AENSI Journals

Australian Journal of Basic and Applied Sciences

ISSN:1991-8178

Journal home page: www.ajbasweb.com

A Review on Approaches, Techniques and Research Challenges in Privacy Preserving Data Mining

[1]M. Prakash and [2]G. Singaravel,

[1]Department of Computer Science and Engineering, K.S.R College of Engineering (Affiliated to Anna University, Chennai), Tiruchengode, Tamilnadu, India.
[2]Department of Information Technology,, K.S.R College of Engineering (Affiliated to Anna University, Chennai), Tiruchengode, Tamilnadu, India.

ARTICLE INFO

Article history:
Received 25 April 2014
Received in revised form
8 May 2014
Accepted 20 May 2014
Available online 17 June 2014

Keywords:
Data Mining, Anonymization, Privacy Preserving, Privacy Preserving Techniques, Data Publishing, Data Security

ABSTRACT

Background: Privacy Preserving Data Mining is a fast growing new era of research due to recent advancements in information, data mining, communications and security technologies. The Privacy Preserving Data Mining's ultimate goal is to develop efficient algorithms that allow one to extract relevant knowledge from large amount of data, while prevent sensitive information from disclosure or inference. Several data mining algorithms, incorporating privacy preserving mechanisms, have been developed over the last years. There is an emerging need of synthesizing literature to understand the nature of problem, identify potential research issues, standardize new research area, and evaluate the relative performance of different approaches. Objective: The main purpose of this study is to review the state-of-the-art in current Privacy Preserving Data Mining research in order to better understand existing algorithms, answer research questions and move forward the field of research. Results: In this paper, we provide a review of the approaches and techniques for privacy and analyze the representative technique for privacy preserving data mining and points out their merits and demerits. Conclusion: A comparison of work done by different authors is presented. Finally, the directions for future research are discussed.

To Cite This Article: M. Prakash, G. Singaravel,., A Review on Approaches, Techniques and Research Challenges in Privacy Preserving Data Mining. ***Aust. J. Basic & Appl. Sci.,*** *8(10): 251-259, 2014*

INTRODUCTION

Data Mining which is sometimes also called as Knowledge Discovery Data (KDD) is the process of analyzing data from different perspectives and summarizing it into useful information. Today, data mining is used by many organizations with a strong consumer focus such as retail, financial, communication, and marketing organizations. Extraction of hidden predictive information from large databases is a powerful new technology with great potential to help organizations focus on the most important information in their data warehouses. Various algorithms and techniques like Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm etc., are used for knowledge discovery from databases. In recent years, data mining has been used widely in the areas of science and engineering, such as bioinformatics, genetics, medicine, education and electrical power engineering. It has been said that knowledge is power, and this is exactly what data mining is about. It is the acquisition of relevant knowledge that can allow making strategic decisions.

Data mining is the process of extracting useful, interesting, and previously unknown information from large data sets. The success of data mining relies on the availability of high quality data and effective information sharing. The collection of digital information by governments, corporations, and individuals has created an environment that facilitates large-scale data mining and data analysis (Prakash, M and G, Singaravel., 2012). Moreover, driven by mutual benefits, or by regulations that require certain data to be published, there is a demand for sharing data among various parties.

Information sharing has a long history in information technology. Traditional information sharing refers to exchanges of data between a data holder and a data recipient. For example, the Electronic Data Interchange (EDI) is a successful implementation of electronic data transmission between organizations with the emphasis on the commercial sector. The development of EDI began in the late 1970s and remains in use today. Nowadays, the terms "information sharing" and "data publishing" not only refer to the traditional one-to-one

Corresponding Author: M. Prakash, Department of Computer Science and Engineering, K.S.R College of Engineering (Affiliated to Anna University, Chennai), Tiruchengode, Tamilnadu, India.
E-mail: mmsprakash.research@gmail.com

model, but also the more general models with multiple data holders and data recipients. Recent standardization of information sharing protocols, such as eXtensible Markup Language (XML), Simple Object Access Protocol (SOAP), and Web Services Description Language (WSDL) are catalysts for the recent development of information sharing technology.

