Review on Frequent Itemset Mining Overall Temporal Data

Ms. Jayashri N. Hajare[1], Prof. Mr. S. M. Rokade[2]
Department of Computer Engineering
[1,2] Pravara Rural Engineering College, Loni

Abstract- *Frequent pattern extraction is important technique in data mining domain. Lot of work has been done on pattern mining in literature. The work includes pattern extraction on static dataset. The whole dataset is used and frequent itemset are extracted. Temporal data is the data containing timestamp information. Timestamp based pattern extraction affects the results. Timestamp based pattern extraction method helps to analyze lifespan of itemset, peak period of purchase, etc. This work includes the analysis of existing system for temporal extraction of frequent patterns, its methodologies and limitations. After analyzing the existing work a new approach is proposed to overcome the existing system problem.*

Keywords- Data mining, Pattern extraction, frequent itemset mining, temporal data, multithreaded programming.

I. INTRODUCTION

Frequent pattern extraction is important technique in data mining domain. Transactions history analysis in shopping mall, banks, stock exchange, etc. is important application in data mining. Frequently occurred patterns in transactions of shopping mall represents frequently sold items. Analysis of such frequent item helps in making decisions such as product promotion, stock maintenance, stock placement, etc. Frequent patterns occurred in bank transactions or in stock exchange helps to predict future patterns. Frequent pattern analysis is also important in intrusion detection, web browsing history, bioinformatics, etc. To define frequent pattern, minimum occurrence count i.e. support value should be mentioned by the user. Itemset occurred more than minimum occurrence count is called as frequent itemset. Itemset is nothing but a co-occurrence of patterns.

All transactions are related to time. Dataset containing time information is called as temporal dataset. Along with occurrence count of transaction the occurrence time i.e. temporal information is again an important aspect. For e.g. transactions containing bread and butter are occurred frequently in the morning time than the whole day transactions. Such interesting patterns are time dependant. Temporal analysis of patterns extracts more useful information. This information includes frequent itemset purchase time span, lifespan of itemset, peak period of purchase, etc.

For frequent pattern extraction, algorithms such as apriori, FP-Growth, eclact, etc. are used. Agrawal et al. [10]. Initially proposes the co-occurrence pattern extraction technique. Apriori technique works in 2 phases. In first phase it identifies the frequent itemset candidate using download closure property and in second phase it filters the candidate patterns based on support value and generates the frequent pattern set. `FP Growth is tree based algorithm. It constructs conditional frequent item tree and uses depth first search for tree traversal and itemset extraction. Eclact algorithm also uses tree structure and depth search traversal technique for itemset extraction. These algorithms do not consider temporal aspect and works on complete dataset.

Sequential and time interval association rules, calendar rules are temporal frequent pattern analysis technique proposed in literature. In the following section related work to pattern discovery is discussed in detail with its techniques and limitations.

II. LITERATURE SURVEY

An association rules define the frequent patterns. Association rule discovery technique was first proposed by Agrawal et al.[10]. Association rules were extracted from 2 view data. Initially frequent items were extracted from each view and then dependency of co-occurance of frequent pattern in each view defines the association rules. The experiment was carried out over purchase history and association among products was identified.

Patterns those are occurred after every specific time interval is called as cyclic patterns. The patterns should occur without any exception. Ozden et al.[9] proposes a cyclic patterns extraction technique. This is a temporal technique. It identifies daily, monthly, quarterly and yearly patterns. The analysis of such cyclic pattern helps to analyze market trends and also helps in sale forecasting. For cyclic pattern extraction sequential algorithm and interleaved algorithm are used.

The extended version of cyclic pattern extraction is proposed by Ramaswamy[7]. This technique finds the user defined patterns in association rules. Calender algebra technique is used to detect patterns. But for execution calendar expression rules must be user defined.

Ale and Rossi[6] proposes a system that extracts frequent itemset over specific time period. The time period is shorter than the whole transaction history timespan. Life time of each item is used to define time interval. To extract temporal pattern a new algorithm is proposed based on apriori algorithm. The concept of temporal support is introduced first time in the algorithm.

In the market, purchase strategy varies very frequently. The product sold in one time span may not be sold with that frequency in next cycle. Hence cyclic patterns discovery over transaction dataset may not give the appropriate patterns. To overcome this issue, Han et al[8] proposes a periodic pattern discovery technique. The technique segments the time span in specific time period. The time period are fixed in length. The frequent patterns extracted from such periodic dataset gives good results for some period but not for all.

In real life scenario the exhibition period varies with respect to item. By considering the exhibition period of item, Progressive-partition-miner- PPM[5] technique is proposed by C. H Lee and M. S. Chen. The technique initially partitions the database in item exhibition period. Then it finds occurrence count of each candidate progressively using intrinsic partitioning characteristics.

Matthews et al. [2] proposes a new generic algorithm that traverses whole dataset and find temporal frequent patterns. It uses evolutionary algorithms in searching for finding association rules and defining optimization parameters. The technique simultaneously searches the rule space and temporal Space. Using this technique, temporal frequent patterns are extracted without prior partitioning the dataset.

Xiao et al. [3] proposes a technique to find association rule with time windows. The main aim of this system is to dynamically find association rules and its occurrence time windows. The time window is arbitrary in length and varies with respect to each association rule.

