Frequent Itemset Mining Using Transaction Splitting

**Vasifa Iqbal Mujawar, Sonali Baliram Goral, Shruti Sharad Patil , Shivali Yuvaraj Sawant ,
Teja Suresh Chandanshive**
Department of Computer Sciences, SBGI, Miraj, Maharashtra, India

ABSTRACT

Frequent itemsets mining (FIM) is a popular data mining technique and used in many important data mining tasks. To provide high data utility and privacy a new system is proposed in this paper which is divided into two phases. First phase takes database as an input, which consists of multiple transactions performed by different users, uses smart splitting method to limit the length of each transaction, and creates transformed database. Using this transformed database and threshold value next phase generates frequent itemsets.

Keywords : Frequent Itemset Mining(FIM), Transaction Database, User specified threshold

I. INTRODUCTION

In this paper I will give an overview about important subfield of data mining. Frequent Itemset Mining (FIM) is an important branch of data mining. It mainly focuses on observing the sequence of actions. In Frequent Itemset Mining, data takes the form of set of transactions called as transaction database, in which each transaction consisting of number of items. For this transaction database and user specified threshold value, the main task of frequent itemset mining is to find the complete set of frequent itemsets. The main goal of frequent itemset mining is to find regularities in the shopping behavior of customers. Using Frequent Itemset Mining Methods we can find set of products that are completely bought together. For Example, If a customer buys bread then customer will probably also buy butter, jam, cheese etc.

Frequent Itemset Mining plays an important role in many data mining tasks that are used to find an interesting pattern from database such as association rule mining, which is helpful in analyzing customer behavior in shopping market. Frequent Itemset Mining tries to find set of items that occur in transactions more frequently than the user specified threshold value. This method is used to find relevant item sets.

The many existing systems have the problem of tradeoff between utility and privacy in designing frequent itemset algorithms. Existing methods does not deal with high utility transaction itemsets Existing methods has large time complexity. Existing system gives comparatively large size output combination. To solve this problem, this project develops a time efficient differentially private FIM algorithm.

II. METHODS AND MATERIAL

A. Related Work

One method for frequent itemset mining is given in one paper [2] called as Matrix Based Algorithm With Tags(MBAT), which is based on transaction matrix and transaction reduction. From Transaction matrix this method finds frequent itemsets and this transaction matrix is generated from database. This method reduces number of candidate set generations. Another paper [3] provides two data structures to avoid excessive memory requirements and expensive disk accesses, CFP tree and CFP array. Using these data structures we can process large dataset in main memory. Another paper [4] gives one algorithm called as Frequent Pattern Tree which is very fast and tree based algorithm. In paper [7], Frequent itemset mining methods are classified into, horizontal data format, vertical data format and projected techniques.

B. Proposed Work

Proposed system is divided into two phases, In phase I , if size of transaction is greater than threshold value then the transaction is splitted into sub transactions using smart splitting method[1] and length of each sub-transaction is under the threshold value. This phase is carried out only once for given database. This phase creates Transformed database. In the phase II, the transformed database and a specified threshold value are given; then the system identifies frequent itemsets. The Transaction splitting might bring frequency information loss, a run-time estimation method[2] is used to estimate the original support of itemsets in the original database and to balance the information loss caused by transaction splitting.
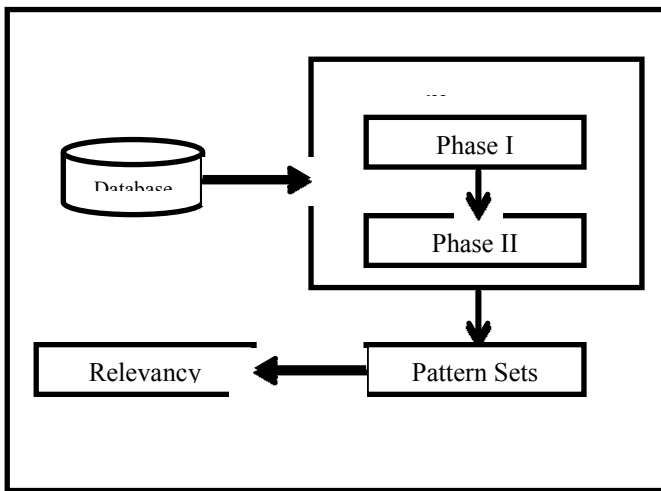