Detailed data in its original form often contain sensitive information about individuals, and sharing such data could potentially violate individual privacy. For example, one may not want to share her web browsing history and the information of items in her intelligent fridge to a third party. The data recipient having access to such information could potentially infer the fridge owner's current health status and predict her future status. The general public expresses serious concerns on their privacy and the consequences of sharing their person-specific information.

Privacy Preserving Data Mining Model:

The Privacy Preserving Data Mining Model is shown in the Fig. 1. This model consists of various components. The data are collated from a number of sources called data sources which includes text, image, audio, video, spatial data, transactional data and more. However, the collated data need to be organized and cleaned to form a data warehouse so that a data mining techniques can be rapidly applied. The data warehouse is a huge collection of data, which can be used for further knowledge extraction process using the mining algorithms. The privacy preserving techniques can be applied on the data to be mined. Data mining algorithms can be now applied to extract the knowledge or patterns (Prakash, M and G, Singaravel., 2014). Here the extracted knowledge cannot contain the private data of any individual, because the privacy preserving techniques have been applied to ensure the privacy.

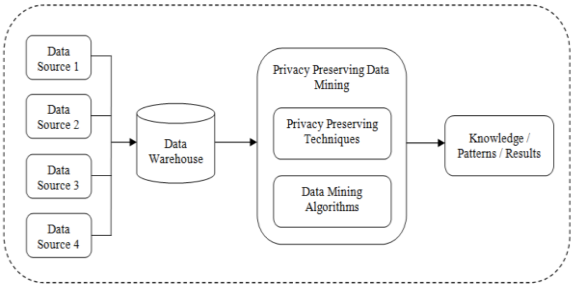
Data Source 1
Data Source 2
Data Source 3
Data Source 4
Data Warehouse
Privacy Preserving Data Mining
Privacy Preserving Techniques
Data Mining Algorithms
Knowledge / Patterns / Results

Fig. 1: Privacy Preserving Data Mining Model.

This model can be used to protect the information leakage of any individual's sensitive information in any forms of attack like, Homogeneity attack, Skewness attack, Similarity attack, Background knowledge, etc.,.

Privacy Preserving Data Mining Techniques:

The taxonomy of Privacy Preserving Data Mining (PPDM) Techniques (Verykios, S., *et al*., 2004) is depicted in the Fig. 2. In general this taxonomy is divided into four levels namely Data Distribution, Hiding Purpose, Data Mining Algorithms and Privacy Preserving Techniques.

Data Distribution:

The PPDM algorithms can be first divided into two major categories, centralized and distributed data, based on the distribution of data (Du, W and Z, Zhan., 2003). In a centralized database environment, data are all stored in a single database; while, in a distributed database environment, data are stored in different databases. Earlier research has been predominately focused on dealing with privacy preservation in a centralized Database. The difficulties of applying PPDM algorithms (Xia, Y., *et al*., 2004) to a distributed databases can be attributed to two reasons: first, the data owners have privacy concerns so they may not willing to release their own data for others; second, even if they are willing to share data for data mining, the communication cost between the sites is too expensive. In today's global digital environment, most data are often stored in different sites, thus, more attention and research should be focused on distributed PPDM algorithms.

Hiding Purpose:

The PPDM algorithms can be further classified into two types, data hiding and rule hiding, according to the purposes of hiding. Data hiding refers to the cases where the sensitive data from original database like identity, name, and address that can be linked, directly or indirectly, to an individual person are hided. In contrast, in rule hiding, the sensitive knowledge derived from original database after applying data mining algorithms are removed. Majority of the PPDM algorithms used data hiding techniques (Yang, Z., *et al*., 2006). This is especially true in a distributed database environment as the techniques can be used to prevent individual information from being discovered by other parties in the joint computational process. Most PPDM algorithms hide sensitive patterns by modifying data. Also, at present, the rule hiding techniques is only being adopted by association rule mining for centralized database.

