A Survey Paper on Frequent Itemset Mining Methods and Techniques

Sheetal Labade[1], Srinivas Narasim Kini[2]

[1]M.E (Computer), Department of Computer Engineering, Jayawantrao Sawant College of Engineering, Hadapsar Pune-411028, India Affiliated to Savitribai Phule Pune University, Pune, Maharashtra, India -411007

[2]Professor, Department of Computer Engineering, Jayawantrao Sawant College of Engineering, Hadapsar Pune-411028, India Affiliated to Savitribai Phule Pune University, Pune, Maharashtra, India -411007

Abstract: ***Recently, there has been growing interest in designing differentially private data mining algorithms. Especially, so in frequent itemset mining. Frequent itemset mining is subset of frequent pattern mining. Additionally, our paper discusses other subsets of frequent pattern mining namely, frequent sequential mining and frequent structured mining. Individual privacy may get affected by exposing frequent itemsets.Therefore, differential privacy is now at central alarm. This paper examines literature analysis on several methods for mining frequent itemsets. Further, it also describes in what manner differential privacy is handled in present systems. The paper concludes by highlighting the importance of protecting individual's data and combining different techniques and methods to achieve differentially private frequent set mining in its truest sense.***

Keywords: Data mining**,** frequent itemset mining, differential privacy, private, frequent pattern mining.

1. Introduction

Data mining is the sighting of secret information found in databases and therefore it can be seen as a very important stage in the knowledge discovery procedure. It includes different functions like clustering, classification, prediction and link analysis i.e. nothing but associations.Mining of association rule is most significant data mining application.Agarwal firstly announced association rules which are supportive for analyzing customer behavior in retail trade, banking system etc. Association rule can be well-defined as {A, B} => {C}. In retail stores if customer buys A, B he is about to by C. Idea of association rule today used in different areas such as intrusion detection system, biometrics, production planning etc. Association rule mining is the procedure of finding out association rules which fulfil the predefined minimum support and confidence from a known data base. The item set which supports the minimum support and confidence is known as frequent itemset.The problem of discovering the association rules can be allocated into two fragments: First, Invention of all frequent item sets. Second, Generation of strong association rules from the frequent item sets.

Frequent pattern mining is the method of mining data in a set of items or some patterns from big databases,which must chains the minimum support threshold. A frequent pattern is a pattern that befalls recurrently in a dataset. These frequent patterns may present in different forms like frequent itemsets,sequential pattern or substructure.Frequent itemset generally indicates that a set of items that frequently occurs together in a transactional data set.For example,milk and sugar.The patterns that customer have a tendency to purchase in subsequences refers to frequent sequential pattern. For example,customer generally first purchase the laptop and then think for purchasing anti-virus software for it.A substructure can take many structural forms like graph,trees or lattices and these forms can be combined with itemsets or subsequences.And if these substructure shows some frequency in operations, then it is known as a frequent structure pattern.

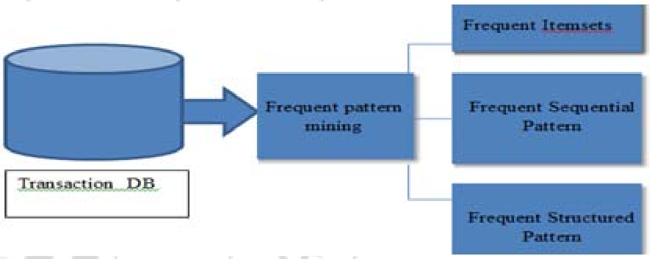

Figure 1: Frequent pattern mining framework

Frequent itemset mining plays vital part in numerous data mining arenas as associationrules, warehousing, correlations, clustering of high-dimensional biological data, and classification.Problems occurring during market basket analysis is one main motives of frequent itemset mining.A set of items purchased by customer in market basket analysis in any transaction is known as tuple. An association rule extracted from market basket database based on the principle that if some items are bought in transaction, then it is possible that some other items are also get purchased.A very important thing while mining the the association rule is the itemsets mining.Therefore various methods are present to generate frequent itemsets,with the help of which we can efficiently mine the association rules. Large number of algorithms are present to mine the frequent itemsets.The absolute vital algorithms are briefly described here.According to the method of generation of candidate itemsets and support count, algorithm may be different.

