Mining Frequent Patterns on Temporal Data using Basic Time Cubes(BTCs)

Shweta S. Holkar[1] & Sachin B. Takmare[2]
[1]ME Student, Department of Computer Science and Engineering, Bharati Vidyapeeth's College of Engineering, Kolhapur, Maharashtra, India.
[2]Professor, Department of Computer Engineering, CT's A P Shah Institute OBBC Technology, Thane (W), Mumbai, Maharashtra, India.

Received: February 13, 2019 Accepted: March 24, 2019

ABSTRACT: *Data mining focuses on using analytical and expert system models to forecast the future and discover the patterns among data. These patterns can help you to determine the further analysis in future. One of the most significant application of data mining is the analysis of transactional data. Transactional data is the data that frequently updated, it means it does not coexist over time so that every particular time interval new information becomes available. Holding patterns and some beneficial information in some of the time interval called temporal data. We wants to extract the patterns and some events that happens at the same time. We are trying to develop an algorithm to mine frequent patterns along with their temporal pattern with time cubes. Time cubes represents time interval with different time hierarchy like time cube (2, 4) Month having time interval between 2nd month to 4th month. In previous work time intervals are defined by the user. In their algorithm they are using basic time cube as an input may not be optimal. Here we are willing to try to reduce the user chosen parameter that is time cubes. To fulfill this concept we are trying to develop a metaheuristic algorithm. The purpose of algorithm is to analyses the database and accordingly generate the time cubes, then partitioning of database is done according to the time cubes and finally we mine valid frequent patterns and their related time interval on temporal data with the help of this time cubes.*

Key Words: *Data mining,Frequentitemset, metaheuristic algorithm, BTCs, hypercube.*

I. INTRODUCTION

In information mining discovering irregularity, examples and connections inside huge informational indexes to anticipate results. A standout amongst the most imperative utilizations of information mining is the examination of value-based information. Databases which created from operational framework. In this information contains time and numeric qualities. Fleeting databases that contain time-stepping data. Utilizing an extensive scope of methods, you can utilize this data to build incomes, cut expenses, improve client connections, and diminish dangers and that's just the beginning.

In this day and age everything is partner with web in a few or different ways. Individuals utilizing distinctive application that creating and playing with enormous measure of information. These producing of information is known as information mining. Information Mining is the way toward finding fascinating examples and usable data from a lot of information. For instance call detail records are ordinarily utilized by police and savvy benefits everywhere throughout the world. PDA go about as an essential substance in any of the wrongdoing scene. The database of call detail records containing time-stepping data. With the assistance of discovering specific records at specific wrongdoing time can assume an essential job in finding the conceivable suspects; along these lines, the time amid which they can be watched is vital. When we find substantial time shapes or time interim along these lines, that we can find the legitimate and productive example.

From the above we may presume that distinctive examples can be found on the off chance that we are viewed as various time interims. Mining such examples that held in a specific time interim may give us valuable data.

Essentially worldly information mining is appeared with the investigation of transient information and for discovering fleeting examples and regularities in sets of worldly information. Likewise fleeting information digging strategies take into account the likelihood of PC driven, programmed investigation of the information.

Some related examinations are mining affiliation rules with time-windows on continuous exchange database [5], timetable based transient affiliation rules [15], and mining affiliation manages in worldly databases [16], showcase bin investigation in a different store condition [17] lately.

The remainder of the paper is sorted out as pursues. In segment II we talk about some related work which are connected with our work. In segment III design graph and proposed calculation is displayed in subtleties. In area IV we present investigation on trial aftereffect of our calculation. In segment V we finish up the paper

II. LITERATURE REVIEW

GhorbaniMazaher, AbessiMasoud[1] have proposed calculation is exhibited to consider time chains of command in information mining process. It empowers us to discover various types of worldly examples. What's more, some minor improvements were presented. Another edge, called thickness, was proposed to mine substantial examples and tackle the issue of overestimating the timespans .Here the constraint of our calculation is that utilizing the underlying fundamental 3D squares may not be ideal.

