Discovery of Hierarchical Temporal Association Rules Using FP – Growth

Bharathwaj.S[1], Gowtham. N[2], Harish.C[3], Mrs.R.Anuradha[4]

Abstract-*In the First scan, it finds support for each item and sort the frequent item in descending order. In the second scan, two transactions are checked at a time and it is overlapped if it follows same path. In each transaction, the maximum common occurrence period is found and is included in the transaction database. From the results it is concluded that the proposed method for calculating lifespan using FP-Tree is efficient than Apriori algorithm.*

Keywords-Association Rule mining, Data Mining ,FP-Tree, Item Lifespan, Maximum Common Occurrence Period, Time granules.

I. INTRODUCTION

In data mining, association rules are useful for analyzing and predicting customer behavior. They play an important role in Market analysis , credit card fraud detection, genomic analysis. Association rules are created for analyzing frequent patterns by using the criteria of support and confidence to identify frequent item sets. Support is an indication of how frequently the items appear in the database. Confidence indicates the value that is found true in relation with support.

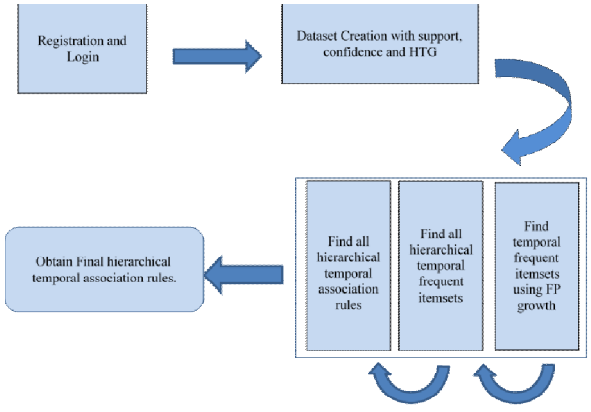
Temporal association rules adds time information on frequent item sets. It is used to describe valuable information among the items in temporal database. It helps us to find the associations among the item sets within its lifespan. In Apriori algorithm, the candidate generates lots of duplicate items and it makes process heavier and so lifespan cannot be predicted efficiently.

II. LITERATURE SURVEY

Association rule mining is a popular approach which can store possible combinations from dataset. It can store large data in form of arrays thus reducing the no of links [1]. According to Ke Sun and Fenshan Bai weighted association rule mining for itemset cannot be counted as transaction without the help of weighted support and it is prioritized in significance with dataset [2]. Multilevel association rules can be mined efficiently using concept hierarchies under a support-confidence framework. [3]Pratima has proposed large item sets level by level and then derive association rules from by using precise match and fuzzy match was used to predict the similarities between two calendar data [4] .Y Chen, K Tang proposed the usage of association rule mining in the stores for monitoring the customer behavior and also to find the frequent item set from the transaction database [5].Frans Coenen, Paul Leng proposed the usage of T tree and P tree with the Apriori algorithm to find the associations between two items [6].Association rules can also be used to identify the gene expression. It is applied in micro array gene expression separated into time series which was proposed by Ki lee and D lee [7]. The Temporal Data is used in the clinical database separated by time series and the values are marked in the particular columns [8]. Guojun Mao proposed the use of Temporal mining to store the associations of the Network traffic at different time intervals from the temporal network traffic database [9].The Incremental association rule proposed by Sultan along with the Association rule mining was used to extract an useful pattern in an incremental process to alter the changes in the previous model [10]. The sequential temporal association rules was used to store the electricity usage of residents using Apriori [11]. The concept of time granularity with temporal database was proposed by G.C.Lan [12]. Our work differs from [12] where we construct FP-tree with the split of time intervals to find the lifespan of each item from the transaction database.

III. SYSTEM OVERVIEW

This paper focuses on the performance of the transaction database and to reduce the database scan and time granules determines the lifespan of particular item. The modules begin with the user interface for login. The new user can register by filling the required details and the existing user can login with their own username and password.



Registration and login:

A user registration form is created in order to register and maintain the newly incoming user details. The register user can login into the system with the proper user details which they have registered already.

Dataset Creation:

In order to proceed the association rule mining the initial dataset has to be generated. The product transaction list is collected and their support, confidence and hierarchy with time granules HTG are computed for each transaction and store in the MySQL database, which is used as the dataset in the proposed project.