The methods of mining frequent itemsets are classified into basic three practices:

1. Horizontal data format
2. Vertical data format
3. Projected database techniques.

1.1 Horizontal Data Format

There are some ways which mines frequent itemsets from the set of transactions in Transaction Identifier-itemset arrangement where, TID is nothing but transaction ID and itemset is the set of items accepted in transaction TID. This is recognized as the horizontal data format.

For example, T1→ X, Y, Z,here T1 is the TID and X, Y, Z are the itemsets in that transaction.

Now we see different algorithms which work in Horizontal data format:

1.1.1 Apriori Algorithm

Apriori algorithm (Agrawal et al. 199), is the utmost central and significant algorithm for mining frequent things.All the frequent things in a given database DB are searched with the help of Ariori algorithm.The keynote of Apriori algorithm is to create several passes above the database. It pays an repetitive procedure called as a breadth-first search which is also known as levelwise search over the search area, where k-things are recycled to discover (k+1)-things.

WORKING OF APRIORI
A] Discovery of all frequent itemsets.
Catch frequent things:
Things whose existence in database is more than or equivalent to the smallest help threshold.
Catch frequent itemsets :
i.Generate candidates from frequent things.
ii.Prune the outcomes to identify the frequent itemsets.
B] Develop robust association rules from frequent itemsets.

It will consists rule which fulfills the minimum support and minimum confidence threshold.

1.1.2 Hash-based system

It is based on the principle that, ak-itemset whose parallel hashing bucket count is under the threshold can"t be frequent.The algorithm comes under this system is known as Direct Hashing and Pruning Algorithm. DHP basically based on apriori.for this purpose CDHP adds some additional control like hash table that targets at controlling the generation of candidates in set as much as probable.DHP also gradually decreases the database by removing the attributes in transaction or even by removing whole transactions when they look as if consequently useless. In this system, support is calculated by plotting the items from the candidate list into the buckets which is separated according to support known as Hash table structure. As the new itemset is come across if item occur earlier then rise the bucket count else add into new bucket. Thus in the termination of this whole procedure the support of bucket is compared and the bucket whose support count is less the minimum support is removed from the candidate set.

1.1.3 Transaction reduction method

Here we reduce the transaction based on the principle that, the transactionwhich does not contain any frequent x-itemsets cannot contain any frequent (x+1)- itemsets. So such a transaction can be marked or removed from additional consideration because succeeding database scans for y-itemsets, where y >x, will not want to reflect such a transaction.

1.1.4 Partitioning the data

Partitioning approach works on the standard that, any itemset that is possibly frequent in database must be present in at least one of the partitions of database. It overrides the memory problem for bulky database which do not suitable for main memory because small parts of database simply suits into main memory. Here, database contains many transactions. This approach works in two phases.

Phase I:
a) Divide whole database into number of partitions.
b) Search the frequent itemsets resident to every partition (1st scan).
c) Group entire resident frequent itemsets to form candidate itemset.

Phase II:
Search global frequent itemsets between candidates (2nd scan).
And at the end of 2nd phase we get frequent itemset in the database.

1.1.5 Sampling Algorithm

The elementary notion of the sampling method is to pick arbitrary sample S of the assumed data D, and then find out frequent itemsets in S as a substitute of D. Like this, we trade off some amount ofexactness in contradiction of efficiency. The S sample size is such that which can be used for searching the frequent itemsets in S can be done in main memory, and so only single scan of the transactions in S is required overall. Because we are interested in findiing frequent itemsets in S rather than in D, it is probable that we will miss some of the global frequent itemsets. To reduce this probability, we use a lower support threshold than minimum support to search the frequent itemsets native to sample. Else; another pass can be performed to search the frequent itemsets that were unused in the first pass. The sampling tactic is particularly advantageous when efficiency is of greatest importance such as in computationally exhaustive applications that must be route repeatedly.

1.1.6 DIC technique

Dynamic itemset counting (DIC) technique was planned in which the database is divided into chunks marked by start points. In this distinction, novel candidate itemsets can be added further at any start point. The technique practices the count-so-far as the minor bound of the real count. If the count-so-far passes the minimum support, the itemset is added further into the frequent itemset assembly and can be recycled to produce the lengthier candidates. and therefore; this takes less database scans than Apriori for searching every frequent itemsets.

