MJRET
Open Access
Multidisciplinary Journal of Research in Engineering and Technology
Journal homepage: www.mjret.in
ISSN:2348 - 6953

A NEW ENERGY EFFICIENT VERTICAL HANDOVER ALGORITHM IN HETEROGENEOUS NETWORKS

Akanksha Bhalerao-Kulkarni, Prof. Soumitra Das

Department of Computer Engineering, Savitribai Phule Pune University
Dr. D.Y. Patil College of Engineering, Charholi, Pune
Pune, India

Abstract: *The mining of successive examples is a central part in numerous information mining undertakings. A lot of examination on this issue has prompted a wide arrangement of efficient and scalable algorithms for mining frequent patterns. On the other hand, discharging these examples is posturing worries on the protection of the clients' participating in the data. In this proposition, we examine the mining of successive examples in a protection saving setting. We propose an approach for differential private frequent itemset mining which is based on LCM algorithm; we refer it as P-LCM algorithm. P-LCM is extended version on PFP growth algorithm which basically has two phases such as preprocessing and mining phase. The preprocessing phase needs to be performed only once and smart splitting method is used in this phase for improving utility as well as privacy trade off. Second phase limits the information loss caused by splitting as well as reduces the amount of noise added during mining process. In addition we propose three algorithms LCMfreq for mining all frequent sets instead of PFP-growth. LCM finds all frequent item sets in polynomial time per item set, and at the same time doesn't store prior obtained closed item sets in memory. The computational experiments on real world and synthetic databases suggest on comparing their performance to the previous algorithms, that LCM algorithms are faster on large real dataset especially in case of high degree of privacy, high utility and high time efficiency.*

Keywords*: PFP-growth, LCMfreq, Differential Privacy, Frequent Itemset Mining, P-LCM.*

1. INTRODUCTION

Currently, associated with the increased use of networking and the ability to collect confidential private data, privacy has become a serious concern for all users. In this paper, the focus lies on privacy issues that arise with regards to find frequent itemsets in "transactional" data. Frequent itemset mining is widely used across applications, and mostly in market basket analysis. The motive behind the study of frequent itemset mining in market basket- analysis is to find sets of items that are frequently purchased together, which is

helpful information in applications for product placement to marketing and beyond. Currently the talk of the town is to develop time efficient algorithms for frequent itemset mining.

Eventually the area of differentially private approach to frequent itemset mining has been away from the eyes of researchers by far until for the recent work done [1]. A frequent itemset mining algorithm is given input a dataset of the transactions by a group of individuals, and it results the frequent itemsets mined as output. But this creates an alert for privacy check immediately. Also the concern raised is regarding how can there be surety that publishing the frequent itemsets in the dataset does not reveal any confidential information about the users whose data is being studied. Another aspect relating to this problem is that we may not be aware about the data the individual users would like to protect. Also we are clueless regarding the background information possessed by an attacker. These compounding factors find their solutions in differential privacy [1].

This paper works into two prime phases of Pre-processing Phase and Mining Phase. The pre-processing phase is carried out only once and its output is subjected to Mining Phase. The final output of Mining Phase is the frequent itemsets from a collection of input data.

2. LITERATURE SURVEY

- **C. Dwork [1],** Author gives a general result which depicts that a formalization of Dalenius' goal as stated in semantic security is not practical. Instead, a variant of the result is a hazard to the privacy of a non-database participant. Authors suggest a novel measure known as differential privacy, which spontaneously captures the increased risk to one's privacy incurred by participating in a database. Most recent papers show that these new techniques can achieve any desired level of privacy under this measure. These measures ensure that accurate information regarding the database can be provided along with high level of privacy maintained.

- **Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam [3]**, Authors recognized severe privacy issues with k-anonymized dataset by studying two simple attacks. Firstly, Authors showed that an attacker can discover the values of sensitive attributes when there is little diversity in them. Secondly, attackers do possess some background knowledge. This shows that k-anonymity does not guarantee complete privacy preservation against such attackers. Thus authors studied both these attack mechanisms in detail and developed an analysis which leads to proposing a unique and powerful privacy definition known as ℓ-diversity. Authors efficiently cultivated a formal foundation for ℓ-diversity, and proved with experimental evaluation that ℓ-diversity is practical and its implementation derives worthy results.