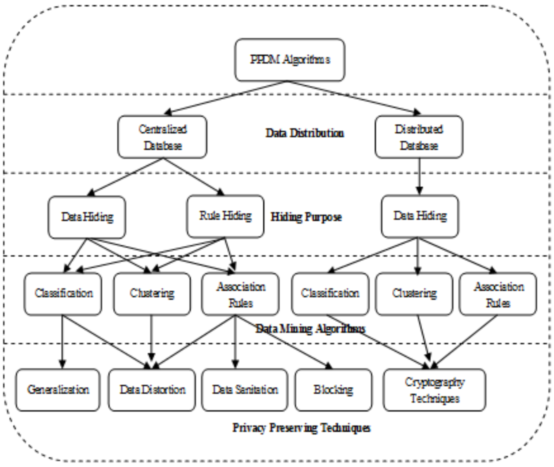

Fig. 2: The Taxonomy of Privacy Preserving Data Mining Techniques.

Data Mining Algorithms:

Currently, the PPDM algorithms are mainly used on the tasks of classification, association rule and clustering. Association analysis involves the discovery of associated rules, showing attribute value and conditions that occur frequently in a given set of data. Classification is the process of finding a set of models (or functions) that describe and distinguish data classes or concepts, for the purpose of being able to use the model to predict the class of objects whose class label is unknown. Clustering Analysis concerns the problem of decomposing or partitioning a data set (usually multivariate) into groups so that the points in one group are similar to each other and are as different as possible from the points in other groups. A majority of the PPDM algorithms used association rule method for mining data followed by classification and then clustering.

Privacy Preserving Techniques:

Four techniques – sanitation, blocking, distort, and generalization – have been used to hide data items for a centralized data distribution. Data sanitation is to remove or modify items in a database to reduce the support of some frequently used item sets such that sensitive patterns cannot be mined. The blocking approach replaces certain attributes of the data with a question mark. In this regard, the minimum support and confidence level will be altered into a minimum interval. As long as the support and/or the confidence of a sensitive rule lie below the middle in these two ranges, the confidentiality of data is expected to be protected. Data distort protects privacy for individual data records through modification of its original data, in which the original distribution of the data is reconstructed from the randomized data. These techniques aim to design distortion methods after which the true value of any individual record is difficult to ascertain, but "global" properties of the data remain largely unchanged. Generalization transforms and replaces each record value with a corresponding generalized value.

The privacy preservation technique used in a distributed database is mainly based on cryptography techniques. SMC algorithms deal with computing any function on any input, in a distributed network where each participant holds one of the inputs, while ensuring that no more information is revealed to a participant in the computation than can be inferred from that participant's input and output. Data distort is the most popular method used in hiding data followed by data sanitation and generalization. If one wants to obtain data mining results from different data sources, then the only method can be used is a cryptography technique. Since the parties who use SMC operators cannot reveal anything from others except final results, it can have benefits of both accuracy of data mining results and the privacy of the database (Bertino, E., *et al*., 2005). The Privacy Preserving Techniques are discussed further in detail in the following to understand in better.

A. Data Perturbation Approach:

Data Perturbation (Liu, J., *et al*., 2011) (Kargupta, H., *et al*., 2003) is a technique for modifying data using random process. This technique apparently distorts sensitive data values by changing them by adding, subtracting or any other mathematical formula. This technique can handle different data types: character type, Boolean type, classification type and integer. In discrete data, it is required to preprocess the original data set. The preprocessing of data is classified into attribute coding and obtaining sets coded data set. The method of average region to disperse the continuous data is used here. Discrete formula prescribed (Lohiya, S and L, Ragha., 2012) is:

A (max) - A (min)/n = length
Where, A is continuous attribute,
n is number of discrete,
length is the length of the discrete interval.

The technique does not reconstruct the original data values, it only reconstructs the distribution. Data distortion or data noise are different names for data perturbation. It is very important and critical to secure the sensitive data and data perturbation plays an important role in preserving the sensitive data. Distortion is done by applying different methods such as adding noise, data transpose matrix, by adding unknown values etc., (Jahan, T., *et al*., 2012). In some perturbation approaches it is very difficult to preserve the original data. Some of these are distribution based techniques. In order to overcome this problem, new algorithm were developed which were able to reconstruct the distributions. This means that for every individual problem in classification, clustering, or association rule mining, a new distribution based data mining algorithm needs to be developed. It develops a new distribution-based data mining algorithm for the classification problem, and developed the methods for privacy-preserving association rule mining.