1.2 Vertical data format

Data can be offered in another way also like item-TID_set, where item is an item name, and group of TIDs indicates the set of transactions iin which item is present. This is known as the vertical data format.

For example, X→ T1, T3, hereX is the item and T1, T3 are the transactions for that item.

1.2.1 Eclat

Eclat algorithm works on the basis of the Vertical data format. First, the item sets are checked in lexicographic order i.e nothing but depth-first traversal of the prefix tree.The search plan is the same as the general pattern for with canonical forms having the prefix assets and holding a picture-perfect extension rule (just produce canonical extensions). Eclat produces many candidate item sets than Apriori, because it does not accumulate the support of every visited item sets.As an importance it cannot completely feat the Apriori property for pruning.Eclat practices a only vertical transaction demonstration. For this algorithm no subset checks and no subset generation are desired to calculate the support. The support of item sets is fairly determined by intersecting transaction lists.

1.3 Projected database techniques

The idea of projected database was suggested and applied to mine the frequentitemsets efficiently because initial styles are capable to mine the frequent itemsets but use great amount of memory. This type of database uses divide and conquer scheme to mine itemsets therefore it totals the support more efficiently then Apriori based algorithms. Tree projected design based methodologies use tree structure to store and mines the itemsets. The projected based design contains the record idwhich is separated by column then record.Tree Projection algorithms created upon two types of organization breadth-first and depth-first.For breath-first order, nodes are assembled level by level in the lexicographic tree for frequent itemsets . In order to calculate frequencies of nodes at k level, tree projection algorithm upheld matrices at nodes of the k-2 level and one database scan was essential for calculating support. Every single transaction is projected by node serially.The projected set of transaction for reduced set is used to assess frequency.

Following algorithms comes under projected database technique:

1.3.1 FP-Growth Algorithm

The most widespread algorithm for frequent itemset mining is nothing but the FP-Growth Algorithm which aims at removing the bottlenecks of the Apriori-Algorithm in producing and testing candidate set. The problem of Apriori algorithm was share out with, by make known to a unique, compact data structure, called frequent pattern tree or FP-tree therefore grounded on this structure an FP-tree-centred pattern fragment growth manner was developed. FP-growth uses a mixture of the vertical and horizontal database arrangement to store the database in main memory. As an alternative of storing the cover for every item in the database, it stores the genuine transactions from the database in a tree structure and each item has a linked list passing through all transactions that contain that item. This novel data structure is meant by FP-tree. Basically, all transactions are stored in tree-like data structure.

The FP-tree is built in the following stages:

1) First the scanning of the transaction database DB is done once. Then, assemble the set of frequent items F and their supports.F is sorted in support downward order as L, the list of frequent items.
2) Construct the root of an FP-tree, T, and tag it as “root”.

1.3.2 Broglet’s FP-Growth

An efficient FP-Growth algorithm using C Language was implemented by Broglet.The steps of preprocessing using Broglet algorithm is as follows:

a) In first scan the frequencies of the items are determined.
b) In next step all rare items, that is, all items that seem in fewer transactions than a user-specifiedminimum number are removed from the transactions, since, noticeably, they cannot be part of the frequent item data set.
c) Finally, the items in each and every transactionis arranged, so that they are in downward fashion with respect to their frequency in the database.

1.3.3 CT-PRO Algorithm

CT-PRO is based on compress tree structure and also it is the variation of classic FP-tree algorithm. It navigates the tree in bottom up style. It is based upon the non-recursive based procedure. Compress tree structure is likewise the prefix tree in which every item is put away in the descending direction of the frequency with the Field_Index, Frequency_F, Pointer, Item_ID.In this every item if the databases afterward searching the frequency of items and items whose frequency is superior than minimum support are plotted into the index forms according to the existence of items in the transaction. Root of the tree is at all times at index „0‟ with extreme frequency components.The compact data structure which is used by CT-PRO is known as CFP-tree i e. compact frequent pattern tree so that each of the item of the transactions can be signified in the main memory.