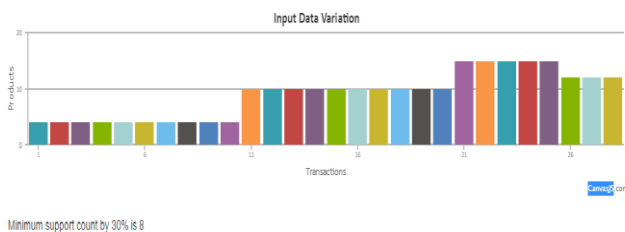**Figure 1.** System Block Diagram

III. RESULTS AND DISCUSSION
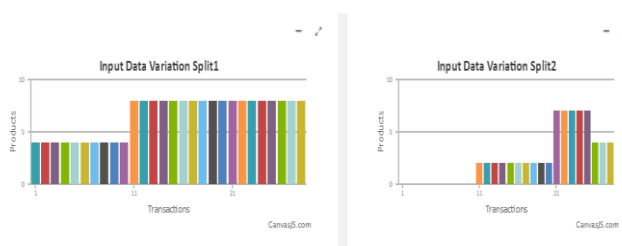

Figure 2. Input Database

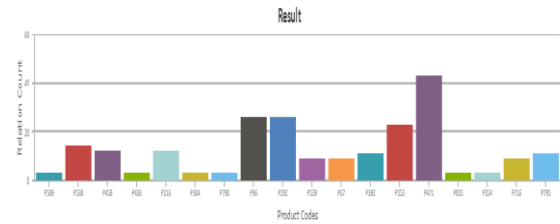

Figure 3. Phase I Result

Phase II

Figure 4. Phase II Result

IV.CONCLUSION

Differentially Private Frequent Itemset Mining is a method of mining frequent itemsets with the guarantee that the presence of any individual transaction data in the database does not reveal much more about that individual.

V. REFERENCES

[1]. Sen Su,Shengzhi Xu,Xiang cheng,Zhengyi Li,Fangchun Yang,"Differentially Private Frequent Itemset Mining via Transaction Spliting",IEEE 2015.

[2]. Harpreet Singh and Renu Dhir,"A New efficient Matrix Based Frequent Itemset Mining Algorithm with Tags",Manuscript received September 9,2012;revised octomber 12,2012.

[3]. Lorain Charlet Annie M.C. and Ashok Kumar D,"Market Basket Analysis For a Supermarket based on Frequent Itemset Mining",International Journal of computer Science Issues 2012.

[4]. Benjamin Schlegel,Rainer Gemulla,Wolfgang Lehner,"Memory-Efficient Frequent-Itemset Mining",EDBT 2011,March 22-24,2011,Uppsala,Sweden.

[5]. H.D.K.Moonesinghe,Moon-jung Chungh,Pang-Ning Tan,"Fast Parallel Mining of Frequent Itemsets",Department of Computer Science & Engineering Michigan State University East Lansing,MI48824.

[6]. Jiawei Han and Micheline Kamber,"Frequent Itemset Mining Methods",http://www.google.co.in/url?sa==t&source=web&cd=1&ved=0ahUKEwiT68GUtqjSAhUFjJQKHf8nBnQQFggeMAA&url=http%3A%2F%2Fwww-ai.cs.uni-dortmund.de%2FLEHRE%2FSEMINARE%2FSS09%2FAKTARBITENDESDM%2FFOLIEN%2FFrequent_Itemset_Mining_methods.pdf&usg=AFQjCNEz3aE8KbOv5IEIqI1KYEgZHtUBvg.

[7]. Sheetal Labade and Srinivas Narasim kini,"a Survey Paper on Frequent Itemset Mining Methods and Techniques",Internatinal Joural of Science and Research(IJSR) ISSN(Online):2319-7064.