for discovering retail knowledge with price information from transaction databases," Expert Syst. Appl., vol. 34, no. 4, pp. 2350–2359, 2008.

[16] R. Agrawal and R. Srikant, "Mining sequential patterns," in Proc. IEEE 11th Int. Conf. Data Eng., 1995, pp. 3–14.

[17] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques. Amsterdam, The Netherlands: Elsevier, 2011.

[18] Srivatsan Laxman and P S Sastry, "A survey of temporal data mining", Sadhana Vol. 31, Part 2, April 2006, pp. 173–198.