

Study on Efficient Algorithm and Differential Private Frequent Item set Mining

M.Banu priya

Department of Computer
Science and Engineering
Jansons Institute of
Technology
Coimbatore, India

Abstract: Frequent item set mining is to find the set of item occur frequently in the database Transactional database are insufficient to analyze the data in current shopping trends and dynamic dataset that update in data set. Discovering frequent item set play an important role in mining association rules, clusters ,web long mining and many other interesting pattern among complex data Efficient algorithm for analyze frequent item set based on the memory utilization and performance at the run time .Differential private FIM to find high data utility and high degree of privacy in the database

Keywords: Frequent item set (FIM), cluster association rule, Differential Private frequent item set

1. INTRODUCTION

Frequent item set play an essential role in many data mining task that try to find out interesting pattern from databases such as association rule, correlation ,sequences, classifiers and clusters .association rule helpful for analyzing customer behavior in retail trade, banking system etc.Frequent item set tries to find item set that occur in transaction more frequently than given threshold.FIM treat all the item having the same unit profit.Differential privacy offer strong privacy of released data without making assumption about an attacker background knowledge . Sequential pattern mining is define to finding statistically relevant pattern . The customer buying first a mobile phone,data cable and memory card if it occur frequently in a shopping history database is a sequential pattern. Association rule can be defined as $\{X,Y\} \Rightarrow \{Z\}$ the customer buys X,Y product and they like to buy Z..

2 RELATED WORK

2.1 Association rule

An association rules provide information in the form “if-then” statements. Association rules are computed from the data and unlike the if-then rules of logic but the association rule are probabilistic in nature .if 90% of transaction that purchase mouse and pen drive ,then also purchase headset.

Antecedent: mouse and pen drive

Consequent: headset

Confidence factor: 90%

As an addition to the antecedent(the “if” part)and the consequent (the “then” part) an association rule has two number that express the degree of uncertainty about the rule .Association analysis the antecedent and consequent are set of item.

of the item set are maintained in a tree-based structure called utility pattern tree. The construction of the UP-tree can be performed

	ASSOCIATION RULE	APERORI	FREQUENT PATTERN
Advantages	Predictability ,support	Easily parallelized	Only 2 pass over data set No candidate generation
Disadvantage	<ul style="list-style-type: none"> • Obtaining non interesting rules • Huge number of discovered rules • Low algorithm performance 	<ul style="list-style-type: none"> • Bottle neck problem • Requires many database scan 	<ul style="list-style-type: none"> • It is tree based structure when the database size increase when the node disconnect difficult to find the item set • It does not fit into the memory

FIGURE 2: ALGORITHM MERITS AND DEMERITS

2.2 Apriori Algorithm

This algorithm uses the association rule to find out frequent item set in transactional database. It uses bottom up approach for finding frequent item set and breath first search method .Hash tree to count the item set effectively

2.3 FP growth Algorithm

Frequent pattern algorithm is fragment growth method .it store all the item set in tree structure .FP growth algorithm use both horizontal and vertical database to store data set in main memory.FP growth algorithm is based on the depth first search method A frequent item pattern tree is a tree structure .

2.4 UP growth Algorithm:

Utility pattern growth algorithm for mining high utility item set. The information

With the two scan original database Step involved while construct UP-tree During First scan of the Database

STEP1: TU (Transaction Unit) of each transaction is computed

STEP2: TWU (Transaction Weight Unit) of each single item is also calculated

STEP 3: Discarding global unpromising items.

STEP 4: Unpromising item are removed from the transaction

STEP 5: The remaining promising item in the transaction are sorted in the Descending order TWU

Second scan

STEP 6: Transaction is inserted into UP (Utility Pattern)

3 DIFFERENTIAL PRIVACY

Differential privacy aim to publishing the frequent item set from the original database without providing the information about individual details. Differential privacy

ensures that the removal or addition of a single database item

Differential Privacy main concept is sensitivity. The amount of injected noise is carefully calibrated to the sensitivity .the Laplace Mechanism add noise drawn roughly from the Laplace distribution. To support multiple differentially private computation sequential composition guarantees overall privacy