F.Benites and E. Sapozhnikova[2] have proposed pairwise summed up affiliation rules mined from crude information can be utilized to interface the ideas of different ontologies. For this situation the things of guidelines are progressively composed a one can utilize the relations between them so as to lessen rule excess. In this they considering two progressive systems on both the precursor and the subsequent sides of a standard.

M. Shaheen, M. Shahbaz, and A. Guergachi[3] have proposed another way to deal with mine setting based positive and negative spatial affiliation leads as they may be connected to hydrocarbon prospection. Numerous specialists are as of now utilizing an apriori calculation on spatial database however this calculation does not use the qualities of positive and negative affiliation rules and of time arrangement examination, thus it misses the disclosure of extremely fascinating and valuable affiliations present in the information.

S. G. Matthews, M. A. Gongora, and A. A. Hopgood [4] have proposed hereditary calculation to discover worldly affiliation runs out of the blue. Hereditary calculation was utilized to all the while look through the standard space and transient space. They utilized the quality of developmental calculations in hunting down affiliation governs and streamlining parameters of enlistment process (backing and certainty esteems). They further broadened their works for finding fluffy fleeting affiliation rules.

Y. Xiao, R. Zhang, and I. Kaku [5] have proposed another type of affiliation rule, i.e., affiliation rule with time windows. The primary reason for their examination was to discover the time interims for affiliation rules which might be discretionary long and not client indicated. They further streamlined the way toward discovering time windows by scientific demonstrating.

B.Saleh and F. Masseglia [6] manage this issue from another perspective. The idea of strong itemset mining was proposed to discover the subsets of database that contain visit itemsets and afterward the proposed calculation was created dependent on that. Like above examinations, the test was to locate the ideal time interim.

B. Shen, M. Yao, Z. Wu, and Y. Gao [7] have proposed the issue of mining dynamic affiliation rule with comments (DAR-C). The remarks determine when to apply the standard. So as to formalize the issue, they exhibited the articulation technique for competitor viable occasions parcels, and afterward proposed two calculation to mine affiliation rules.

Y.- L. Chen and C.- H. Weng[8] have proposed how to apply the affiliation mining strategies to break down poll information. They present the poll information mining issue and characterize the standard examples that can be mined from survey information. A bound together methodology is created dependent on fluffy strategies with the goal that every unique datum types can be handle in a uniform way.

A.K.Mahanta, F.A.Mazarbhuiya, and H.K.Baruah [9] have displayed a technique which can separate diverse kinds of intermittent examples that may exist in a transient dataset. The client don't have to indicate the periods ahead of time. Set task called set superimposition was utilized for putting away periods related with thing sets.

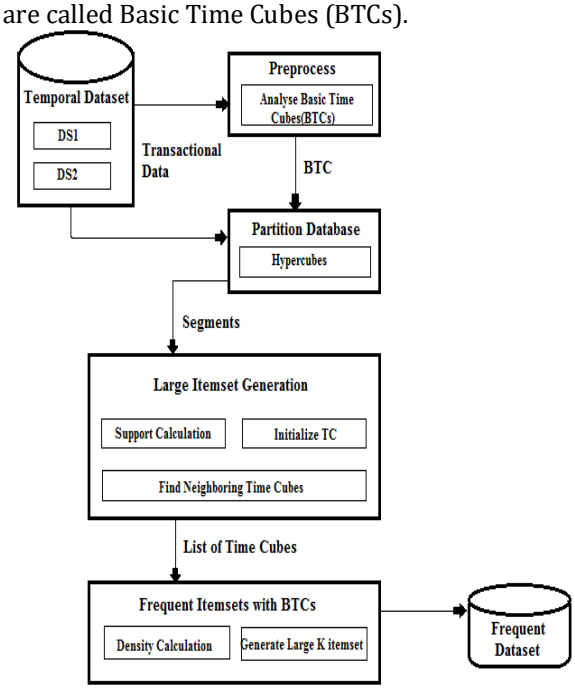
C.- H. Lee, J. C. Ou, and M.- S. Chen [10] have proposed distinctive presentation periods, dynamic segment excavator (PPM) calculation to find designs in databases. The fundamental thought of dynamic segment digger is to initially parcel the database in light of display times of things and afterward dynamically amass the event check of every hopeful 2-itemset dependent on the natural dividing qualities. They thinking about the significance of time interims.