A new approach in data perturbation was introduced (Jahan, T., *et al*., 2012). It was based on Singular Value Decomposition (SVD) and Sparsified Singular Value Distribution (SSVD) technique and having the feature of selection to reduce the feature space. In this method, different matrices have been introduced to compare or measure the difference between original dataset and distorted dataset. SSVD is efficient approach in keeping data utility, SVD also works better than other standard data distortion methods which add noise to the data to make it perturbed.

The perturbation approach has a drawback. The distribution of each data dimension is reconstructed independently. This means that any distribution based data mining algorithm works under an implicit assumption to treat each dimension independently. In many cases, a lot of relevant information for data mining algorithms such as classification is hidden in inter-attribute correlations.

B. Blocking Based Technique:

In blocking based technique (Lohiya, S and L, Ragha., 2012), authors state that there is a sensitive classification rule which is used for hiding sensitive data from others. In this technique, there are two steps which are used for preserving privacy. First is to identify transactions of sensitive rule and second is to replace the known values to the unknown values (?). In this technique, there is scanning of original database and identifying the transactions supporting sensitive rule. And then for each transaction, algorithm replaces the sensitive data with unknown values. This technique is applicable to those applications in which one can save unknown values for some attributes. Authors (Lohiya, S and L, Ragha., 2012) want to hide the actual values; they replace '1' by '0' or '0' by '1' or with any unknown (?) values in a specific transaction. The replacement of these values does not depend on any specific rule. The main aim of this technique is to preserve the sensitive data from unauthorized access. There may be different sensitive rules according to the requirements. For every sensitive rule, the scanning of original database is done. When the left side of the pair of rule is a subset of attribute values pair of the transaction and the right hand side of the rule should be same as the attribute class of the transaction then only transaction supports any rule. The algorithm replaces unknown values in the place of attribute for every transaction which supports that sensitive rule. These steps will continue till all the sensitive attributes are hidden by the unknown values.

C. Cryptographic Technique:

Cryptography is a technique through which sensitive data can be encrypted. It is a good technique to preserve the data. The cryptographic technique is introduced which is very popular because it provides security and safety of sensitive attributes. There are different algorithms of cryptography available. But this method has many disadvantages. It fails to protect the output of computation. It prevents privacy leakage of computation. This algorithm does not give fruitful results when it talks about more parties. It is very difficult to apply this algorithm for huge databases. Final data mining result may break the privacy of individual's record.

D. Condensation Approach:

Another approach used is Condensation approach. It was introduced (Aggarwal, C., *et al*., 2004) to build the constrained clusters in the data set and after that produces pseudo-data. The basic concept of the method is to contract or condense the data into multiple groups of predefined size. For each group, certain statistics are maintained. This approach is used in dynamic data update such as stream problems. Each group has a size of at least '*k*', which is referred to as the level of that privacy-preserving approach. The higher the level, the high is the amount of privacy. They use the statistics from each group in order to generate the corresponding pseudo-data. This is a simple privacy preservation approach but it is not efficient because it leads to loss of the information.

E. Anonymization:

When releasing micro data for research purposes, one needs to limit disclosure risks to an acceptable level while maximizing data utility. To limit disclosure risk, the k-anonymity privacy requirement is introduced (Sweeney, L., 2004), which requires each record in an anonymized table to be indistinguishable with at least *k* other records within the dataset, with respect to a set of quasi-identifier attributes. To achieve the k-anonymity requirement, they used both generalization and suppression for data anonymization. Unlike traditional privacy protection techniques such as data swapping and adding noise, information in a *k*-anonymous table through generalization and suppression remains truthful.

F. Randomized Response Techniques:

Consider a set of data records denoted by $X = \{x_1, x_2 \ldots x_n\}$. For record x_i, we add a noise component which is drawn from the probability distribution fY(y). These noise components are drawn independently, and are denoted as $y_1, y_2 \ldots y_n$. Thus, the new sets of distorted records are denoted by $(x_1 + y_1), (x_2 + y_2) \ldots (x_n + y_n)$. We denote this new set of records by $z_1, z_2 \ldots z_n$. In general, it is assumed that the variance of the added noise is large enough, so that the original record values cannot be easily guessed from the distorted data. Thus, the original records cannot be recovered, but the distribution of the original records can be recovered. Thus, if X be the random variable denoting the data distribution for the original record, Y be the random variable describing the noise distribution, and Z be the random variable denoting the final record, we have:

