International Journal of Information Technology and Management
Vol. XI, Issue No. XVII, November-2016, ISSN 2249-4510

Data Mining Method for Use Identify Deliberate Deception and Secure Privacy

Kale Deepali Anil[1]* Dr. Suneel Kumar[2]

[1]Ph.D. Research Student, MUIT, Lucknow

[2]Research Guide, MUIT, Lucknow

Abstract – In new era the information and data communication technologies are highly used in the Business Industry. The data warehouse is used in the significant business value by improving the effectiveness of managerial decision-making. Naturally such a process may open up new assumption dimensions, detect new invasion patterns, and raises new data security problems. Due to the explosive development of Internet browsers can extract desired data from large amount of database. Data mining is a collection of technique helps the miner can retrieve exact knowledge from it. Privacy Preservation has been one of the greater concerns in data mining; there are many methods and techniques for privacy Preservation data mining. This paper discussed very keen points of various Privacy Preservation data mining algorithms and analyzing techniques and concludes the advantages and disadvantages, and a way to direct the problems. Recent developments in information technology have enabled collection and processing of enormous amount of personal data, such as criminal records, shopping habits, banking, credit and medical history, and driving records. This information is undoubtedly very useful in many areas, including medical research, law enforcement and national security. Data Storage, data access efficiently and speedily is not only the key for competitiveness but the date security and privacy is important.

Keywords: Data Mining Identify, Deliberate, Deception, Secure Privacy, Important, etc.

----------X----------

INTRODUCTION

Data mining consists of number of techniques for manufacture automatically and interestingly to retrieve the information from the large amount of database which consists of sensitive information also. It requires data preparation which can uncover information which may compromise confidentiality and privacy obligations. Efficient data mining technique has increases the disclosure risks of sensitive data. The general way for this to happen because of data aggregation. Data aggregation is used for when the data are accrued, possibly from various sources and put together, so that they can be analyzed. But the data mining by it result the preparation of data before, for the purposes of analysis. To make up a publicly available scheme secure , must ensure not only the private sensitive data to be fit out, but also to build sure that certain inference channels have to prevented as well. An individual's privacy comes under a struggle when the data previously composed, that data may roots the data miner, or newly created data sets, are able to make out specific individuals, especially when initially the data were anonymous. Security to sensitive data against unauthorized access has been a long term goal for the database security research community. Hence, Privacy preservation data mining is novel research direction in data mining. It composed of number of effective method and techniques to make sure that might result in information loss, side effects, improve accuracy, utility and efficiently.

REVIEW OF LITERATURE:

The advent of information technology in various fields of human life has led to the large amount of data storage in various formats like records, documents, images, sound recording, videos, scientific data and many new data formats. The Data collected from different applications require proper mechanism of extracting knowledge/ information from large repositories for better decision making Knowledge discovery in Databases (KDD), often called data mining, aims at the discovery of useful information from large collection of Data (Jiawei, et. al., 2002). The field of data mining is gaining significance recognition to the availability of large amounts of data, easily collected and stored via computer systems. Recently, the large amount of data, gathered from various channels, contains much personal information. When personal and sensitive

data are published and/or analyzed, one important question to take into account is whether the analysis violates the privacy of individuals whose data is referred to. The importance of privacy is growing constantly. For this reason, many research works have focused on privacy-preserving data mining, proposing novel techniques that allow extracting knowledge while trying to protect the privacy of users. Some of these approaches aim at individual privacy while others aim at corporate privacy. Data mining, popularly known as Knowledge Discovery in Databases (KDD), it is the nontrivial extraction of implicit, previously unknown and potentially useful information from data in databases. Though, data mining and knowledge discovery in databases (or KDD) are frequently treated as synonyms, data mining is actually part of the knowledge discovery process (Verykios, et. al., 2004), (Zhang, 2006), (Agrawal and Srikant, 2000), Usually, data mining e.g. data or knowledge discovery is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases (Evfimievski, et. al., 2004). Although data mining is a comparatively new term but the technology is not. Companies have used powerful computers to filter through volumes of superstore scanner data and analyze market research reports for many years. However, continuous innovations in computer processing power, disk storage, and statistical software are dramatically increasing the accuracy of analysis while driving down the cost. Data mining, the discovery of new and interesting patterns in large datasets, is an exploding field. One aspect is the use of data mining to improve security, e.g., for intrusion detection. A second aspect is the potential security hazards posed when an adversary has data mining capabilities.

