Data Mining Ethics in Privacy Preservation - A Survey

S. M. Mahajan and A. K. Reshamwala

Abstract—**When most people think of ethics or morality, it refers to well-defined standards of right and wrong, usually in terms of rights, obligations, and benefits to society, fairness, or specific virtues. As technology advances, computers continue to have a greater impact on society. The world-wide-web in its current form, linking heterogeneous data from distributed sources, has led to a number of concerns about issues related to privacy, copyright, and intellectual property. The benefits of such a Web are plenty but threats to personal information in such as social networking sites also abound. Privacy is essential for the proper functioning of a liberal, democratic society.**

Another area where computer ethics play an important part is data mining, where the prolific data generated from various sources, used for identifying patterns. With increasing usage of data mining in the public and private sectors, privacy assumes paramount importance. Sequential pattern mining is commonly defined as finding the complete set of frequent subsequences in a set of sequences, is currently one of the most active areas of research. We provide an overview of privacy preserving association rule as well as in sequential mining, which is one of the rapidly emerging research areas of privacy preservation.

Index Terms—**Association rule mining, computer ethics, data mining, sequential pattern mining, privacy preservation.**

I. Introduction

Data mining technology has emerged as a means for identifying patterns and trends from such large quantities of data. For instance, shopping centers conclude that male customers who buy diaper usually shop beers by analyzing consuming lists. This forms the relation between diaper and beer through rearranging these goods. In addition, there is also the relation between milk and bread. This improvement of goods arrangement after analysis not only makes customers convenient but increases expenditure. Credit card centers of banks find out behavior features and consuming models of high-quality clients from lots of trade data so as to seek potential clients, stimulate clients' consumption and create more opportunities of overlapping sell.

However, data mining also brings some problems. For example, credit card centers may intentionally or unconsciously make sensitive information of clients leak while mining relating information of clients. With the Internet popularity, because more and more information can be obtained in electronic form, that people have their own privacy confidential is becoming increasingly urgent. According to statistics, even if privacy protection measures, about one-fifth of Internet users don't like to provide their own information to the Web site and more than the half investigators only in good privacy-preserving measures are willing to provide their own information to the Web site. Among the potential consumers shopping in internet browser, there are almost half who gave up the hope for internet shopping because of worrying about no protection of their privacy. Therefore, how to ensure personal privacy in data mining has become a need to be addressed. It requires significant research on how to extract valuable knowledge in data and at the same time, prevent private or sensitive information in data mining process from leaking. Thus techniques of data mining without leaking the private information are needed. Research on privacy preserving data mining is developed for this purpose. Correspondingly the privacy preserving data mining and knowledge discovery should be developed aimed at these problems.

The association rule mining problem was first proposed by Agrawal et al [1]. In order to make a publicly available system secure, we must ensure not only that private sensitive data have been trimmed out, but also to make sure that certain inference channels have been blocked as well. Under privacy constraints, the association rule mining problem was extensive researched. Many effective methods for privacy preserving association rule mining have been proposed [2-13]. But most of those methods may result in information loss and side-effects in some extent, such as non-sensitive rules falsely hidden and spurious rules falsely generated, may be produced in the sensitive rule hiding process. That is, an essential problem under the context is trade-off between the data utility and the disclosure risk.

Sequential pattern mining is commonly defined as finding the complete set of frequent subsequences in a set of sequences [14]. Sequential pattern mining provides a means for discovering meaningful sequential patterns among a large quantity of data. For example, let us consider the sales database of a bookstore. The discovered sequential pattern could be "70% of people who bought Harry Porter also bought the Lord of Rings at a later time". The bookstore can use this information for shelf placement, promotions, etc.

The paper is organized as follows, Section II gives the classification frame of privacy-preserving algorithm in data mining, we provide an overview of privacy preserving association rule mining, which is one of the most popular pattern discovery methods in the new and rapidly emerging research area of privacy preserving data mining in section III and in section IV, we summarize various proposals and algorithms designed in the research area of privacy preserving sequential pattern mining.