$$Z = X + Y$$
$$X = Z - Y$$

Now, we note that N instantiations of the probability distribution Z are known, whereas the distribution Y is known publicly. For a large enough number of values of N, the distribution Z can be approximated closely by using a variety of methods such as kernel density estimation. By subtracting Y from the approximated distribution of Z, it is possible to approximate the original probability distribution X. In practice, one can combine the process of approximation of Z with subtraction of the distribution Y from Z by using a variety of iterative methods. Such iterative methods typically have a higher accuracy than the sequential solution of first approximating Z and then subtracting Y from it. The basic idea of randomized response is to scramble the data in such a way that the central place cannot tell with probabilities better than a pre-defined threshold whether the data from a customer contain truthful information or false information. Although information from each individual user is scrambled, if the number of users is significantly large, the aggregate information of these users can be estimated with decent accuracy. One key advantage of the randomization method is that it is relatively simple, and does not require knowledge of the distribution of other records in the data. Therefore, the randomization method can be implemented at data collection time, and does not require the use of a trusted server containing all the original records in order to perform the Anonymization process.

G. Hybrid Technique:

Privacy preservation is a very huge field. Many algorithms have been proposed in order to secure the data. Hybrid technique is a new technique through which one can combine two or more techniques to preserve the data. A hybrid technique (Lohiya, S and L, Ragha., 2012) is proposed in which they used randomization and generalization. In this approach first they randomize the data and then generalized the modified or randomized data. This technique protects private data with better accuracy; also it can reconstruct original data and provide data with no information loss. Many other techniques can also be combined to make a hybrid technique such as Data perturbation, Blocking Based Method, Cryptographic Technique, Condensation Approach etc.

Privacy Preserving Data Mining Techniques – A Comparission:

There are many different techniques proposed in the field of Privacy Preserving Data Mining but one outperforms over other or vice versa on different criteria. Algorithms are classified on the basis of performance, utility, cost, complexity, tolerance against data mining algorithms etc. We have shown a tabular comparison of the work done by different authors is shown in Table 1. We have taken the parameters like technique used for PPDM, its approach, results and accuracy.

Table 1: Comparison of Privacy Preserving Data Mining Techniques.

Authors and Year of Publication	Technique Used for PPDM	Approach	Result and Accuracy
Aggarwal, C and Yu. P.S., 2004	Condensation Approach	This approach works with pseudo-data rather than with modifications of original data, this helps in better preservation of privacy than	The use of pseudo-data no longer necessitates the redesign of data mining algorithms, since they have the same format as the original data.