Seasonal product purchase behavior analysis technique is proposed by Saleh and Masseglia[4].This technique partition the database and analyse each part separately. The technique works on assumption that some products are sold in specific time period than a whole year like items purchased in winter are different than summer. Based on item frequency in transaction dataset, dataset is get partitioned for further analysis. Itemset support value increases with partitioned dataset as compared to the whole dataset and more frequent itemset are extracted for specific time window.

M. Ghorbani and M. Abessi[1] proposes a temporal frequent itemset extraction based on time intervals. A basic time cubes are defined by taking 3-D parameters such as day-month-year or, hour-day-month. The technique assumes that patterns are found in specific time interval or else it may exist over complete dataset. For temporal pattern extraction apriori algorithm is modified in which time cube based large one itemsets are extracted. The technique is time consuming because each time cube is processed separately.

III. ANALYSIS AND PROBLEM FORMULATION

There is need to develop frequent itemset extraction from a given dataset with temporal information. The item selling frequency varies with respect to time. The efficient system should be developed that finds frequent itemset over specific time period.

Problem Statement: To design, develop and test efficient frequent itemset extraction technique over temporal dataset.

IV. SYSTEM OVERVIEW

Following figure shows the architecture of the system:

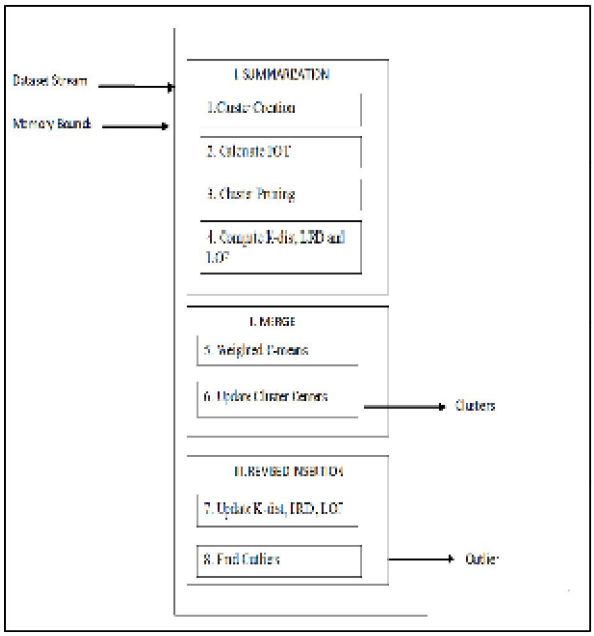

Fig: System Architecture

System extracts the frequent itemset from a given dataset with specific time interval .The dataset contains number of transactions with timestamp information. Transaction contains unique ids of items purchased in that transaction.

The dataset is partitioned in number of data cubes. The cubes are called as basic time cubes- BTC . These time cubes are generated based on the timestamp information. The partitioning strategy is initially defined by the user. For cube generation 3 dimensions are selected by user like: day-month-year, hour-month-year, hour-day, month, etc.

Apriori algorithm is used to find frequent itemset over each time cube. Initially large 1 itemset is generated and then based on support value frequent items are filtered. Processing each time cube is time consuming task. To improve system efficiency parallel processing is used. A multithreaded application simultaneously process time cubes and finally merged the results.

V. CONCLUSION

Lot of frequent itemset extraction technique with timestamp information are studied. After analyzing the existing system new approach is proposed. System finds frequent patterns over temporal dataset. The Basic time Cubes divides the temporal dataset in number of cubes. To find valid frequent items from time cube data cube density concept is proposed. Due to cube density overestimation of time period is avoided. Multithreaded application improves the system efficiency by parallel processing of time cubes data.

ACKNOWLEDGMENT

Authors would like to thanks Dr. Y. R. Kharde, Principal Pravara Rural Engineering College, Loni for their kind support and suggestions. We would also like to extend my sincere thanks to all the faculty members of the department of computer engineering and colleagues for their help.

REFERENCES

[1] Mazaher Ghorbani and Masoud Abessi, "A New Methodology for Mining Frequent Itemsets on Temporal Data", in IEEE Transactions on Engineering Management, Vol. 64, Issue. 4, pp. 566 - 573, Nov 2017

[2] S. G. Matthews, M. A. Gongora, and A. A. Hopgood, "Evolving temporal association rules with genetic algorithms," in Research and Development in Intelligent Systems XXVII. New York, NY, USA: Springer, 2011, pp. 107–120.

[3] Y. Xiao, R. Zhang, and I. Kaku, "A new framework of mining association rules with time-windows on real-time transaction database," Int. J. Innov. Comput., Inf. Control, vol. 7, no. 6, pp. 3239–3253, 2011.

[4] 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, 2011.

[5] 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.

[6] J. M. Ale and G. H. Rossi, "An approach to discovering temporal association rules," in Proc. ACM Symp. Appl. Comput.-vol. 1, 2000, pp. 294–300.

[7] 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.

[8] J. Han, W. Gong, and Y. Yin, "Mining segment-wise periodic patterns in time-related databases," in Proc. Int. Conf. Knowl. Discovery Data Mining, 1998, pp. 214-218.

[9] B.O zden, S. Ramaswamy, andA. Silberschatz, "Cyclic association rules," in Proc. IEEE 14th Int. Conf. Data Eng., 1998, pp. 412-421.

[10] R. Agrawal, T. Imielinski, and A. Swami, "Mining association rules between sets of items in large databases," ACM SIGMOD Rec., vol. 22, no. 2, pp. 207-216, 1993.