Manuscript received January 5, 2011; revised July 5, 2011.

S. M. Mahajan, principal, Institute of Computer Science, M.E.T, Banda, Mumbai, India (e-mail: sunitam_ics@met.edu).

A. K. Reshamwala, Assistant Professor, Research Scholar, Computer Department, Mukesh Patel School of Technology Management and Engineering, SVKM's NMIMS University, Mumbai, India (e-mail: alpa.reshamwala@nmims.edu).

II. Data Mining

The data mining target is to find knowledge, and knowledge is presented through certain patterns. Association rule is the most frequently used method in data mining, which finds out the association between data and different objects by discovering the potential dependence among data. Classification and clustering are to sort out things by characterizing the common significance of different things. According to Y. Shen et al. [15], the privacy-preserving classification mainly has the following two ways: (1) privacy preserving technology of centralized data and distributed data depending on the data distribution. The latter one can be further classified into privacy preserving technology of horizontal partition and vertical partition. Distributed privacy preserving data mining algorithm is realized through the employment of secure multi-party computation (SMC) as discussed by Y. Shen et al. in [16]. (2) According to the data mining algorithm classification, privacy-preserving technology can be classified.

The two basic methods by Verykios et al. in 17], is associated with association rule privacy-preserving technology, also has been put forward by Y. Shen et al. in [15]. The first one is to prevent from producing association rules by hiding frequent item sets; the second one is to avoid producing important rules by making the belief degree of important rules achieve the lowest belief appointed by users. Oleveira and Zaiane in [18], even proposed one heuristic privacy-preserving method which realizes the protection of sensitive rules through one kind of single scanning algorithm. This algorithm mainly takes the method of removing part of information to realize data clearing, and then to hide sensitive rules, which won't append any noise to raw database. The disadvantage of rule mining technology in centralized database, mainly have the several following points: network traffic is considered a little, mining efficiency is low and the degree of spatial complexity is high. Data perturbation way is very efficient used in data mining alone in centralized environment, but it will produce some problems in distributed environment, Vaidya and Clifton in [13], has proposed association rule privacy preserving algorithm which is based on data vertical distributing, which gains support counting of item sets by securely computing scalar product delegating sub-item sets.

The most typical classification data mining are classification methods based on distance, classification methods based on decision tree, Bayesian classification and so on. Agrawal and Srikant in [10], proposed one algorithm of preserving data privacy which first adds noise to raw data, this randomized management will not influence data distribution, and then based on reconstruction technology concludes distributed information similar with raw data and meanwhile constructs decision tree. Du and Zhan [19] proposed privacy-preserving K-nearest classification algorithm relying on vertical distribution. While decision tree classification method under the condition of data distribution is to construct decision tree by transferring middle computation result. Lind ell and Pink as [20] first proposed privacy-preserving ID3 classification tree distributed algorithm which adopts computing tool of security involving the participation of semi-honest third party. In addition, based on Bayesian classification algorithm, M. Kantarcioglu and Clifton [12] established a Naive Bayesian Classification model of horizontal partition to realize privacy preservation through the secure sum method.

Privacy-preserving clustering mining relying on data perturbation is to make real sensitive data unknown by transforming data, and then to process clustering analysis. However, privacy-preserving clustering based on SMC is to make one party who participates only under the condition of owning its own personal information become fully aware of the whole clustering information, mainly through constructing secure multiparty protocol. Oliveiral and Zaiane [21] proposed rotational based transformation (RBT) method to transform data, which realizes the isometry transformation of points in multi-dimension and achieves an excellent privacy preserving result. Merman and Ghost [22] proposed a privacy-preserving method solving distributed clustering analysis, in which the raw or disturbed data shared by every site will form suitable parameter in each local site, transmit parameter to the central site and accomplish high quality distributed clustering through suitable samples.