TFI and hierarchical temporal frequent itemset using FP-Growth and TPPI:

The temporal frequent itemsets within the intervals are first found, and then the item sets are identified as candidate temporal frequent ones in all the time granules of the upper level of the hierarchy. For Finding-Individual-TFI (Temporal Frequent Item set) FP Growth is adopted to derive frequent item sets from the transactions within a time period.

Hierarchical temporal association rules:

The possible candidate association rules are derived from the temporal frequent itemsets at each level. Their confidence values are then calculated and compared with the minimum confidence value to get the final temporal association rules.

IV. THE PROPOSED ALGORITHM

In the proposed system the transaction database is scanned only twice and the whole process will be completed within that scan period. By constructing FP-Tree, sub-nodes of two transactions are combined and formed into a single tree ,if the paths are equal.

Step-1: First the value is initialized to zero and each transaction is split into periods by using Hierarchical time granules. This table consists of two fields namely Items and Time periods.

Step-2: By using Frequent itemset , the total number of transactions is entered in the table and the FP-Tree constructed for each item.So the Hierarchical frequent itemsets is also kept empty at the beginning.

Step-3: For each period, the constructed FP-Tree is merged with the corresponding transaction and then similar itemsets are merged.

Step-4: The Maximum Common Occurrence Period (MCOP) is calculated for the combined FP-Tree . The Total occurrence of the item for all period is entered in the table.

Step-5: For each itemset , we calculate the relative support using the actual count of X and maximum occurrence of item in each period within the total transaction.

RS = Actual count of X + Maximum occurrence of item in each period / Total Transaction.

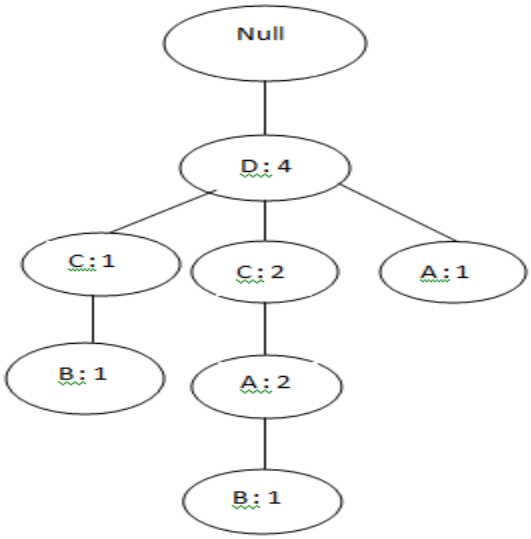
Step-6: Generate all itemset and check whether common period exist and include the relative support value and tabulate it.

Step-7: The relative confidence is calculated by the addition of relative support of two itemset by the relative support of the particular item.

Example of FP-Tree is given below
P1:
Minimum support = 2

T.ID	ITEMS	ORDERED ITEMS
Transaction 1	A,C,D	D,C,A
Transaction 2	A,B,C,D	D,C,A,B
Transaction 3	B,C,D	D,C,B
Transaction 4	A,D	D,A

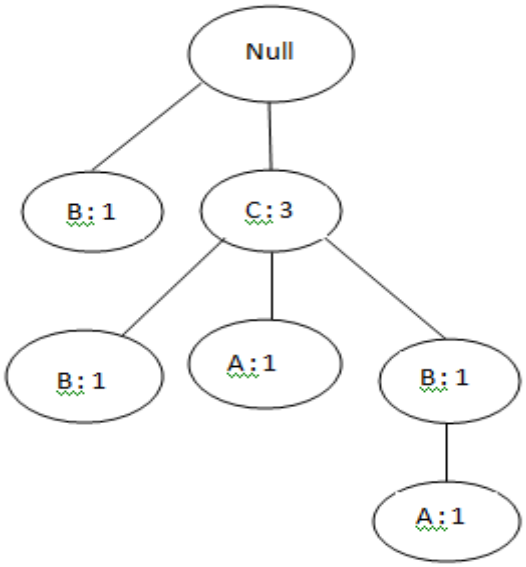
(A:3, B:2, C:3, D:4)



P2:
Minimum Support = 2

T.ID	ITEMS	ORDERED ITEMS
Transaction 5	B	B
Transaction 6	A,C	C,A
Transaction 7	A,B,C	C,B,A
Transaction 8	B,C	C,B