CLASSIFICATION OF PRIVACY DATA MINING:

Table 1 - Privacy Data Mining

Data Hiding	Data Perturbation	Value Distortion	Additive Perturbation Multiplicative Perturbation Data Microaggregation Data Anonymization Data Swapping Other Randomization Techniques
		Probability Distribution	Sampling Method Analytical Method
	Secure Multi-Party Computation (SMC) / Cryptographic Protocols		
	Distributed Data Mining (DDM)		
Rule Hiding	Association Rule Hiding		Data Perturbation Data Blocking
	Classification Rule Hiding		Parsimonious Downgrading

Privacy Preservation Data Mining: Privacy has been gaining more attention to handle the terrorism, the government needed to examine, using data mining technology, more information about individuals to detect unusual disease outbreaks, financial fraudulent behaviors, network intrusions, etc. While all of these applications of data mining can benefit our society, there is also a negative side to this technology because it could be a threat to the individuals' privacy. Overcome the "limitations" of data mining techniques including areas like data security and privacy preserving data mining, which are actually active and growing research areas (Kargupta, et. al., 2003).

Data Distribution: The PPDM algorithm can be divided into two major categories, Centralized and distributed data. In a centralized database, data are stored in a single database, in distributed data can further classified into horizontal a vertical data distributions. In Horizontal data distribution from different records of the same data attributes are resided in different places. In vertical data distributor different attributes of the same record of data are resided in different places most research occurred on a centralized data base. Applying PPDM algorithm to a distributed database privacy concerns, communication cost is too expensive.

Purpose of PPDM: The PPDM algorithms main purpose is hiding is data hiding and rule hiding. In Data Hiding the sensitive data from original database like identify name and address are linked, directly or indirectly to an individual person are hided. In rule hiding the sensitive data (rule) from original database after applying data mining algorithm is removed. Most of the PPDM algorithms hide sensitive patterns by modifying data hiding.

PPDM Algorithm: The PPDM algorithm are specifically on the tasks of classification, association rule and clustering classification is the process of finding a set of models that describe and distinguish data classes or concepts, for the purpose of the model is used for prediction the class of objects whose label is unknown clustering analysis concerns the problem of separating a data set in one group which are similar top each other and are different as possible in other group.

PPDM Techniques: PPDM techniques used by four categories Sanitation, it can remove or modify items for a database to reduce the support of some frequently used items sets that sensitive patterns are not to be mined. Blocking it can replace certain attributes the data with a question mark. According to this the minimum support and confidence level will be altered into a minimum interval. In distort, the support and the confidence of a sensitive rule lie below the middle the two and the confidentiality of data is expected to be protected and also known as data perturb action or data randomization, where individual data records are modified from original data, and reconstructed from randomized data (Huang, et. al., 2005). This technique aims to design for distortion methods after which the true value of any individual

record is difficult to ascertain, but unchanged for danger data. In generalization transforms and replaces each record value with a correspondery generalized value.

DATA SECURITY ISSUES:

One of the key issues raised by data mining technology is not a business or technological one, but a social one. It is the issue of individual privacy. Data mining makes it possible to analyze routine business transactions and glean a significant amount of information about individuals buying habits and preferences. Another issue is that of data integrity. Clearly, data analysis can only be as good as the data that is being analyzed. A key implementation challenge is integrating conflicting or redundant data from different sources. For example, a bank may maintain credit cards accounts on several different databases. The addresses (or even the names) of a single cardholder may be different in each. Software must translate data from one system to another and select the address most recently entered. Finally, there is the issue of cost. While system hardware costs have dropped dramatically within the past five years, data mining and data warehousing tend to be self-reinforcing (Mishra, et. al., 2012). The more powerful the data mining queries, the greater the utility of the information being gleaned from the data, and the greater the pressure to increase the amount of data being collected and maintained, which increases the pressure for faster, more powerful data mining queries. This increases pressure for larger, faster systems, which are more expensive. Data mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining tools can answer business questions that traditionally were too time consuming to resolve (Guo, et. al., 2007). They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations.