III. Association Rules

Let $I = \{ I_1, I_2 \ldots, I_m \}$ be a set of items [23]. Let D be a database of transactions where each transaction T is a set of items such that $T \subseteq I$. Each transaction is associated to an identifier, call TID. A transaction T is said to contain A if and only if $A \subseteq T$. An association rule is an implication of the form $A \Rightarrow B$, where $A \subseteq I$, $B \subseteq I$, and $A \cap B = \phi$. The rule $A \Rightarrow B$ holds in the transaction set D with support s as given in equation (1), where s is the percentage of transactions in D that contain $A \cup B$. The rule $A \Rightarrow B$ has confidence c in the transaction set D as given in equation (2).

$$\text{Sup}\ (A \Rightarrow B) = \text{P}\ (A \cup B) = \frac{|A \cup B|}{|D|} \quad (1)$$

$$\text{Conf}\ (A \Rightarrow B) = \text{P}\ (B \mid A) = \frac{|A \cup B|}{|A|} \quad (2)$$

where $|A|$ is named as the support count of the set of items A in the set of transactions D, as denoted by *sup_ count(A)* . A occurs in a transaction T, if and only if $A \subseteq T$. Rules that satisfy both a minimum support threshold (*min_ sup*) and a minimum confidence threshold (*min_conf*) are called strong. A set of items referred to as an item set. An item set that contains k items is a k -item set. Item sets that satisfy *min_ sup* is named as frequent item sets. All strong association rules result from frequent item sets.

According to privacy protection technologies discussed in [24], at present, privacy preserving association rule mining algorithms commonly can be divided into three categories [25]: (1) Heuristic-Based Techniques, (2) Reconstruction-Based Association Rule and (3) Cryptography-Based Techniques.

Heuristic-based techniques are to modify data for the selected data sets and take into account the effectiveness of data security and privacy. The methods of Heuristic-based modification include perturbation, which is accomplished by the alteration of an attribute value by a new value (i.e., changing a 1-value to a 0-value, or adding noise), and blocking, which is the replacement of an existing attribute

value with a "?". There is a basic principle of choosing the transaction or the item of item set to be modified that we should reduce the influence of the original database as far as possible. Dasseni et al. [4] extends the sanitization of sensitive large item sets to the sanitization of sensitive rules. Oliveira and Zaiane also in [5] aims at balancing between privacy and disclosure of information by trying to minimize the impact on sanitized transactions or else to minimize the accidentally hidden and ghost rules. Wang et al. propose a matrix based sanitization approach to hide the sensitive patterns in [26]. It is the first paper to involve the consideration of avoiding the Forward-Inference Attacks [27], which can also be avoided in the sanitized database generated by our sanitization process. Oliveira et al. propose a novel method to modify databases for hiding sensitive patterns in [8]. Multiplying the original database by a sanitization matrix yields a sanitized database with private content. The method can avoid the question of the Forward-Inference Attacks. Lin and Cheng in [28] describes a technique that uses a queue and a random number generator to generate the items so that each item has an approximately equal frequency of being added to transactions. In replacement-Based Techniques, Saygin et al. in [29], discusses specific examples with the use of an uncertain symbol used in association rule mining. Agrawal et al. improve on the distribution reconstruction technique presented in [30] by using the Expectation Maximization (EM) method and they propose novel metrics for the quantization and measurement of privacy preserving data mining algorithms. Xiao and Tao in [31] presents a new generalization framework on the concept of personalized anonymity in order to perform the minimum generalization for satisfying everybody's requirements, the core of personalized anonymity is the concept of personalized anonymity. Ming and Ye in [32] proposes a personalized anonymity model on the base of (α,k)-anonymization model in order to resolve the problem of privacy self management and proposes corresponding anonymity method by using local recoding and sensitive attribute generalization.