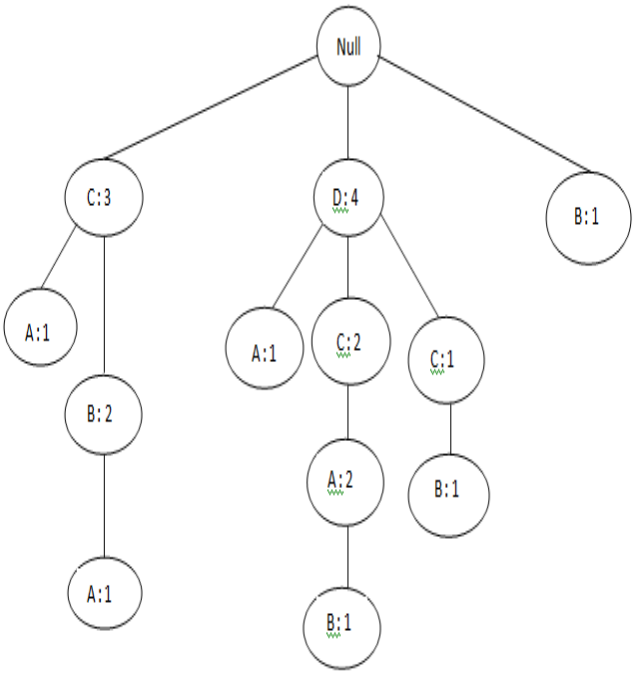
(B:3, C:3, A:2)



The combined FP-Tree within period P1 and P2 is given below :

(A:5, B:5, C:6, D:4)

P1, P2 :



V. CONCLUSION

In this paper , we mine the result by performing the database scan twice. The optimal support and confidence value can be minimized by using Maximum Common Occurrence Period. We used one of the efficient mining algorithms such as FP-Tree to mine the frequent itemset with the time granules.

REFERENCES

[1] Tzung-Pei Hong, Guo-Cheng Lan, Ja-Hwung Su, Pei-Shan Wub, Shyue-Liang Wang, Discovery of temporal association rules with hierarchical granular framework, Applied Computing and Informatics (2016) 12, 134–141 .

[2] Sheethal Abrahama, Sumy Joseph, A Coherent Rule Mining Method for Incremental Datasets based on Plausibility, Procedia Technology 24 (2016) 1292 – 1299 .

[3] Guo-Cheng Lan, Chun-Hao Chen, Tzung-Pei Hong, Shih-Bin Lin, A Fuzzy Approach for Mining General Temporal Association Rules in a Publication Database, 11th International Conference on Hybrid Intelligent Systems (HIS). 2011, pp. 611-615.

[4] Akash Rajak, and Kanak Saxena, Modeling Clinical Database using Time Series based Temporal Mining, International Journal of Computer Theory and Engineering, Vol. 2, No. 2 April, 2010. 1793-8201.

[5] Chun-Wei Lin, Tzung-Pei Hong*, Yi-Fan Chen, Tsung-Ching Lin, Shing-Tai PanJournal of Universal Computer Science, vol. 19, no. 4 (2013), 521-538.

[6] Frans Coenen, Paul Leng, and Shakil Ahmed , Data Structure for Association Rule Mining: T-Trees and P-Trees, IEEE transactions on knowledge and data engineering, vol. 16, no. 6, june 2004.

[7] Hojung Nam1, KiYoung Lee2 and Doheon Lee*1, Identification of temporal association rules from time-series microarray data sets, BMC Bioinformatics 2009, 10(Suppl 3):S6.

[8] Akash Rajak, and Kanak Saxena, Modeling Clinical Database using Time Series based Temporal Mining, International Journal of Computer Theory and Engineering, Vol. 2, No. 2 April, 2010 1793-8201.

[9] Ahmed Sultan Al-Hegami, A Framework for Incremental Mining of Interesting Temporal Association Rules, International Journal of Computer Applications (0975 – 8887) Volume 131 – No.8, December2015.

[10] Hông-Ân Cao, Tri Kurniawan Wijaya, Karl Aberer, Temporal Association Rules For Electrical Activity Detection in Residential Homes, 2016 IEEE International Conference on Big Data (Big Data).

[11] V.Vidya , Mining Weighted Association Rule using FP – tree , International Journal on Computer Science and Engineering (IJCSE).

[12] Guo-Cheng Lan1 , Chun-Hao Chen2 , Tzung-Pei Hong3, 4*, and Shih-Bin Lin4, A Fuzzy Approach for Mining General Temporal Association Rules in a Publication Database, 2011 11th International Conference on Hybrid Intelligent Systems .