- **Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke [11],** A framework for mining association rules from transactions consisting of categorical items having data subjected to randomization for privacy preservation of individual transactions has been presented by Authors. On one hand it is quite feasible to recover association rules and maintain privacy using a straightforward ‘‘uniform’’ randomization but these the discovered rules on other hand can lead to being exploited to find privacy

breaches. The nature of privacy breaches along with a proposed class of randomization operators are analyzed and have proven more effective than uniform randomization in limiting the breaches. They derive formulae for an unbiased support estimator and its variance, which allow us to recover itemset supports from randomized datasets, and depict their working into mining algorithms. Authors have successfully presented experimental results validating the algorithm by applying it on real datasets.

- **N. Li, W. Qardaji, D. Su, and J. Cao [14],** In this paper authors studied the problem of performing frequent itemset mining on transaction databases while satisfying differential privacy. We propose an approach called PrivBasis, which influences a fresh notion called basis sets. A θ-basis set has the property that any itemset with frequency higher than θ is a subset of some basis. They introduce algorithms for privately constructing a basis set and then using it to find the most frequent itemsets. Experiments show that our approach greatly outperforms the current system.

- **C. Dwork, F. McSherry, K. Nissim, and A. Smith [16],** Authors continue a line of research initiated on privacy-preserving statistical databases. Initially a trusted server with a database of sensitive information is considered to derive results. Consider Given a query function *f* mapping databases to reals, the so-called *true answer* is the result of applying *f* to the database. In order to protect privacy, the true answer is distributed by the adding random noise generated as per chosen distribution, and the total sum of the true answer plus noise, is returned to the user. Earlier work focused on the case of noisy sums, in which $f = \sum_i g(x_i)$, where x_i denotes the *i*th row of the database and *g* maps database rows to [0,1]. We extend the study to general functions *f*, which proves that privacy can be maintained by calibrating the standard deviation of the noise according to the *sensitivity* of the function *f*. This is the amount that any single argument to *f* can change its output. The new analysis shows that for several particular applications substantially less noise is needed than was previously estimated. The first step characterizes privacy in terms of distinguishability of transcripts. The separation results show the increased value of interactive sanitization mechanisms which establish the above.

3. PROPOSED WORK

In this project Authors propose the PFP growth algorithm, which satisfies differential privacy [1]. The PFP-growth algorithm is combination of a preprocessing phase and a mining phase in totality. The first phase being the preprocessing phase, it aims to improve the utility and privacy balance. To do so a new smart splitting method is proposed to transform the database. For a given database under consideration, the preprocessing phase is a one-time activity. Next is the mining phase, under which the information loss is caused by transaction splitting, hence we devise a run-time estimation method which estimates the actual support of itemsets in the original database.

In this paper, we investigate an efficient algorithm LCM for enumerating all frequent closed item sets instead of PFP-growth. LCM is an abbreviation of ***L****inear time* ***C****losed item set* ***M****iner*. Existing algorithms enumerate frequent item sets with cutting off unnecessary frequent item sets by pruning. However, if pruning is not complete, the algorithms operate on

unnecessary frequent item sets and may lead to data loss. In LCM, we define a parent-child relationship between frequent closed item sets. This relationship induces tree-shaped transversal routes composed of all the frequent closed item sets only. Our algorithm traverses the routes taking linear time of the number of frequent closed item sets. This algorithm is obtained from the algorithms for enumerating maximal bipartite cliques, which is designed based on reverse search technique. The result of computer experiments performed on real and artificial datasets using the previous algorithms depict that LCMfreq significantly outperforms the above.