C.- Y. Chang, M.- S. Chen, and C.- H. Lee[11] have proposed calculation fragmented dynamic channel (SPF) was acquaint with first portion the database into sub databases so that things in each sub database will have either beginning time or completion time. At that point, for each sub database, SPF continuously channels applicant 2-itemsets with total separating limits either forward or in reverse in time. Lee and Lee likewise utilized schedule variable based math to find affiliation rules. Since person will in general be questionable.

By inspecting diverse frameworks and strategies proposed by various creator, the proposed framework has been planned with transient dataset and segment database with regular thing set age.

III. SYSTEM ARCHITECTURE

In proposed engineering we are endeavoring to improve the proficiency of mining regular thing on worldly information. Since example can hold in either all or a portion of the interims. Calculation to limit time interim which is called visit item set mining with time blocks. The documentation of time solid shape is proposed to deal with time chain of command so thusly design that happen occasionally, amid a period interim or both are perceived. Knowing the time progressive system apportioning is utilized by the base interim characterized by the client. The underlying blocks are called Basic Time Cubes (BTCs).



The system has been divided into following different modules as –

Module 1: Preprocessing

In proposed work the dataset used for mining frequent itemsets is the dataset oftransactional data. The dataset required for this project is collected from various means.The dataset are collected from various websites searched on Google. Data in the real world is dirty meaning it is incomplete, noisy and inconsistent. For making quality decisions quality data is must require. Duplicate or missing data may cause incorrect or even misleading statistics. Therefore preprocessing is must require for good data mining results.

In this we have the dataset that contain the transactional records. So our proposed work is on the set of items we consider I= $\{i_1, i_2 \ldots\ldots i_n\}$ be the set of items, DB be a database of transactions. Each transaction $t_{rec}$is belongs to database (DB) and have associated with unique identifier called transaction ID(TID), a time stamp $t_{trec}$and a set of items. (t_{trec}belongs to T).

We consider here T is a total time span of the database. Let t_{st}, t_{et}belongs to T, where t_{st} is the start time and $t_{et}$is the end time be the time interval in each hierarchy. For example $(1, 3)_{Month}$ shows the interval between 1st and 3rd months. We consider here time cubes (TC) to represents time interval in between t_{st} and t_{et}. We use here cubic structure to represents time hierarchies. Basically cubic structure for time hierarchy helps us to easily generate frequent itemsets on temporal data within less time and with less comparisons.We implement the algorithm to generate the basic time cubes automatically according to the dataset. The output of our algorithm is the time cubes.

Module 2: Partition Database with BTC

In this proposed work we get the basic time cubes as an input, so that partitioning is done on database. In this module database is partition into small segments according to the basic time cubes. We wants to generate the Basic time cubes automatically to reduce the user chosen parameters. In previous system BTCs are given by the user. Our aim is fulfill in the preprocessing module we generate the BTC automatically. The small segments generated in this proposed work is act as an input for next module.

Module 3: Large Itemset Generation

In this module we implement the algorithm for mining large itemset. Large itemsets means those itemsets which appear sufficiently often in the database. This module takes segments as an input. For mining large itemsets we calculate support for itemset Y. The support of an item set Y can be defined as proportion of transactions in the data set which contain the itemset. This module takes dataset, minsupport and mindensity and segments from the previous module.

$$Support(Y) = NYCube/Ncube$$

Here, NCube be the number of transactions occurred in the cube (time interval), NYCube be the number of transactions that contains the itemset Y.