		techniques which simply use modifications of the original data.	
Alotaibi, K., *et al*., 2012	Multi-Dimensional Scaling	A non linear dimensionality reduction technique used to project data on lower dimensional space.	The application of nonmetric Multi Dimensional Scaling transformation works efficiently and hence produces better results.
Beck, M and M, Marhofer., 2012	Anonymizing Demonstrator	Making a demonstrator with user friendly interface and performs Anonymization.	Swapping and Recording can be applied to enhance the utility.
Deivanai, P., *et al*., 2011	Hybrid Approach	Hybrid Approach is a combination of different techniques which combine to give an integrated result.	It uses Anonymization and Suppression to preserve data.
Huang, H.C and W. C. Fang., 2011	Histogram Based Reversible Data Hiding	A concept of reversibility which states that an original data can easily be hidden and the hidden data can also be recovered perfectly. Sensitive data is embedded into medical images which is very good technique for hiding secret data.	Histogram technique is basically used for X-Ray or CT medical images and it has the potential to be integrated into databases for managing the medical images in the hospital.
Jahan, T., *et al*., 2012	Data Perturbation Using Sparsified Singular Value Distribution	An analyzing system used to transform original dataset into distorted data set using Sparsified Singular Value Decomposition.	Use of Sparsified Singular Value Distribution than Singular Value Distribution is more successful.
Kargupta, H., *et al*., 2003	Data Perturbation	They tried to preserve data privacy by adding random noise, while making sure that the random noise still preserves the “signal” from the data so that the patterns can still be accurately estimated.	Randomization – based Techniques are used to generate random matrices.
Karthikeswarant, D., *et al*., 2012	Association Rule	Sanitizes datasets using Sliding Window Algorithm and preserves data.	A novel approach that modifies the database to hide sensitive rules.
Kisilevich, S., *et al*., 2010	Anonymization	Anonymization is a technique for hiding individual’s sensitive data from owner’s record. *k*-anonymity is used for generalization and suppression for data hiding.	Background Knowledge and Homogeneity attacks of *k*-Anonymity Algorithm do not preserve sensitivity of an individual.
Komishani. E.G and M, Abadi., 2012	Trajectory data	Approach for privacy Preservation in trajectory data publishing in which trajectories and sensitive attributes are generalized with respect to different privacy requirements of moving objects.	It is able to provide personalized privacy preservation in trajectory data publishing, but also it is resistant to all three identity linkage, attribute linkage, and similarity attacks.
Kumbhar, M. N and R, Kharat., 2012	Association Rule by Horizontal and Vertical Distribution	Different approaches in the field of Association rule are reviewed.	The performance of all models is analyzed in terms of privacy, security and communications.
LeFevre, K., *et al*., 2006	Multidimensional *k*-Anonymity	A new multidimensional recoding model and a greedy algorithm for *k*-Anonymization	The greedy algorithm is substantially more efficient than proposed optimal *k*-Anonymization algorithms for single-dimensional models. The time complexity of the greedy algorithm is *O*(*nlogn*), while the optimal algorithms are exponential in the worst case. The greedy multidimensional algorithm often produces higher-quality results than optimal single dimensional algorithms.
Lindell, Y and B. Pinkas., 2000	Cryptographic Technique	A technique through which sensitive data can be encrypted. There is also a proper toolset for algorithms of cryptography.	This approach is especially difficult to scale when more than a few parties are involved. Also it does not hold good for large databases.
Liu, J., *et al*., 2011	Rating Based Privacy Preservation	A novel algorithm which overcomes the curse of dimensionality and provides privacy.	It is better than *k*-Anonymity and *l*-Diversity.
Lohiya, S and L, Ragha., 2012	Hybrid Approach	This approach uses the combination of *k*-Anonymity and Randomization.	It has a better accuracy and original data can be reconstructed.
Machanavajjhala, A., *et al*., 2006	*l*-Diversity Algorithm	If there are ‘*l*’ ‘well represented’ values for sensitive attribute then that class is said to have *l*- Diversity.	It is better than *k*-Anonymity in Privacy Preserving Data Mining.
Mathew, G and Obradovic, Z., 2011	Decision Tree	An approach which is technical, methodological and should give judgmental knowledge.	A graph-based framework for preserving patient’s sensitive information.
Mumtaz, S., *et al*.,	Distortion Based	Data perturbation technique which is	This distribution of distortion

2011	Perturbation Technique in OLAP Data Cube	also called uniformly adjusted distortion is proposed which initially distorts one cell of a cube and then distortion occurs in whole cube.	technique not only preserves, but also provides utmost accuracy with range sum queries and high availability.
Ninghui Li., *et al.*, 2007	*t*-Closeness	The distribution of sensitive attribute in any equivalence class to be close to the distribution of the attribute in the overall table. The distance between the two distributions should be no more than a threshold *t*	This effectively limits the amount of individual-specific information an observer can learn. However, an analysis on data utility shows that *t*-closeness substantially limits the amount of useful information that can be extracted from the released data.
Parmar, A., *et al.*, 2011	Blocking Based Technique	Finding sensitive attribute and then they replace known sensitive values with unknown values ("?"). Finally the sanitized dataset is generated from which sensitive classification rules are no longer mined.	Unknown Values help in preserving privacy but reconstruction of original data set is quite difficult.
Sweeney, L., 2002	*k*- Anonymity	A record from a dataset cannot be distinguished from at least *k*-1 records whose data is also in the dataset.	*k*-Anonymity Approach is able to preserve privacy.
Vaidya, J and C. Clifton., 2002	Association Rule	Distribution of data vertically into segments.	This Distribution Based Association Rule Data Mining provides privacy.