4 EFFICIENT METHOD FINDING FREQUENT ITEM SET AND DIFFERENTIAL PRIVACY

Cheng –Wei Wu[3] :paper proposed lossless and compact representation named high utility item set(frequent item set).To mine the representation three algorithm are proposed AprioriHC(Apriority-based approach for mining high utility closed item set),AprioriHC-D (AprioriHC algorithm with discarding unpromising and isolated item) and CHUD(Closed High utility Item set).AprioriHC discard the global unpromising item and Isolated Item Discarding Strategy for finding itemset.AprioriHC perform breadth-first search in horizontal database and CHUD perform depth first search in vertical database .Derive High Utility Item set (DAHU) for efficiently recovering all the High utility item set from the CHUD

Vincent Tseng[4]:Proposed two efficient algorithm UP-growth and UP-growth+ for mining high utility item set from transactional database .In this algorithm construct UP-tree for maintaining information about high utility Item set .Potentially High Utility Item(PHUI) set have been generated from the UP-tree with only two scan of the original database

Freddy Chong Tat Chua [5]: Proposed a social correlation framework that incorporates a probabilistic social correlation into a latent space approach. Two generative models are used they are Sequential Generative model describe the social correlation matrix and Unified Generative Model. In this method efficient parameter estimation solution based on the expectation maximization. In this paper is focused on item adoption predication based on the social links Differential private transaction splitting methods

Sensu[6]: proposed Private FP-growth (PFP-growth) Algorithm. It consist of two phase they are preprocessing phase and mining phase .In the preprocessing phase improve the utility–privacy Tradeoff and smart splitting method to transform the database. In the mining phase a run time estimation method is used to estimate the actual support of item set Downward closure property used in dynamic reduction method to dynamically reduce the amount of noise added to provide privacy during the mining process

Cynthia Dwork:[7]:proposed privacy mechanism add noise to the transferring data item set. In this Gaussian noise are used instead of Laplacian.A finer analysis of realistic attack and to provide differential privacy are used sharpen the result and difficult to analyze the techniques used for transforming the data sets.

5 CONCLUSIONS

Frequent item set is very important to find out from the large data set. Online transaction has increases need to find out which item has frequently access.Privacy mechanism discussed adds an amount of noise in the data set. Summary of the in-depth analysis of few algorithm s is done which made a significant contribution to the search

of improving the efficiency of frequent item set mining algorithm

6 FUTURE WORK

In this paper, we done comparative study and analysis of efficient algorithm techniques, but these techniques have pros and cons, therefore there is necessity to develop such technique to overcome the entire disadvantage to find frequent item and to provide privacy accessing data from the database

7 ACKNOWLEDGEMENTS

The authors would like to thanks to the earlier work regarding different efficient algorithm method that contribute the work made in this paper .All work done in this paper will surely help to the researchers for future work on finding frequent item set

8 REFERENCES

- [1] R.Agrawal, T.Imelinski and A. Swami “Mining association rule between set of items in large databases” , in proceedings of the ACM SIGMOD Int’l Conf.on Management of, data pp 207-216 , 1993.
- [2] Rakesh Agrawal and Ramakrishnan Srikant , “Fast Algorithms for Mining Association rules” ,in Proceedings 20th Int.conf.very large Data Bases ,pp,487-499 ,1994
- [3]Cheug-wei wu, Philippe Fournier –viger, Philip S.Yu “Efficient Algorithms For Mining The Concise and Lossless Representation of Closed+ High Utility Item sets” Unpublished
- [4] Bai-En, Shie, Cheng-Wei Wu, Philip S.Yu and Vincent S.Tseng “Efficient Algorithm for Mining High Utility Itemset from

Transactional Database”, IEEE, Vol 25 No8 ,2013

[5] Freddy ChongTat Chua, Hady W.Lauw,Ee-peng Lim “Generative Models for Item Adoption using Social Correlation”, unpublished

[6]Sen Su,Shengzhi Xu,Xiang Chang,Zheng Li and Fangchun Yang “Differentially private Frequent Itemset Mining via Transaction Splitting”,IEEE 2014

[7]Denning,D.E.”Secure Statistical Databases Sample Queries “, ACM, 1998