Agrawal et al. in [10] first proposed the method of distribution reconstruction on numeric data which is disturbed by Bayesian algorithm in 2000. Then, Dakshi and Charu in [33] improve the work over the Bayesian-based reconstruction procedure by using an Expectation Maximization (EM) algorithm for distribution reconstruction. The work presented in [34] by Agarwal et al., deals with binary and categorical data in the context of association rule mining. Shariq J. et al. in [35] present a scheme called MASK, which attempts to simultaneously provide a high degree of privacy to the user and retain a high degree of accuracy in the mining results. Its scheme is based on a simple probabilistic distortion of binary data, employing random numbers generated from a pre-defined distribution function. The work by Agarwal et al. in [34], is based on the "select-a-size" and "cut-and paste" random transform operation to hide the original data set method, and then convert the transformed data into project item sets support counting, in order to identify frequent item sets.

Many Cryptography-based approaches have been proposed in the context of privacy preserving data mining algorithms. Cryptography-based approaches like SMC are secure at the end of the computations. No party knows anything except its own input and the results. SMC method is a typical technique. Clifton et al. in [36], presents four secure multiparty computations based on the methods that can support privacy preserving data mining which include the secure sum, the secure set union, the secure size of set intersection, and the scalar product. Theory for performing linear regression on vertically partitioned data has also been developed. Snail et al. in [37, 38], describe two different perspectives. The paper in [37] relies on quadratic optimization to solve for coefficients. The paper in [38] uses a form of secure matrix multiplication to calculate off-diagonal blocks of the full-data covariance matrix. Another way for computing the support count utilizes the secure size of set intersection method described in [36]. If the transactions are vertically partitioned across the sites, this problem can be solved by generating and computing a set of independent linear equations [13]. Shang and Hammerlock in [39], develops a log-linear model approach for strictly vertically partitioned databases and a more general secure logistic regression for problems involving partially overlapping data bases with measurement error. Kantarcioglu and Clifton in [40] use a secure multi-party computation to model the horizontal partitioning of transactions across sites, and present algorithms that incorporate cryptographic techniques to minimize the shared information without incurring much overhead in the mining process. Cheung at al. in [41] proposes an efficient distributed algorithm FDM (Fast Distributed Mining of association rules) for mining association rules.

IV. Sequential Pattern Mining

In the sequential pattern mining, we are given a database of customer transactions. Each transaction consists of the following fields: customer-ID, transaction-time, and the items purchased in the transaction. No customer has more than one transaction with the same transaction-time. We do not consider quantities of items bought in a transaction: each item is a binary variable representing whether an item was bought or not. An item set is a non-empty set of items. A sequence is an ordered list of item sets according to time. The support for a sequence is defined as the fraction of total customers who support this sequence. The problem of mining sequential patterns is to find the sequences with maximal length (e.g., maximal sequence) among all sequences that have a certain user-specified minimum support. Each such maximal sequence represents a sequential pattern.

A pattern-set is a non-empty set of patterns. A sequence is an ordered list of pattern-set. Without loss of generality, we assume that the set of patterns is mapped to a set of contiguous integers. We denote a pattern-set **a** as $(a_1a_2\ldots a_n)$, where a_j is a pattern. We denote a sequence **S** by $< s_1s_2\ldots s_n>$, where s_j is a pattern-set. A sequence $< a_1a_2\ldots a_n >$ is contained in another sequence $< b_1b_2\ldots b_m >$ if there exist integers $i_1 < i_2 < \ldots < i_n$ such that $a_1 \subseteq bi_1$, $a_2 \subseteq bi_2 \ldots a_n \subseteq bi_n$. For example, the sequence < (3) (4 5) (8) > is contained in < (7) (3 8) (9) (4 5 6) (8) >, since (3) $\subseteq$ (3 8), (4 5) $\subseteq$ (4 5 6) and (8) $\subseteq$ (8). However, the sequence < (3) (5) > is not contained in < (3 5) > (and vice versa). The former represents patterns 3 and 5 occurred one after the other, while the latter represents pattern 3 and 5 occurred together.

W. Ouyang et al. has proposed a randomized [43] and data perturbation [42] approach, which is simple and easy to be implemented, and has a rather good precision of support reconstruction. The latter approach does not decrease the support of true frequent sequences and can easily be combined with existing sequential pattern mining algorithms. The proposed work by Mhatre, A. et al. [44] demonstrates the effect of adding noisy data on sequential pattern mining algorithm over progressive database which is scalable from a single node system to a multi-party scenario.

