International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 02 | Feb-2018 www.irjet.net p-ISSN: 2395-0072

MINING CO-OCCURENCE ELEMENT SETS ON SEQUENTIAL STATISTICS

RANJANI .G[1], SINDHU .S[2], SWATHI .V[3], Ms .RAJAVIZHI .N[4]

[1,2,3,4] *Dept of Information technology, Jeppiaar SRR Engineering College, Tamil Nadu, Chennai*

Abstract - *The* Databases present here are the Partial database and then the Vertical database. Temporary Products and the Permanent Products are combined and then later stored in a Vertical database. The admin is created for each and every product. The Product admin should have to create the transaction id for their products. The Product admin has to update their products with the transaction id to the common data owner. The product database will be in encrypted format. The Product admin gives the key to the data owner for viewing their product. The Product databases will be in decrypted format by using the key for the data owner. The data owner has to update their product in the horizontal database.

***Key Words*: Data mining, frequent itemset, frequent pattern, temporal data.**

1. INTRODUCTION

DATA MINING is the process of discovering interesting patterns and knowledge from large amounts of data One of the most important applications of data mining is the analysis of transactional data. The application is, Capturing the co-occurrence of items in transactions was first proposed. Databases originates from transactions in a supermarket, bank , etc. These are called temporal databases . It contains time-stamping information The time consuming part of the algorithm is discovering frequent item sets and generating association rules is straight forward. The main feature is , a new notion of TCs is presented to consider time hierarchies in data mining process. It develops an efficient algorithm to mine frequent patterns and their related time interval from transactional database.

LITERATURE SURVEY

TITLE: Web usage mining with evolutionary extraction of temporal fuzzy association rules

AUTHOR: Stephen G. Matthews, Mario A. Gongora , Adrian A. Hop good

YEAR: 2013

DESCRIPTION: In Web usage mining, fuzzy association rules that have a temporal property can provide useful knowledge about when associations occur. However, there is a problem with traditional temporal fuzzy association rule mining algorithms. Some rules occur at the intersection of fuzzy sets' boundaries where there is less support (lower membership), so the rules are lost. A genetic algorithm (GA)-based solution is described that uses the flexible nature of the 2-tuple linguistic representation to discover rules that occur at the intersection of fuzzy set boundaries. The GA-based approach is enhanced from previous work by including a graph representation and an improved fitness function. A comparison of the GA-based approach with a traditional approach on real-world Web log data discovered rules that were lost with the traditional approach. The GA-based approach is recommended as complementary to existing algorithms, because it discovers extra rules.

TITLE: Optimizing Frequent Time-window Selection for Association Rules Mining in a Temporal Database Using a Variable Neighborhood Search

AUTHOR: Yiyong Xiao, Yun Tian, Qiuhong Zhao

YEAR: 2014

DESCRIPTION: In this study, we investigate the problem of maximum frequent time-window selection (MFTWS) that appears inthe process of discovering association rules time-windows (ARTW). We formulate the problem as a mathematical model using integer programming that is a typical combination problem with a solution space exponentially related to the problem size. A variable neighborhood search (VNS) algorithm is developed to solve the problem with near-optimal solutions. Computational experiments are performed to test the VNS algorithm against a benchmark problem set. The results show that the VNS algorithm is an effective approach for solving the MTFWS problem, capable of discovering many large-one frequent item set with time-windows (FITW) with a larger time-coverage rate than the lower bounds, thus laying a good foundation for mining ARTW.

TITLE: Mining dynamic association rules with comments

AUTHOR: Bin Shen · Min Yao · ZhaohuiWu · Yunjun GAO

YEAR: 2012

DESCRIPTION: In this paper, we study a new problem of mining dynamic association rules with comments (DAR-C for short). A DAR-C contains not only rule itself, but also its comments that specify when to apply the rule. In order to formalize this problem, we first present the expression method of candidate effective time slots, and then propose several definitions concerning DAR-C. Subsequently, two algorithms, namely ITS2 and EFP-Growth2, are developed for handling the problem of mining DAR-C. In particular, ITS2 is an improved two-stage dynamic association rule mining algorithm, while EFP-Growth2 is based on the EFP-tree structure and is suitable for mining high-density mass data. Extensive experimental results demonstrate that the efficiency and scalability of our proposed two algorithms (i.e., ITS2 and EFP-Growth2) on DAR-C mining tasks, and their practicability on real retail dataset.

SCOPE OF THE PROJECT

Our focus is developing an efficient algorithm for this mining problem by extending thewell-known a priori algorithm. The notion of time cubes is proposed to handle time hierarchies. This is the way by which the patterns that happen periodically, during a time interval or both, are recognized. A new density threshold is also proposed to solve the overestimating problem of time periods and also make sure that discovered patterns are valid. We evaluate our algorithms via experiments.

EXISTING SYSTEM

The Recommended products will always be shown for the user. These Products are in a miscollapsed format. The Products cannot be mined by the data owner. The Data owner will buy every Products for the user. The Data owner will expense more money for their indeed products. More Preferences are given for the time interval products. The time intervals during which frequent patterns hold and the discovery of possible periodicities that patterns include. Traditional techniques for finding frequent item sets assume that datasets are static and the induced rules are relevant across the entire dataset. Frequent itemsets are discovered and those with neighboring time intervals are merged. Empirical experiments conducted on synthetic datasets to evaluate the performance of the algorithm.

1.2 LIMITATIONS

Less efficient to improve a frequent item sets on data.

User name will be created by the user.

The entire day will be needed for checking the product details.