Also calculate density to overcome the overestimating problem of time period.

For generating large itemsets with list of tine cubes it satisfies the following conditions:

1. $Y \rightarrow^{cube} Z$ has support value greater than or equals to the minsupport value given by the user.
2. The time interval must be dense.

By using this we implement algorithm which gives us list of time cubes with the large itemsets. We also find here the neighboring Time cubes (TCs).

Module 4: Frequent Itemsetwith BTCs

Frequent itemset mining means finding the most frequent and relevant pattern in large dataset. In this module we generate the frequent itemsets with BTCs. In previous module list of time cubes are generated. This time cubes are act as an input for this module. We also calculate the density here to overcome the overestimating problem of time periods.

Density of a time interval is calculated as follows:

$$\text{Density} = \alpha * A$$

Here, α is the user specified parameter is called density rate. A is the average transaction per time cube. A is calculated as follows:

$$A = N/NTC$$

Here, N is the total number of records and NTC is the number of basic time cubes.

In this way we calculate density in this module. For mining frequent itemsets on temporal data with basic time cubes first we find candidate itemsets then calculate the support count for each candidate. With the help of this algorithm we candidate itemsets$C_k$is generated where k is the size of the itemset. The algorithm gives the candidate itemsets as an output. Higher level itemsets are generated by joining lower level itemsets. Then calculate support count for each candidate. After that we calculate the frequent itemsets with basic time cubes. Secondly we implement the algorithm to achieve our aim which satisfy the following conditions:

1. Support count of candidates is greater than or equal to the minimum support defined by the user.
2. Power set of time cubes that is P (TC) must be dense.

With the help of this we implement our algorithm which takes list of time cubes as an input and it gives frequent itemsets with time interval in less time as an output.

Algorithm for Frequent Itemsetwith BTCs:

Input: Database (D), Min sup, Min den, Basic time cubes(BTC)
Output: Frequent itemset with BTC Fbtc
1: Partitions database according to the BTC.
2:Initialize largeitemset L_1.
3: Repeat for all itemset Y belongs to I
4: For all segment of database belongs to D
5: Count support of Y
6: End for
7: Initialize TC
8: For all basic time cubes BTC
9: If {(sup(Y^{BTC})≥min sup)^(TR^{BTC}≥min den)} then
10: Set TC with union of TC_s and BTC_s.
11: Else
12: Set large itemset $L_1=L_1+Y^{TC}$
13: Initialize TC
14: End if.
15: End for.
For ($K = 2, LK-1$ _= ∅,K++)
16. Initialize C_k= ∅
17: For all pair of L_i,L_j belongs to Large itemset L_{k-1}.
18: Cand=$L_i \bowtie L_j$.
19: If Mod of Cand is equal to k then,
20: Put Cand into C_k.
21: End If.
22: End For
For all candidates $CT\ C \in CK$
23: Count Sup($CT\ C$)
24: End for
25:For all time hierarchies
26. LK={CTC∈CK|sup(CTC)≥min sup^P(TC)≥min den}
27: End for
28: End for
Fbtc= Fbtc$\cup LT\ C$

IV. EXPERIMENTAL SETUP

The instances are created using the windows 8.the java language has been used with NetBeans IDE.The data has been stored into MySQL database.

A. RESULT ANALYSIS

In table 1 considering time hierarchies. In result support and time is shown. In Table 1 shows time insec and support in percentage. Here whenever support increases time in second decreases. In