V. Conclusion

Ethics refers to well-defined standards of right and wrong that prescribe what humans ought to do, usually in terms of rights, obligations, benefits to society, fairness, or specific virtues. Ethical standards also include those that enjoin virtues of honesty, compassion, and loyalty. Computer ethics is a set of moral principles that regulate the use of computers. Some common issues of computer ethics include intellectual property rights such as copyrighted electronic content, privacy concerns, and how computers affect society. The benefits of world-wide-web are plenty but threats to personal information in such as social networking sites also abound. Another area where computer ethics play an important part is the prolific data generated from various sources. Data mining technology has emerged as a means for identifying patterns and trends from such large quantities of data. With increasing usage of data mining in the public and private sectors, privacy assumes paramount importance. We provide an overview of privacy preserving association rule mining, which is one of the most popular pattern discovery methods in the rapidly emerging research area of privacy preserving data mining. Also, we have focused on of the most active areas of research in data mining – sequential pattern mining. We have summarized various proposals and algorithms designed in the research area of privacy preserving data mining and survey current existing techniques, and analyze their advantages and disadvantages.

References

[1] R. Agrawal, T. Imielinski, and A. Swami, "Mining Association Rules between Sets of Items in Large Databases," ACM SIGMOD Record, New York, 1993, pp. 207-216.

[2] A. Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke, "Privacy Preserving Mining of Association Rules," Information System, vol. 29, Apr. 2004, pp. 343-364.

[3] M. Atallah, E. Bertino, A. Elmagarmid, M. Ibrahim3, V. Verykios, "Disclosure Limitation of Sensitive Rules," Proc. IEEE Knowledge and Data Engineering Workshop, Chicago, Illinois, 1999, pp. 25-52.

[4] E. Dasseni, V. S. Verykios, A. K. Elmagarmid, and E. Bertino, "Hiding Association Rules by Using Confidence and Support," Proc. the 4th Information Hiding Workshop, Pittsburg, PA, 2001, LNCS 2137, pp. 369-383.

[5] S. R. M. Oliveira, and O. R. Zaiane, "Privacy Preserving Frequent Itemset Mining," Proc. IEEE ICDM Workshop on Privacy, Security, and Data Mining, Maebashi, Japan, 2002, pp. 43-54.

[6] S. R. M. Oliveira and O. R. Zaiane, "Algorithms for Balancing Privacy and Knowledge Discovery in Association Rule Mining," Proc. the 7th International Database Engineering and Applications Symposium, Hong Kong, China, 2003, pp. 54-63.

[7] Y. Wu, C.M. Chiang, and A.L.P. Chen, "Hiding Sensitive Association Rules with Limited Side-effects," IEEE Transactions on Knowledge and Data Engineering, vol. 19, Jan. 2007, pp. 29-42.

[8] S.R.M. Oliveira, O.R. Zaiane, and Y. Saygin, "Secure Association-Rule Sharing," Advances in Knowledge Discovery and Data Mining, Springer Berlin, Heidelberg,Vol. 3056,2004, pp. 74-85.

[9] V.S. Verykios, A.K. Elmagarmid, E. Bertino, Y. Saygin, E. Dasseni, "Association Rule Hiding," IEEE Transactions on Knowledge and Data Engineering, vol. 16, Apr. 2004, pp, 434-447.

[10] R. Agrawal, and R. Srikant, "Privacy-preserving data mining," ACM SIGMOD Record, New York, vol. 29, Feb. 2000, pp.439-450.

[11] Y. Saygin, V. S. Verykios, and C. Clifton, "Using Unknowns to Prevent Discovery of Association Rules," ACM SIGMOD Record, New York, vol. 30, Apr. 2001, pp. 45-54.

[12] M. Kantarcioglu, and C. Clifton, "Privacy-Preserving Distributed Mining of Association Rules on Horizontally Partitioned Data," IEEE Transactions on Knowledge and Data Engineering, vol. 16, Sep. 2004, pp. 1026-1037.