EXPLOITATION OF DATA MINING:

Define Data mining is used for a variety of purposes in both the private and public sectors. Industries such as banking, insurance, medicine, and retailing commonly use data mining to reduce costs, enhance research, and increase sales. For example, the insurance and banking industries use data mining applications to detect fraud and assist in risk assessment. Using customer data collected over several years, companies can develop models that predict whether a customer is a good credit risk, or whether an accident claim may be fraudulent and should be investigated more closely. The medical community sometimes uses data mining to help predict the effectiveness of a procedure or medicine. Pharmaceutical firms use data mining of chemical compounds and genetic material to help guide research on new treatments for diseases. Retailers can use information collected through affinity programs to assess the effectiveness of product selection and placement decisions, coupon offers, and which products are often purchased together. Companies such as telephone service providers and music clubs can use data mining to create a “churn analysis,” to assess which customers are likely to remain as subscribers and which ones are likely to switch to a competitor.

CONCLUSION:

This paper carries out various approaches for privacy preservation data mining and analysis techniques and method what are existing. All the proposed methods are just approximate to achieve the goal of privacy upon some extend. Firstly, new algorithm with better approximation ratio and/or time complexity in this framework needs to be under development, still introduce considerable information loss with high-dimensional metric space involved. Data mining has become one of the key features of many homeland security initiatives. Often used as a means for detecting fraud, assessing risk, and product retailing, data mining involves the use of data analysis tools to discover previously unknown, valid patterns and relationships in large data sets. In the context of homeland security, data mining can be a potential means to identify terrorist activities, such as money transfers and communications, and to identify and track individual terrorists themselves, such as through travel and immigration records. While data mining represents a significant advance in the type of analytical tools currently available, there are limitations to its capability. One limitation is that although data mining can help reveal patterns and relationships, it does not tell the user the value or significance of these patterns. These types of determinations must be made by the user. A second limitation is that while data mining can identify connections between behaviors and/or variables, it does not necessarily identify a causal relationship. Successful data mining still requires skilled technical and analytical specialists who can structure the analysis and interpret the output. Data mining is becoming increasingly common in both the private and public sectors. Industries such as banking, insurance, medicine, and retailing commonly use data mining to reduce costs, enhance research, and increase sales. In the public sector, data mining applications initially were used as a means to detect fraud and waste, but have grown to also be used for

purposes such as measuring and improving program performance. One issue is data quality, which refers to the accuracy and completeness of the data being analyzed. A second issue is the interoperability of the data mining software and databases being used by different agencies. A third issue is mission creep, or the use of data for purposes other than for which the data were originally collected.

REFERENCES:

A. Evfimievski, R. Srikant, R. Agrawal, J. Gehrke (2004). "Privacy Preserving Mining of Association Rules", Information System, vol.29, no.4, pp. 343-364.

Guo, S. Guo, X. Wu (2007). "Privacy Preserving Market Basket Data Analysis", In Proceedings the 11th European Conference on Principles and Practice of Knowledge Discovery in Databases, pp. 103-114.

H. Kargupta, S. Datta, Q. Wang, K. Sivakumar (2003). "On the Privacy Preserving Properties of Random Data Perturbation Techniques", In Proceedings of the 3rd International Conference on Data Mining, pp. 99-106.

Han Jiawei, M. Kamber, and Data Mining (2002). Concepts and Techniques, Beijing: China Machine Press, Privacy, Security, and Data Mining, pp.1-8.

N. Zhang (2006). "Privacy-Preserving Data Mining", Texas A&M University, pp.19-25.

Pragyaban Mishra, Neelamadhab Pandhy and Rasmita Panigrahi (2012). " The Survey of Data Mining Applications and Feature Scope", Asian Journal of Computer Science And Information Technology 2-4, pp. 68-77.

R. Agrawal, R. Srikant (2000). "Privacy-Preserving Data Mining", ACM SIGMOD Record, New York, vol.29, no.2, pp.439- 450.

V. S. Verykios, E. Bertino, I. N. Fovino, L. P. Provenza, Y. Saygin, Y. Theodoridis, (2004). "State-of-the-art in Privacy Preserving Data Mining", New York, ACM SIGMOD Record, vol.33, no.2, pp.50-57.

Z. Huang, W. Du, B. Chen (2005). "Deriving Private Information from Randomized Data", In Proceedings of the ACM SIGMOD Conference on Management of Data, Baltimore, Maryland, USA, pp. 37-48.

Corresponding Author

Kale Deepali Anil*

Ph.D. Research Student, MUIT, Lucknow

E-Mail – d_a_kale@yahoo.co.in