1.3.4 H-mine Algorithm

H-mine algorithm is the enhancement above FP-tree algorithm as in H-mine projected database is formed using in-memory pointers. The hyperlinked structure is used by H-mine algorithm for mining the items. It is used upon the dynamic tuning of pointers which benefits to preserve the processed projected tree in main memory therefore H-mine offered for frequent pattern data mining for data sets that can suitable for main memory. It has polynomial space complexity hence it is extra space efficient than FP-growth and also aimed for fast mining purpose.If the database is compact then it take part in FP-Growth dynamically by spotting the exchanging the condition and creating the FP tree.This functioning confirms that it is scalable for both huge and average size databases and for both light and compact datasets.The improvement of expending in-memory pointers is that their projected database does not want any memory, the memory is compulsory simply for the set of in-memory pointers.

1.4 Privacy Preservation with Differentially Private Mechanisms

Although the mining of frequent itemsets has been well studied in innumerable techniques, the privacy alarms expanding in the community are posturing different challenges on this problem. The quick growth of digital information from diverse sources for e.g. web, institutions, individuals etc has been generating great chances for revealing own information. While the worries of revealing sensitive and private evidences have directed researchers to improve a multiple privacy representations, therefore the project of effective and efficient data mining algorithms with privacy is really great task. Privacy alarms present where individually recognizable information is gathered and kept in digital form, and data mining programs are capable of accessing such data, even through data preparation. Inadequate or absent exposure control can be the main reason of privacy concerns. To handle such worries, plentiful data security-improving skills have been developed. And also, there has been abundant amount of recent efforts are taken over developingprivacy preserving data mining systems. In this segment, we gaze at some of the progresses indefending privacy and data security in data mining.

1.4.1 Data security enhancing procedures

Many such procedures have been developed to service for defending the data. Databases can hire a multilevel security model to categorize and limit data rendering to several security levels, with users allowed access to only their approved level. It has been observed that, however, users are performing precise queries at their approved security level can still infer great amount of sensitive facts, and that a related possibility can arise over data mining. Encryption is alternative procedure in which individual data items might be encrypted. This may encompass blind signatures which uses public key encryption, biometric encryption where the copy of a person‟s fingerprint, iris, retina is used to encrypt his or her personal information, and anonymous databases which license the links various databases but edge access to personal information simply to those who want to see; personal information is converted and stored at altered sites. Intrusion detection is added energetic region to research which helps to defend the privacy of individual data.

1.4.2 Privacy-preserving data mining

It is a part of data mining research which studies, protection of privacy in data mining. Privacy-enhanced or Privacy sensitive are also another two terminologies of privacy preserving data mining. It works for gaining valid data mining outcomes without revealingthe essential sensitive data values. Most privacy-preserving data mining approaches usesome form of alteration on the data to achieve privacy preservation.Normally,such approaches decreases the granularity of illustration to preserve privacy. For example, they may generalize the data from distinct customers to customer crowds. This drop in granularity results in the damage of information and probably the utility of the datamining results. This is the usual trade-off between loss & privacy of useful information.

Following are the categories of Privacy-preserving data mining techniques:

(a)Randomization techniques: These approaches add noise to the data to cover some feature of records. The noise added should be satisfactorily huge so that individual record features, specifically sensitive ones, cannot be extracted. However, it should be added using some diplomacy so that the concluding outcomes of data mining are fundamentally preserved. Procedures are considered to develop combined distributions from the worrieddata. Afterward, data mining skills can be settled to work with these combined distributions.

(b)The k-anonymity and l-diversity approach: These ways modify individual records so that they cannot be individually recognized. In the k-anonymity scheme, the granularity of data illustration is reduced appropriately so that every specified record plots on minimum k another records in the data set. It practices skills like generality and suppression. The k-anonymity way is fragile in that, if there is equality in sensitive values in any set, then those values may be undirected for the different records. The l-diversity Procedure was studied to overcome this weakness by using intragroup diversity of sensitive values to guarantee anonymization. The basic objective is to create it satisfactorily tough for adversaries to use groupings of record features to exactly identify individual records.

(c)Distributed privacy preservation: We know that big data sets could be split and spread either horizontally or vertically, or may in the fusion of both. While the distinct locations may not want to share their whole data sets, they may give permission to restricted information sharing with the usage of a variability of protocols. The total effect of such ways is that to preserve privacy for every individual entity, while arising combined outcomes over whole data set.