[13] J. Vaidya, and C. Clifton, "Privacy Preserving Association Rule Mining in Vertically Partitioned Data," Proc. the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, D.C. 2002,pp. 639-644.

[14] R. Agrawal and R. Srikant, Mining sequential patterns, In Proc. Of the 11th Int'l Conference on Data Engineering, pp3-14, Taipei, Taiwan, March 1995.

[15] Yanguang Shen, Junrui Han, HuiShao, "Research on Privacy-Preserving Technology of Data Mining", Second International Conference on Intelligent Computation Technology and Automation, pp 612-614 (2009).

[16] Yanguang Shen, Hui Shao, Yan Li, "Research on the Personalized Privacy Preserving Distributed Data Mining", Second International Conference of Future Information Technology and Management Engineering, 436 – 439 (2009).

[17] Vassilios S. Verykios, Ahmed k. Elmagarmid, Bertino Elisa, Yucel Saygin, Dasseni Elena, "Association Rule Hiding", IEEE Transactions on knowledge and Data Engineering (2003).

[18] Oleveira S R M, Zaiane O R, "Protecting Sensetive Knowledge by Data Sanitization." In:Proceedings of the Third IEEE International Conference on Data Mining, 2003.

[19] Wenliang Du, Zhijun Zhan. Building Decision Tree Classifier on Private Data [J]. Workshop on Privacy,Security and Data Mining at the 2002 IEEE International Conference on Data M ining,2002, Vol.14

[20] Y. Lindell, B.Pinkas. Privacy Preserving Data Mining [J]. Journal of Cryptology, 2002, 15:177–206.

[21] S. R. M. Oliveira and O. R. Zaiane, "Achieving Privacy Preservation When Sharing Data For Clustering", In Proceedings of the International Workshop on Secure Data Management in a Connected World(SDM'04)in conjunction with VLDH2004[16].Toronto,Canada: August,2004.

[22] S. Merngn, J. Ghosh. Privacy-Preserving Distributed Clustering Using Generative Models. In Proc. of the 3rd IEEE International,Conference on Data Mining (ICDM' 03), pages:211-218, Melbourne, Florida, USA, November 2003.

[23] Han Jiawei, and M. Kamber, Data Ming: Concepts and Techniques. Beijing: China Machine Press, 2006, pp. 227-250.

[24] Yongcheng Luo , Yan Zhao Jiajin Le, "A Survey on the Privacy Preserving Algorithm of Association Rule Mining", Second International Symposium on Electronic Commerce and Security, 241 – 245(2009)

[25] Vassilios S. Verykios, Elisa Bertino,et al., "State-of-the-art in Privacy Preserving Data Mining," SIGMOD Record, Vol. 33, No. 1, March 2004, pp.50-57.

[26] E.T. Wang, G. Lee, Y.T. Lin, "A novel method for protecting sensitive knowledge in association rules mining," In: Proceedings of the 29th IEEE Annual International Computer Software and Applications Conference (COMPSAC'05), Edinburgh, Scotland, 2005, pp.511–516.

[27] E.T. Wang, G. Lee, "An efficient sanitization algorithm for balancing information privacy and knowledge discovery in association patterns mining," Data Knowl.Eng. (2008), doi:10.1016/j.datak.2007.12.005.

[28] Jun Lin Lin, Yung Wei Cheng, "Privacy preserving item set mining through noisy items," Expert Systems with Applications, 2009, pp.5711– 5717.

[29] Yucel Saygin, Vassilios S. Verykios, and Ahmed K. Elmagarmid, "Privacy preserving association rule mining," In Proceedings of the 12th International Workshop on Research Issues in Data Engineering (2002), 151–158.

[30] L. Sweeney, "k-anonymity: a model for protecting privacy", International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 2002, pp.557–570.