Research Challenges:

The challenge is to extract useful information while, at the same time, maintaining privacy. Several efforts are under way that attempt to preserve privacy throughout data mining. The different techniques, like randomization, cover stories, or multiparty policy enforcement, can be used to preserve privacy while data mining. While there is some progress in this area, the effectiveness of such techniques needs further evaluation. Some issues impacting privacy constraints on data mining are discussed so far. Here the key is an ability to measure privacy. Since privacy has many meanings, it required a set of metrics. Another use of this metric is to evaluate the inherent loss of privacy caused by data mining results or outcomes. The use of conditional privacy enables to estimate how much privacy is lost by knowing the data mining results even with a "perfect" privacy-preserving technique such as secure multiparty computation. The literature has not yet addressed this issue; the assumption has generally been that the data mining results do not of themselves violate privacy.

Conclusion:

Privacy Preserving Data Mining is a fast growing new era of research due to recent advancements in information, data mining, communications and security technologies. In this study it concludes that the Privacy Preserving Data Mining's ultimate goal is to develop efficient algorithms that allow one to extract relevant knowledge from large amount of data, while prevent sensitive information from disclosure or inference. The state-of-the-art in current Privacy Preserving Data Mining research is reviewed in order to better understand existing algorithms, answer research questions and move forward the field of research. Several data mining algorithms, incorporating privacy preserving mechanisms, have been developed over the last years are discussed. The emerging need of synthesizing literature is learned to understand the nature of problem, identify potential research issues, standardize new research area, and evaluate the relative performance of different approaches. In this paper, the approaches and techniques for privacy and the representative technique for privacy preserving data mining are reviewed and analyzed and their merits and demerits are pointed out. A comparison of work done by different authors is presented and the directions for future research are also discussed.

REFERENCES

Aggarwal, C. and P.S. Yu, 2004. "A condensation approach to privacy preserving data mining," *Proceedings of International Conference on Extending Database Technology (EDBT),* pp: 183-199.

Alotaibi, K., V.J. Rayward Smith, W. Wang and Beatriz de la Iglesia, 2012. "Non-linear Dimensionality Reduction for Privacy-Preserving Data Classification," *Proceedings of 2012 ASE/IEEE International Conference on Social Computing, Privacy, Security, Risk and Trust.*

Bertino, E., I. Fovino and L. Provenza, 2005. "A Framework for Evaluating Privacy Preserving Data Mining Algorithms," *Journal of Data Mining and Knowledge Discovery,* 11(2): 121-154.

Deivanai, P., J. Jesu Vedha Nayahi and V. Kavitha, 2011. "A Hybrid Data Anonymization integrated with Suppression for Preserving Privacy in mining multi party data," *Proceedings of International Conference on Recent Trends in Information Technology, IEEE.*

Du, W. and Z. Zhan, 2003. "Using Randomized Response Techniques for Privacy-Preserving Data Mining," *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, D.C.*

Huang, H.C. and W.C. Fang, 2011. "Integrity Preservation and Privacy Protection for Medical Images with Histogram-Based Reversible Data Hiding," *Proceedings of IEEE/NIH Life Science Systems and Applications Workshop, IEEE.*

Jahan, T., G. Narsimha and C.V. Guru Rao, 2012. "Data Perturbation and Features Selection in Preserving Privacy," *Proceedings of Ninenth International Conference on Wireless and Optical Communications Networks (WOCN), IEEE.*

Kargupta, H., S. Datta, Q. Wang and K. Sivakumar, 2003. "On the Privacy Preserving Properties of Random Data Perturbation Techniques," *Proceedings of the Third IEEE International Conference on Data Mining, IEEE.*

Karthikeswarant, D., V.M. Sudha, V.M. Suresh and A.J. Sultan, 2012. "A Pattern based framework for privacy preservation through Association rule Mining," *Proceedings of International Conference on Advances in Engineering, Science and Management (ICAESM -2012), IEEE.*

Kisilevich, S., L. Rokach, Y. Elovici and B. Shapira, 2010. "Efficient Multi-Dimensional Suppression for *k*-Anonymity," *Proceedings of IEEE Transactions on Knowledge and Data Engineering,* 22(3): 334-347.