(d)Dropping the usefulness of data mining results: In many circumstances, even though the data may not be present, the yield of data mining may results in the destructions of privacy. The answer could be to dropping the usefulness of data mining by either altering data or mining outcomes, such as thumping some association rules or slightly changing certain sorting representations.Researchers just suggested fresh concepts in privacy-preserving data mining such as the perception of differential privacy. The overall clue is that, for any two data sets that are close to one another i.e., that vary only on a little data set such as a single element, a specified differentially private algorithm will act almost the identical on both data sets. This definition springs a solid assurance that the appearance or non-appearance of a small data set (e.g., representing a singular) will not touch the concluding productivity of the query pointedly. Based on this opinion, a set of differential privacy-preserving data mining algorithms have been established. Research on this track is going on. We imagine more dominant privacy-preserving data issuing and data mining algorithms in the nearby future.

2. Literature Survey and Related Work

As the prior section tells several approaches for mining the frequent itemsets and also give focus on the various privacy issues and the categories of privacy preserving, but still there is a vast gap to reach at the perfection. So, as a step in the direction of this, this paper tried to clutch several conceptions so that a new and efficient technique can be suggested. The thorough studies are as follows.

In [1], author gives general dead end result showing that a validation of Dalenius‟s objective along the lines of semantic security cannot be reached. Conflicting to insight, a different result threatens the privacy though somebody not present in the database. This state of undertakings suggests a fresh degree which is known as differential privacy, which, spontaneously, captures the enlarged risk to one‟s privacy acquired by contributing in a database. The method developed in order of papers, concluding in those described in, can reach any desired level of privacy below this measure. In several belongings, extremely perfect information about the database can be provided while concurrently warranting very great degree of privacy.The work debated herein was initially motivated by the problem that how to expose valuable information about the core population, as characterized by the database, while taking care the individual‟s privacy.

Data holders functioning independently and with partial knowledge are left withthe trouble of freeing information that does not negotiate privacy, confidentiality or country - wide benefits. L. Sweeney [2] this paper contains official protection model entitled k-anonymity and a setof associated plans for distribution. An issue provides k-anonymity safety ifthe information for every individual enclosed in the release cannot be notable from atleast k-1 individuals whose information also looks in the publication. These paper alsostudies re-identification occurrences that can be grasped on publications that obey to k anonymityunless additional plans are appreciated. The k-anonymity protectionmodel is vital because it forms the origin on which the real-world schemes known asData fly, m-Argus and k-Similar deliver agreements of privacy protection.

A. Machanavajjhala, Gehrke, Kifer & Venkitasubramaniam [3] shows that k anonymized dataset cannot avoid major attacks because diversity is not present in the sensitive features.Here, they announced a framework known as l-diversity,which gives robust privacy agreements. They also verified that l-diversity andk-anonymity have adequate resemblance in their arrangement that k-anonymity algorithms can be altered to work with l-diversity.

Advancement in bar-code knowledge has made it probable to retail groups to gather and store vast quantities of sales data, denoted as the basket data. A record in such data normally involves the date and the items purchased in the transaction. Big groups sight such databases as central part of the selling organization. Agrawal and Srikant in [4] offered 2 fresh algorithms, first, Apriori and second, AprioriTid, for totally searching important association rules in the items which is present in the big transaction databases. They also equated these algorithms to the formerly identified algorithms, the AIS and SETM algorithms. They also open how the finest features of the both offered algorithms can be mixed-up into a new hybrid algorithm, called AprioriHybrid, which afterward becomes the algorithm of excellent quality for this difficulty.

Frequent pattern mining perform an important activity in mining associations in different areas of sales transactions. The Apriori method gains better performance by decreasing the size of candidate sets generated. An Apriori algorithm can suffer from following two issues:

1) It is expensive to handle a vast no. of candidate data sets.
2) It is really a tedious activity to frequently scan the database and checked a big set of candidates by matching the patterns.