[31] Xiao X, Tao Y, "Personalized privacy preservation", Proceedings of ACM Conference on management of Data (SIGMOD). ACM Press, New York: 2006, pp.785–790.

[32] LIU Ming, Xiaojun Ye, "Personalized K-anonymity", Computer Engineering and Design, Jan.2008, pp.282–286.

[33] Dakshi Agrawal and Charu C. Aggarwal, "On the design and quantification of privacy preserving data mining algorithms," In

Proceedings of the 20th ACMSymposium on Principles of Database Systems (2001), pp.247–255.
[34] Alexandre Evfimievski, Ramakrishnan Srikant, Rakesh Agrawal, and Johannes Gehrke, "Privacy preserving mining of association rules," In Proceedings of the 8th ACM SIGKDDD International Conference on Knowledge Discovery and Data Mining (2002).
[35] Rizvi S J, Haritsa J R., "Maintaining data privacy in association rule mining," In: Proceedings of t he 28th International Conference on Very Large Data Bases , Hong Kong , China , August 2002.
[36] Chris Clifton,Murat Kantarcioglou,XiadongLin,and Michaed Y.Zhu, "Tools for privacy preserving distributed data mining," SIGKDD Explorations 4 (2002), no. 2.
[37] A. Sanil, A. Karr, X. Lin, and J. Reiter, "Privacy preserving regression modelling via distributed computation," In Proc. Tenth ACM SIGKDD Internat. Conf. on Knowledge Discovery and Data Mining, 2004, pp.677–682
[38] A. Sanil, A. Karr, X. Lin, and J. Reiter, "Privacy preserving analysis of vertically partitioned data using secure matrix products," Journal of Official Statistics, 2007. Revised manuscript under review.
[39] ZongBo Shang; Hamerlinck, J.D., "Secure Logistic Regression of Horizontally and Vertically Partitioned Distributed Databases," Data Mining Workshops, ICDM Workshops 2007. Seventh IEEE International Conference on 28-31 Oct. 2007, pp.723–728.
[40] M. Kantarcioglu, C. Clifton, "Privacy-preserving distributed mining of association rules on horizontally partitioned data," The ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery (DMKD'02). ACM SIGMOD'2002 [C]. Madison, Wisconsin, 2002, pp.24–31.
[41] David W. Cheung, Jiawei Han, Vincent T. Ng,Ada W. Fu, and Yongjian Fu, "A fast distributed algorithm for mining association rules," In Proceedings of the 1996 International Conference on Parallel and Distributed Information Systems (1996).
[42] Wei-Min Ouyang; Hong-Liang Xin; Qin-Hua Huang, "Privacy Preserving Sequential Pattern Mining Based On Data Perturbation", International Conference on Machine Learning and Cybernetics, 3239 – 3243(2007).
[43] Weimin Ouyang; Qinhua Huang; Hongliang Xin,"A Randomization Approach To Mining Sequential Pattern With Privacy Preserving" International Symposium on Computational Intelligence and Design, 65 – 68(2008).
[44] Mhatre, A.; Verma, M.; Toshniwal, D., "Privacy Preserving Sequential Pattern Mining In Progressive Databases Using Noisy Data", International Conference on Information Visualisation, 456 – 460(2009).

Sunita M. Mahajan received her M.Sc. degree from Mumbai University in Physics in 1966, Ph D. in parallel processing from SNDT Women;s University, India in 1997. Currently she is Principal, Institute of Computer Science, at M.E.T, Mumbai, India. She worked in Bhabha Atomic Research Centre for 31 years. Her research areas are Parallel Processing, Distributed Computing, Data Mining and Grid Computing.

Alpa K. Reshamwala received her B.E degree in Computer Engineering from Fr. C.R.C.E, Bandra, Mumbai University in 2000, M.E degree in Computer Engineering from TSEC, Mumbai University in 2008. She is currently pursuing Ph D. degree in Computer Engineering from NMIMS University, Mumbai India. Presently, she is working as Assistant professor, at Mukesh Patel School of Technology and Management and Engineering, NMIMS University, Mumbai, India.