Komishani, E.G. and M. Abadi, 2012. "A Generalization-Based Approach for Personalized Privacy Preservation in Trajectory Data Publishing," *Proceedings of Sixth International Symposium on Telecommunications (IST'2012), IEEE.*

Kumbhar, M.N. and R. Kharat, 2012. "Privacy Preserving Mining of Association Rules on horizontally and Vertically Partitioned Data: A Review Paper," *Proceedings of Twelfth International Conference on Hybrid Intelligent Systems (HIS), IEEE.*

LeFevre, K., D.J. DeWitt and R. Ramakrishnan, 2006. "Mondrian: Multidimensional k-Anonymity," *Proceedings of International Conference on Data Engineering.*

Lindell, Y. and B. Pinkas, 2000. "Privacy preserving data mining," *Proceedings of Journal of Cryptology, 5(3).*

Liu, J., J. Luo and J.Z. Huang, 2011. "Rating: Privacy Preservation for Multiple Attributes with Different Sensitivity requirements," *Proceedings of Eleventh IEEE International Conference on Data Mining Workshops.*

Lohiya, S. and L. Ragha, 2012. "Privacy Preserving in Data Mining Using Hybrid Approach," *Proceedings of Fourth International Conference on Computational Intelligence and Communication Networks, IEEE.*

Machanavajjhala, A., J. Gehrke, D. Kifer and M. Venkitasubramaniam, 2006. "*l*-Diversity: Privacy Beyond *k*-Anonymity," *Proceedings of International Conference on Data Engineering (ICDE).*

Martin Beck and Michael Marhofer, 2012. "Privacy-Preserving Data Mining Demonstrator," *Proceedings of Sixteenth International Conference on Intelligence in Next Generation Networks, IEEE.*

Mathew, G. and Z. Obradovic, 2011. "A Privacy-Preserving Framework for Distributed Clinical Decision Support," *Proceedings of First International Conference on Computational Advances in Bio and Medical Sciences (ICCABS), IEEE.*

Mumtaz, S., A. Rauf and S. Khusro, 2011. "A Distortion Based Technique for Preserving Privacy in OLAP Data Cube," *Proceedings of First International Conference on Computer Networks and Information Technology (ICCNIT), IEEE.*

Ninghui Li, Tiancheng Li and Suresh Venkatasubramanian, 2007. "*t*-Closeness: Privacy Beyond *k*-Anonymity and *l*-Diversity", *Proceedings of International Conference on Data Engineering (ICDE),* pp: 106-115.

Parmar, A., U.P. Rao and D.R. Patel, 2011. "Blocking based approach for classification Rule hiding to Preserve the Privacy in Database," *Proceedings of International Symposium on Computer Science and Society, IEEE.*

Prakash, M. and G. Singaravel, 2012. "A New Model for Privacy Preserving Sensitive Data Mining", *Proceedings of Third International Conference on Computing Communication and Networking Technologies (ICCCNT), IEEE.*

Prakash, M. and G. Singaravel, 2014. "An Analysis of Privacy Risks and Design Principles for Developing Countermeasures in Privacy Preserving Sensitive Data Publishing," *Journal of Theoretical and Applied Information Technology, 62(1): 204-213.*

Sweeney, L., 2002. "*k*-Anonymity: A Model for Protecting Privacy," *Proceedings of International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems.*

Vaidya, J. and C. Clifton, 2002. "Privacy preserving association rule mining in vertically partitioned data," *The Eighth ACM SIGKDD International conference on Knowledge Discovery and Data Mining, Edmonton, Alberta, CA, IEEE.*

Verykios, S., E. Bertino, I. Fovino, L. Provenza, Y. Saygin and Y. Theodoridis, 2004. "State-of-the-art in Privacy Preserving Data Mining," *ACM SIGMOD Record,* 33(1): 50–57.

Xia, Y., Y. Yang, Y. Chi and R.R. Muntz, 2004. "Mining Association Rules with Non-uniform Privacy Concerns," *Technical Report CSD-TR No. 040015, University of California.*

Yang, Z., S. Zhong and R.N. Wright, 2006. “GrC – Privacy Preserving Model Selection,” *Proceedings of the IEEE International Conference on Granular Computing.*