Table 1support against time

Support (%)	Time(s)
10	3586
15	1432
20	601
30	83
50	41
80	28

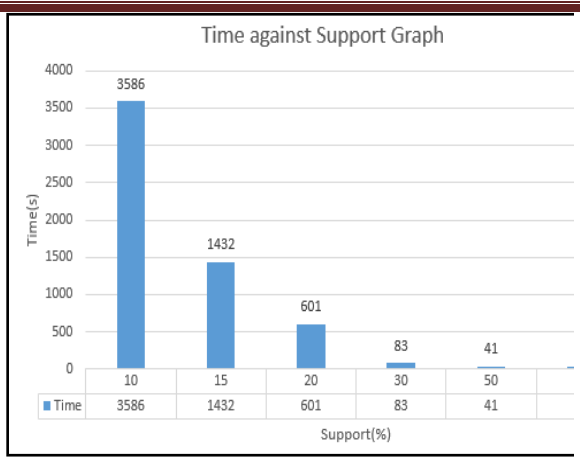


Fig. 2support versus time

Above fig 1 shows the solution time and minimumsupport relationship. As we can see, minimum support, at the maximum the time is shown in fig 2.

The obtained result from numerous runs of the algorithm with different number of transactions with different value of support.

V. CONCLUSION

We examine and analyze the features, opinion and reviews of different literatures and studied different techniques for mining frequent item set on temporal data. The main aspect of our proposed algorithm is methodical and effectively mine the frequent patterns on temporal data with basic time cubes and also this algorithm requires the minimum user chosen parameters. For extracting frequent patterns from uncertain database with basic time cubes. We use here the concept of time cubes which is related with the time fields and time hierarchy available in our database. We achieve our goal by using the concept of time cubes (TC). We also reduce the user chosen parameter that is basic time cubes by developing an efficient algorithm which automatically generates the time cubes. By this we obtained the better result. In this paper we achieved model for extracting frequent patterns on uncertain data that provide an adequately good solution.

REFERENCES

1. GhorbaniMazaher, AbessiMasoud "A new methodology for mining frequent itemset on temporaldata", in IEEE Transactions on Engineering Management, vol. 64 Nov 2017.
2. F. Benites and E. Sapozhnikova, "Hierarchical interestingness measures for association rules with generalization on both antecedent and consequent sides," Pattern Recognit. Lett., vol. 65, pp. 197–203, 2015.
3. 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, 2013.
4. M. Shaheen, M. Shahbaz, and A. Guergachi, "Context based positive and negative spatio-temporal association rule mining," Knowl.-Based Syst., vol. 37, pp. 261–273, 2013.
5. 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.
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, 2011.
7. .B. Shen, M. Yao, Z. Wu, and Y. Gao, "Mining dynamic association rules with comments," Knowl. Inf. Syst., vol. 23, no. 1, pp. 73–98, 2010.
8. Y.-L. Chen and C.-H. Weng, "Mining fuzzy association rules from questionnaire data," Knowl.-Based Syst., vol. 22, no. 1, pp. 46–56, 2009.
9. A. K. Mahanta, F. A. Mazarbhuiya, and H. K. Baruah, "Finding calendarbased periodic patterns," Pattern Recognit. Lett., vol. 29, no. 9, pp. 1274– 1284, 2008.
10. C.H. Lee, J. C. Ou, and M.-S. Chen, "Progressive weighted miner: An efficient method for time-constraint mining," in Advances in Knowledge Discovery and Data Mining. New York, NY, USA: Springer, 2003, pp. 449–460.
11. 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.
12. http://www.statsoft.com/Textbook/Data-Mining-Techniques
13. https://economictimes.indiatimes.com/definition/data-mining
14. https://en.wikipedia.org/wiki/Data mining
15. Y. Li, P. Ning, X. S. Wang, and S. Jajodia ,"Discovering calendar welbasedtemporal Association rules, " *Data Knowl. Eng.*, vol. 44, no. 2, pp. 193–218, 2003.
16. Xinfen Ye and J.A. Keane, "Mining association rules in temporal databases,"1998 IEEEInternational Conference on Systems, Man, and Cybernetics. (Cat.No.98CH36218).
17. Y.-L. Chen, K. Tang, R.-J. Shen, and Y.-H. Hu, "Market basket analysis in a multiple store environment," *Dec. Support Syst.*, vol. 40, no. 2, pp. 339–354, 2005.