To overcome these problems of Apriori authors J. Han, Pei, & Yin [5] develop and assimilate the following fresh three approaches:.

a) A new, condensed data structure, called frequent-pattern tree is created, which is an prolonged prefix-tree structure for storing critical, measurable information of patterns which occurs frequently.
b) Then, an FP-tree-grounded pattern-fragment growth mining technique isderived, which begins with length-1 frequent pattern, studies their conditional-pattern base, and executes mining in recursive manner on that tree.
c) Finally, the partitioning based, divide and conquer method is offered for search mining procedure. All these methods pay to significant reduction in the search costs.

Associated with the growing capability to assemble personal data, privacy has turn out to be a major alarm. Therefore,Zeng, et al.[6]focus on privacy concerns that raises due to task of finding frequent itemset in databasetransaction.Here they offered a new algorithm which is based in differential privacy .And also, exactly quantified the privacy - utility trade-off by decreasing the length of transaction while mining frequentitemset .

The key contributions is given by Vaidya & Clifton [7] for the privacy preservation in vertically partitioned databases. They consider secure computation of scalar product to search overall frequent itemsets. They also offered that it is probable to achieve better distinct security with communication cost equivalent to that essential to construct a central warehouse of data.

Large collections of data is used by data mining to find out useful information from it.But many times it is observed that these databases are divided across different sites.These sites can not directly expose their own data due some privacy concerns. Kantarcioglu and Clifton [8] suggested new approach which overcomes the privacy issue in horizontally divided data. This approach includes cryptographic procedure for reducing the sharing of information but which adds some overhead of mining task.

The owner of data has to consider different things while outsourcing the data for example, basic computational resources and cost of transporting transactions between several sources. If the service supplier is nottrustworthy, he should be disallowed from using the original data because that data may be sometimes very private and important.To avoid all these problems in outsourcing, Wong et al. [9] suggested the set of encryption techniques for database transaction. Initially one-to-one substitution cipher was used which is vulnerable to attacks, and then one-to-n item plotting method also developed.

Wong et al. [10] then studied the integrity issue in outsourcing the mission of mining frequent itemset .In this work they established a new outline of the problem with the set of malicious activities that a malicious service provider might use. One synthetic itemset implanting procedure for formulating an audit environment was offered.

The sighting of frequent itemsets can attend valued commercial and research determinations. So how to release such valuable frequent itemset information was really a big problem. N. Li et al. [14] offered a fresh technique known as PrivBasis, which influences a new idea called basis sets. Here new algorithm was proposed for secretly building a basis set and then finding frequent itemsets which are most useful.

Set-valued information, like transaction data, web search information means that the data in which everyrecord owner is linked with a set of items pinched from a group of items. Handling such set-valued information gives many chances for different data mining activities inseveral application areas.But, such data may have some sensitive information within it whose privacy should be preserved. In [17], they handled problem of releasing set-valued information thatinstantaneously guards privacy while considering differentially and also offers certain utility to the miners.

Frequent sequential pattern mining is a principal assignment in several areas like biology and economics. In[19] Bonomi and Xiong suggested new technique to excellently mine the top-k substring without trailing information of the frequent prefixes. They first established two prefix tree mining procedures which make use of statistical infoto wisely adjust the noise. And they Second worked on local transformation method on the basic string records which minimizes the noise added by the Laplace methodology. And, finally these two techniques are combined to form the new algorithm called two phase mining algorithm.

Pattern mining which is based on frequent graph is one of the keythemes in data mining research. It has been gradually applied for several application areas like Bioinformatics, cheminformatics and social n/w analysis. Graph patterns may contain sensitive data of mobile phone-call graph and so privacy issue again arises. Differential privacy has now became important standard for private data enquiry due to its verifiable privacy assurance. In [20] Shen and Yustudied the differentially private algorithm for frequent graph mining patterns, which was the first privacy algorithm on graph patterns.

Algorithms which is based on association rule mining are very much important in e-commerce area.Zheng et al.[27] used different association rule mining algorithm on real-world and one synthetic dataset of IBM. Finally, they observed that whether the real world and synthetic datasets have some closeness in their results or not.

3. Conclusion

The basic objective of data mining procedure is to mine information from databases and convert it into reasonable arrangement for further usage. Association rules indicate an effective practice for frequent pattern matching from many decades. This paper provides a brief review on several methodologies for mining frequent itemsets. In conclusion we see there is lot of interest in protecting individual‟s data and ensuring that it is not revealed in the course of data mining and also achieving this by means of inducing spurious data and further doing transaction splitting. Hence, there is a need to combine different techniques and methods mentioned in this paper appropriately and wherever relevant to achieve differentially private frequent itemset mining in its truest sense.