Frequent Products cannot be mined.

All the indeed products will be shown for every users.

2. PROPOSED WORK

The Time Cubes(TCs) ,a new technique to analyse the profitable dataset. The more preferences will be given for the paired product. Frequent item set mining and the association rule mining are introduced here for the time saving purposes. The product admin will update their product details to the data owner with the transaction id. The Data owner will view the products using the key. Before viewing the products , the updated products will be in decrypted format. We Should use the key to convert the decrypted format into encrypted format. The Product Details will be stored in a horizontal database by the data owner.

TECHNIQUE USED

Frequent item set mining with time cubes

The objective of the proposed algorithm is to find item set X in a contiguous subset of database, where the support of X is above the minimum support and the size of the time interval is optimal. Also by introducing density threshold, validity of the rules is ensured (i.e., excluding time intervals which are not dense). Considering time hierarchy, our approach toward discovering frequent item sets is to first partition the database into many small segments. We use cubes (hyper cubes for time hierarchies more than three) to show these segments. Candidates that have support more than the minimum support in at least one TC are considered to be frequent. Neighboring TCs of the same item sets are merged if they are frequent.

ALGORITHM

Step 1: Generate the candidate item sets in C1

Step 2: Save the frequent item sets in L1

Step 3: Generate the candidate item sets in Ck from the frequent item sets in Lk-1

Step 4: Join Lk-1 p with Lk-1q,Insert into Ck .Selectp.item1, p.item2, . . . , p.itemk-1, q.itemk-1 from Lk-1 p, Lk-1q where p.item1 = q.item1, . . . p.itemk-2 = q.itemk-2, p.itemk-1<q.itemk-1

Step 5: Generate all (k-1)-subsets from the candidate item sets in Ck

Step 6: Scan the transaction database to determine the support for each candidate item set in Ck

Step 7: D, a transaction database

minsup_the minimum support current threshold

Step 8: if Tree contains a single path P then

foreachcombination (denoted a β) of nodes in the path P.

generate pattern βÜα with support_count=minimum support count of nodes in β.

else {

generate pattern β=ạÜα with support_count=ạ.support_count

}

Step 9: foreach frequent itemset ι do

foreach subset ς of ιdo

if(support (ι) / support (ι- ς)>=minconf) then output the rule(ι- ς)=>ς,

with confidence=support(ι)/support(ι- ς)

and support=support(ι)

ADVANTAGES

Solving the problem of overestimating time periods.

User name will be created by the Admin.

The data owner will take the product , which is more purchased during the time intervals.

The key will be given for the data owner by the particular product admin.

Miscollapse products will never occur.

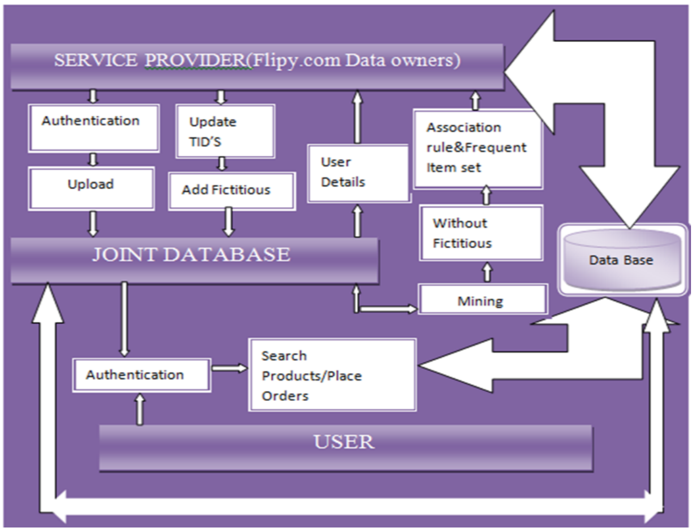

Fig -1: System Architecture

3. CONCLUSIONS

In this paper, we studied the mining of frequent item sets along with their temporal patterns. some patters are held during some time intervals, while others may happen periodically. The main feature of our proposed algorithm is that a new notion of Time Cubes (TCs) is presented to consider time hierarchies in data mining process. It enables us to find different kinds of temporal patterns .In addition , some error enhancements were introduced. A new threshold , called density, was proposed to mine valid patterns and solve the problem of overestimating the time periods .Experiments on synthetic datasets show that the proposed algorithm is quite effective.

REFERENCES

[1] Mazaher, M.Ghorbani, M. Abessi "A New Methodology for Mining Frequent Item on Temporal Data". Accepted May 27, 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] Yiyong Xiao, Yun Tian, Qiuhong Zhao "Optimizing Frequent Time-window Selection for Association Rules Mining in a Temporal Database Using a Variable Neighborhood Se arch" 2014

[4] Stephen G. Matthews, Mario A. Gongora , Adrian A. Hop good "Web usage mining with evolutionary extraction of temporal fuzzy association rules".2013

[5] Stephen G. Matthews, Mario A. Gongora , Adrian A. Hop good "Web usage mining with evolutionary extraction of temporal fuzzy association rules".2013

[6] 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.

[7] S. G. Matthews, M. A. Gongora, A. A. Hopgood, and S. Ahmadi, "Web usage mining with evolutionary extraction of temporal fuzzy association rules," Knowl.-Based Syst., vol. 54, pp. 66–72, 2013.

[8] Bin Shen · Min Yao · ZhaohuiWu · Yunjun GAO "Mining dynamic association rules with comments" 2012

[9] 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.

[10] 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.

[11] 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.