International Journal of Research in Advent Technology, Vol.7, No.5S, May 2019 E-ISSN: 2321-9637 Available online at www.ijrat.org

Apriori Based Frequent Itemset Mining Algorithm

S.Sumathi

Dept of Computer Science, Arulmigu Kalasalingam College of Arts and Science, Krishnankoil sumathisindhu4@gmail.com

Abstract - Now a days, designing differentially private data mining algorithm shows more interest because item mining is most facing problem in data mining. To achieve privacy, utility and efficiency, Apriori based Frequent Itemset Mining algorithm is proposed which is based on the Frequent Pattern growth algorithm. It is divided into two phases namely preprocessing phase and Mining phase. The preprocessing phase consists to improve utility, privacy and novel smart splitting method to transform the database; the preprocessing phase is performed only once. The mining phase consists to offset the information lost during the transaction splitting and calculates a run time estimation method to find the actual support of itemset in a given database. Further dynamic reduction method is used dynamically to reduce the noise added to guarantee privacy during the mining process of an itemset.

Key Words: Itemset, Frequent Itemset Mining, Data mining.

1. INTRODUCTION

It is useful in most applications like decision support, Web usage mining, bioinformatics, etc. In a given a database, each transaction consists a set of items, FIM tries to find itemset that occur in transactions multiple times

A variety of algorithms are already implemented for mining sequence itemsets. The Apriori and FPgrowth algorithm are the two mostprominent ones. In particular, Apriori algorithm is a breadth-first search algorithm. It needs '1' database scans if the maximal length of frequent itemsets is '1'.

FP-growth algorithm is a depth-first search algorithm, which requires no candidate generation. While FP-growth only performs two database scans, which makes FP-growth algorithm an order of magnitude faster than Apriori. The features of FP-growth inspire to design a differentially private FIM algorithm based on the FP-growth algorithm. t has been shown that utility privacy can be improved by reducing the length of transactions.

2. RELATED RESEARCH

A. Data Mining

Data mining[2][3]is the process of finding useful information that is not easily exposed in vast amounts of data. It is a methodology for finding patterns and trends of specific types that extract patterns from data and generate models. A particular data model forms a kind of cluster that describes the relationship between data sets. In other words, the data is analyzed by the method that easily refines data, statistically analyzes and presents hypotheses.

B. Association Rule

Association rule learning[4] is a rule-based machine learning method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using some measures of interestingness. It can find frequent patterns, associations, correlations, or causal structures among sets of items in transaction databases. By using associations, it is possible to understand customer buying habits and correlations between the different items that customers place in their shopping basket.

3. PROPOSEDSYSTEM

A. Problem Definition

To design the algorithm, which is divided into two phases namely Preprocessing phase and Mining phase. The preprocessing phaseconsists to improve utility, privacy and novel spliting method to transform the database, it is perform only one time. The mining phase consists to offset the information lost during the transaction splitting and calculates a run time estimation method to find the actual support of itemset in a given database.

1.Splitting phase

In a splitting phase the long transactions are splitted rather than truncated. It is nothing but dividing long running database transactions into more than one subset.Given a transaction $t = \{a; b; c; d; e; f\}$

Instead of processing transaction t solely, divide t into $t1 = \{a; b; c\}$ and $t2 = \{d; e; f\}$. Doing so results in to the support of itemsets $\{a; b; c\}$, $\{d; e; f\}$ and their subsets will not be affected.

2. Run-time Estimation

This method finds weights of the sub transactions. While splitting the transactions there is data loss. To overcome this problem, a run-time estimation method is proposed. It consist of two steps: 1) first estimate its actual support in the transformed database, and 2) then compute its actual support in the original database.

International Journal of Research in Advent Technology, Vol.7, No.5S, May 2019 E-ISSN: 2321-9637 Available online at www.ijrat.org

3.Dynamic Reduction

Dynamic reduction is the proposed lightweight method. The main idea is to leverage the downward closure property (i.e., the supersets of an infrequent itemset are infrequent), and dynamically reduce the sensitivity of support computations by decreasing the upper bound on the number of support computations.

4. CONCLUSION

The formal privacy analysis and the results of extensive experiments on real datasets show that Apriori based frequent itemset mining algorithm is time-efficient and can achieve both good utility and good privacy.

REFERENCES

[1] Shailza Chaudhary, Pardeep Kumar, Abhilasha Sharma, Ravideep Singh, "Lexicographic Logical Multi-Hashing For Frequent Itemset Mining", International Conference on Computing, Communication and Automation (ICCCA2015)

[2] Lei Xu, Chunxiao Jiang, Jian Wang, Jian Yuan, Yong Ren,"Information Security in Big Data: Privacy and Data Mining", 2014 VOLUME 2, IEEE 29th International Conference on Information Security in Big Data

[3] O.Jamsheela, Raju.G, "Frequent Itemset Mining Algorithms : A Literature Survey", 2015 IEEE International Advance Computing Conference (IACC)

[4] Washio T, Nakanishi K, Motoda H (2005) Association rules based on levelwise subspace clustering. In:Proceedings. of 9th European conference on principles and practice of knowledge discovery in databases.LNAI, vol 3721, pp. 692–700 Springer, Heidelberg

[5] Yang Q, Wu X (2006) 10 challenging problems in data mining research. Int J Inform Technol DecisMaking 5(4):597–604