References

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

[2] L. Sweeney, “k-anonymity: A model for protecting privacy,” Int. J.Uncertainity Fuzziness Knows.-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.

[18] X. Zhang, X. Meng, and R. Chen, “Differentially private setvalued data release against incremental updates,” in Proc. 18th Int. Conf. Database Syst. Adv. Appl., 2013, pp. 392–406.

[19] L. Bonomi and L. Xiong, “A two-phase algorithm for mining sequential patterns with differential privacy,” in Proc. 22nd ACM Conf. Inf. Knowl. Manage, 2013, pp. 269–278.

[20] E. Shen and T. Yu, “Mining frequent graph patterns with differential privacy,” in Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2013, pp. 545–553.

[21] R. Chen, B. C. M. Fung, and B. C. Desai, “Differentially private transit data publication: A case study on the montreal transportation system,” in Proc. 18th ACM SIGKDD Int.Conf. Knowl. Discovery Data Mining, 2012, pp. 213–221.

[22] R. Chen, G. Acs, and C. Castelluccia, “Differentially private sequential data publication via variable-length n grams,” in Proc.ACM Conf. Comput. Commun. Security, 2012, pp. 638–649.

[23] A. Ghosh, T. Roughgarden, and M. Sundararajan, “Universally utility-maximizing privacy mechanisms,” SIAM J. Comput.,vol. 41, no. 6, pp. 1673–1693, 2012.

[24] L. Parsons, E. Haque, and H. Liu, “Subspace clustering for high dimensional data: A review,” SIGKDD Explorations, vol. 6, no. 1, pp. 90–105, 2004.

[25] Pramod S, O.P. Vyas “Survey on Frequent Item set Mining Algorithms “,International Journal of Computer Applications (0975 - 8887) Volume 1 – No. 15.

[26] C. Borgelt. An Implementation of the FP- growth Algorithm. Proc. Workshop Open Software for Data Mining

[27] (OSDM‟05 at KDD‟05, Chicago,IL),1–5.ACMPress, New York, NY, USA 2005

[28] Z. Zheng, R. Kohavi, and L. Mason, “Real world performance of association rule algorithms,” in Proc. 7th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2001, pp. 401–406.

[29] Data mining concepts and techniques, Jawai Han, Michelline Kamber, Jiran Pie, Morgan Kaufmann Publishers, 3rd Edition.

[30] Pramod S, O.P. Vyas “Survey on Frequent Item set Mining Algorithms “,International Journal of Computer Applications (0975 - 8887) Volume 1 – No. 15.

[31] Jian Pei , Jiawei Han , Hongjun Lu , Shojiro Nishio , Shiwei Tang , Dongqing Yang “H-Mine: Hyper-StructureMining of Frequent Patterns in Large Databases”.

Author Profile

Sheetal Labade, is currently pursuing M.E (Computer) from Department of Computer Engineering, Jayawantrao Sawant College of Engineering, Hadapsar, Pune-411028, India Affiliated to Savitribai Phule Pune University, Pune, Maharashtra state, India -411007. She received her B.E (Computer) Degree from GSMCOE, Pune, India Affiliated to Savitribai Phule Pune University, Pune, Maharashtra state, India -411007. Her area of interest is data mining, information retrieval and information security.

Prof. Dr. Srinivas Narasim Kini, received his PhD Degree from Cochin University of Science and Technology, Thrikkakara, South Kalamasserry, Cochin. He received his M.E (Computer Science and Engineering) Degree from B.M.S. College of Engineering, Basavanagudi, Bengaluru, India. He received his B.E (Computer Science and Engineering) Degree from K L E Society's College of Engineering Udyambaug Belgaum, India. He is presently working as Professor at Department of Computer Engineering, Jayawantrao Sawant College of Engineering, Hadapsar, Pune-411028, India Affiliated to Savitribai Phule Pune University, Pune, Maharashtra state, India -411007. His area of interest is data mining, web mining, information retrieval, distributed computing and N/W security.