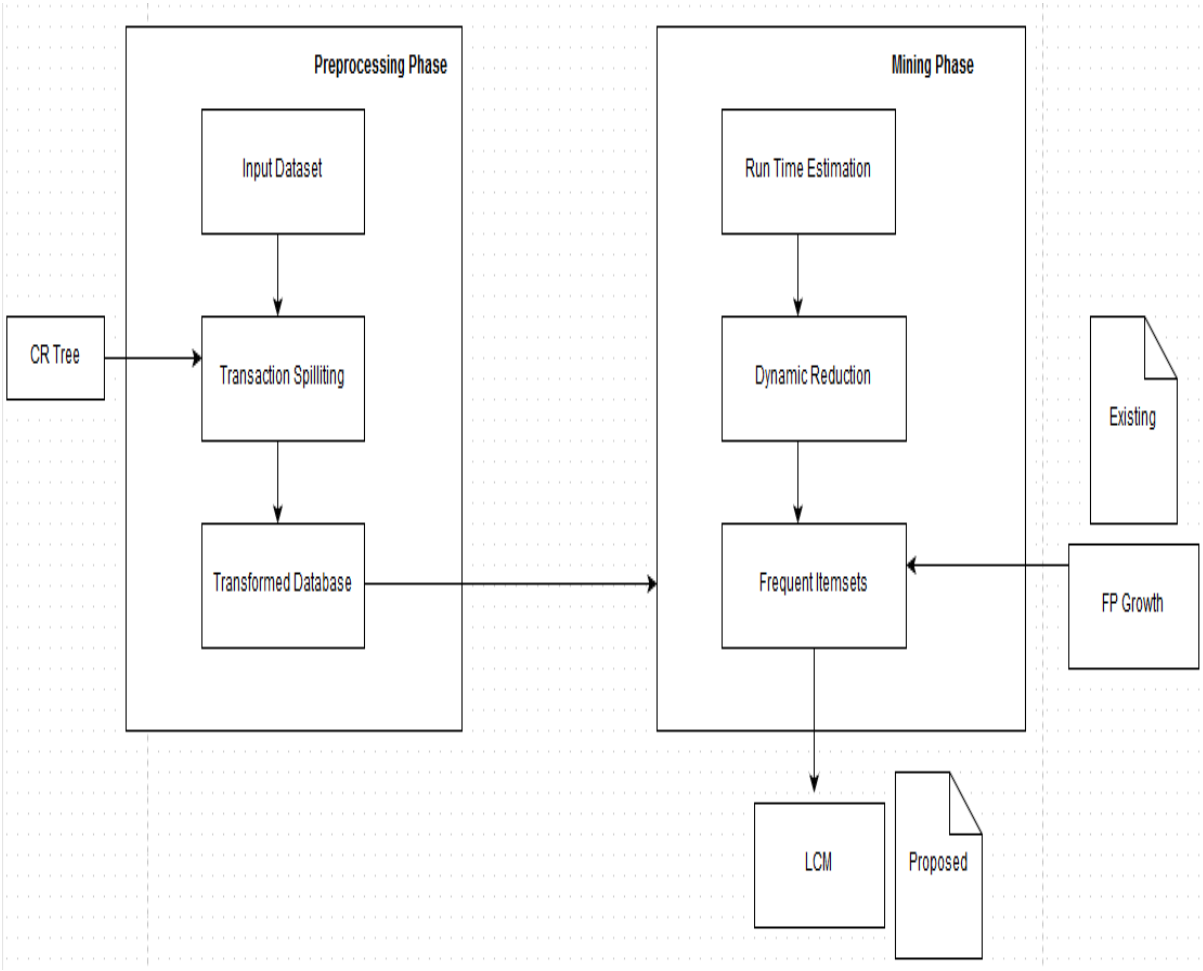

Fig.1: Architecture of the Project (Copyright Akanksha Bhalerao)

4. EXPECTED RESULTS

We propose the LCMfreq algorithm instead of FPF-growth algorithm as worked upon by Authors. LCMfreq algorithm is expected to reduce the net processing time of finding frequent itemsets as well as enhance the quality of operation. Our study proves that LCM is far accurate while comparing with existing Frequent Itemset Mining techniques such as FP-growth.

5. CONCLUSION AND FUTURE WORK

The need for designing differentially private data mining algorithms has seen growth as for frequent itemset mining purposes. It is the backbone of Data Mining. The most traditional and not much effective algorithms have been the cause behind this development. Thus through this project we intend to provide better and time saving results of frequent itemset mining along with maintaining the security of long transactional datasets. An effort to considerably replace the traditional FP-growth algorithm with LCM algorithm is tested for results. The concept of Differential Privacy, Transaction splitting and Run Time Estimation are studied in depth. Our future work will be to apply same techniques on high dimensional dataset of transactions.

ACKNOWLEGEMENT

It gives us immense pleasure to thank MJRET for being a wonderful platform for the research students to publish their work. Also extending Thanks to our guide & teaching staff for their constant support and motivation. Our sincere gratitude to our learning institution; Dr. D.Y.PATIL SCHOOL OF ENGINEERING CHARHOLI, PUNE for providing a strong platform to develop our skill and capabilities for research work.

REFERENCES

[1] C. Dwork, "Differential privacy," in Proc. Int. Colloquium Automata, Languages Programm., 2006, pp. 1–12, http://link.springer.com/chapter/10.1007%2F11787006_1

[2] L. Sweeney, "k-anonymity: A model for protecting privacy," Int. J. Uncertainity Fuzziness Knowl.-Base Syst., vol. 10, no. 5, pp. 557–570,2002.

[3] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam, "l-diversity: Privacy beyond k-anonymity,"in Proc. 22nd Int. Conf. Data Eng., 2006, p. 24.

[4] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in Proc. 20th Int. Conf. Very Large Data Bases, 1994, pp. 487–499.

[5] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2000, pp. 1–12.

[6] C. Zeng, J. F. Naughton, and J.-Y. Cai, "On differentially private frequent itemset mining," Proc. VLDB Endowment, vol. 6, no. 1, pp. 25–36, 2012.

[7] J. Vaidya and C. Clifton, "Privacy preserving association rule mining in vertically partitioned data," in Proc. 8th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2002, pp. 639–644.

[8] M. Kantarcioglu and C. Clifton, "Privacy-preserving distributed mining of association rules on horizontally partitioned data," IEEE Trans. Knowl. Data Eng., vol. 16, no. 9, pp. 1026–1037, Sep. 2004.

[9] W. K. Wong, D. W. Cheung, E. Hung, B. Kao, and N. Mamoulis, "Security in outsourcing of association rule mining," in Proc. 33rd Int. Conf. Very Large Data Bases, 2007, pp. 111–122.

[10] W. K. Wong, D. W. Cheung, E. Hung, B. Kao, and N. Mamoulis, "An audit environment for outsourcing of frequent itemset mining," Proc. VLDB Endowment, vol. 2, no. 1, pp. 1162–1173, 2009.

[11] A. Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke, "Privacy preserving mining of association rules," in Proc. 8th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2002, pp. 217–228.

[12] Maurizio Atzori, F. Bonchi, F. Giannotti, and D. Pedreschi, "Anonymity preserving pattern discovery," VLDB J., vol. 17, no. 4, pp. 703–727, 2008.

[13] R. Bhaskar, S. Laxman, A. Smith, and A. Thakurta, "Discovering frequent patterns in sensitive data," in Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2010, pp. 503–512.

[14] N. Li, W. Qardaji, D. Su, and J. Cao, "Privbasis: Frequent itemset mining with differential privacy," Proc. VLDB Endowment, vol. 5, no. 11, pp. 1340–1351, 2012.

[15] F. McSherry and K. Talwar, "Mechanism design via differential privacy," in Proc. 48th Annu. IEEE Symp. Found. Comput. Sci., 2007, pp. 94–103.

[16] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in Proc. 3rd Conf. Theory Cryptography, 2006, pp. 265–284.

[17] R. Chen, N. Mohammed, B. C. M. Fung, B. C. Desai, and L. Xiong, "Publishing set-valued data via differential privacy," in Proc. Int. Conf. Very Large Data Bases, 2011, pp. 1087–1098.