I J C T A, 9(28) 2016, pp. 211-216

SECLUSION OF DATA IN DATABASES USING DISTINCT METHODOLOGY

Ishwarya. M.V* and K. Ramesh kumar**

Abstract: Data mining and knowledge discoveries are the two recent research areas which enquire the extraction of large quantity of data of unknown patterns. Data mining is where non-trivial and useful knowledge are drawn out from large databases.Sequential data mining techniques are successful in various areas such as science, engineering and medicine. The need for data mining that is both distributed and parallel has been in existence over the past years. Data mining research is concerned with obtaining information from various areas such as bio informatics, customer relationship management etc. The information obtained can be in the form of patterns or clusters. Consider the example where association rules in a super market which gives the relationship among the items bought together. Here customers could be clustered in form of segments. Data mining techniques has its use in various security applications for identifying behavior, link analysis which deals with multi -part databases. The major part of data mining is to extract the information which is hidden from large databases which results in elevated effort for collection of data by companies. Obviously this concerns about the data which is collected as in [2]. This has resulted in developing data mining techniques by the data mining researchers under privacy preserving data mining which can be used without interrupting the private policies of people and companies. Under the new era of research, algorithms which were already present should be reviewed. Various techniques have been proposed for the existing privacy preserving data mining. Names, Identifiers must be changed as they are sensitive data. The important thing in privacy in tuple is to develop algorithms for changing the original data. Database interface problem occurs when confidential information can be obtained from data released by unauthorized users.

Keywords: Slicing, Shuffling, Sensitive Information, Data Partitioning, Generalization, Bucketization

1. INTRODUCTION

The challenge faced in privacy of data is to share data by avoiding disclosure. The sharing technique is done in such a way that the no-public information are hidden. In addition to this, privacy preserving in a multidimensional record is of major concern. Overlapping slicing is a new approach for data anonymization and provides better utility while protecting against privacy threat. This overcomes the limitation of bucketisation and generalization.

2. LITERATURE REVIEW

The idea of privacy preservation, by overcoming the bucketization and generalization has been designed. Anonymity is very powerful technique. But there are huge losses of data in generalization technique when it comes to High Dimensional data which are encountered in medicine when a large amount of measurements are produced at once and so it is unreliable in High dimensional data. If Bucketisation does not provide a disclosure on the memberships on data and cannot work on data, which is incapable of separating the quasi and sensitive attributes. The technique defines that the sliced data where the efficiency is improved when compared with the existing techniques. Data slicing is efficient than generalization and bucketization and there is no need of separation of the quasi identifier and the sensitive attributes. The sliced data is most efficient which also handles data disclosure. Slicing of data is capable of giving

* Research Scholar, HITS, Assistant Professor, CSE Dept, Sri SaiRam Engineering College, West Tambaram, Tamil Nadu, India
Email: ishwarya.cse@sairam.edu.in
** Associate Professor, IT Dept, HITS, Tamil Nadu, India, Email: krkumar@hindustanuniv.ac.in

better utility of data since sensitive attributes are involved. Today mainly organization are involved in publishing micro data which contains tuples of which the data is an entity or real world object. This is done using generalization and bucketisation with l-diversity.

In the methods the columns are divided into three classifications with identifiers such as social security number or name. The second category is the quasi-identifiers, which is set of elements, which in combination with external information can be used to identify the records details such as birth date, sex. The third classification is non public information to make it unknown to others. In every methods first the attributes are divide to columns then to buckets and the data in the tuples are transformed to less specific semantically constant which cannot distinguish itself from rest of the tuples. While sharing patient data it is necessary to preserve them to some extent but in these two methods the datasets are considered to have only one sensitive attribute but in reality there, more than one and so the data slicing is the new technique introduced to overcome that disadvantage in existing methods.

3. SLICING

The privacy preserving techniques in the existing system can be easily infringed by various sources. Slicing is a privacy preserving technique that partitions the data both horizontally and vertically and is capable of handling high dimensional data. Overlapping slicing is done on the data that is to be stored in the database. The attributes are sliced both horizontally and vertically. In vertical portioning highly correlated attributes are grouped together and in horizontal portioning tuples are grouped together. This provides clear separation between the quasi and sensitive attributes. Overlapping slicing provides improved data utility. It provides privacy for person specific data. Highly correlated attributes are placed in one column. Tuples are grouped together to form buckets.

Nowadays collection of data is indeed a great task.Data along with the information needed exist in such a way that they a vast and varied. Since the data available and in abundance the must be privacy techniques available to preserve the data collected.These privacy techniques should be suited to most of the vast amount of data and should be applicable too. One of such techniques to apply privacy to data is privacy preserving publishing. However, the major concepts involved in preserving these data differ. Therefore, it is very difficult to preserve data. Some data contain the information, which added up to personal details.

When such data are needed to released then they may arise some problems. Micro data is one such kind of privacy preserved data where releasing of such information may create some problems to the individual and to the society too. Therefore, such type of data has to be maintained in privacy preservation techniques. Census data is one such collection of data, which has to be preserved properly. Medical data also holds information about the personal details of a person, which in collection elaborately called as medical report in regarding of the person. Therefore, these details have to maintain so carefully so that it can be preserved and maintained. The data collected should be accessible at a faster rate.

Once the data collected, organized, and then released many agencies and organizations come forward to release it. The collected information can be used for the research purposes and other public works such as collection of senses etc. Data sometimes may contain personal information regarding a person or some other additional information, which should be protected for maintaining the privacy. So these type of should not be public and should be preserved privately.Many techniques has be organized to maintain privacy of data. Such techniques in order to list are atomization techniques, which consist of the generalization and bucketisation for preserving the microdata. These techniques have some advantages and more of some disadvantages too. However, the disadvantages outline the advantages.let us discuss about the disadvantages that arise out of these techniques at first. When generalization method is employed, the

major factor is that the data are lost at a considerable amount. These data are particularly high dimensional data.

Therefore, steps must be taken in order to minimize the loss of data that occurs during the method of generalization. On the other hand, when the next concept called the bucketisation is taken into account it does not prevent the disclosures between the membership data. There are two kinds of attributes. They are correspondingly named as quasi attributes and sensitive attributes. Quasi attributes are entirely different from that of the sensitive attributes. Clear separation has to be organized between these two attributes. Data which are clearly distinguishable are termed as the correct data. In bucketisation concept the data are not accepted when they clearly do not distinguish between these two attributes. Then in order to overcome these difficulties a new concept called slicing comes into play. Slicing is found to be one of the efficient methods to distinguish between these data and to prevent the disclosure of data. To add more meaning to the concept of slicing and to brief it, it is the most effective concept as data utility is done it the most needed way it should be and so this is the most primary advantage that over rules the concept of generalization to enhance membership disclosure among data the concept of slicing has been carried out. Once when the data are being utilized threats may prone to occur. These threats have to maintained and minimized. Steps should be taken to minimize the threats.

When the membership disclosure is carried out the additional benefit is that the identity disclosure and attribute disclosure is also maintained as well as prevented. Sensitive information has to be maintained but when it is regarding or maintained for a particular individual it should be secretly maintained and when this information is taken into consideration it should be clear that whether it is the necessary information about that particular person or a thing.When the details is regarding the person it should be clear and safe. So these conditions are taken into account slicing is the best technique that furnishes these above mentioned facilities and satisfies the needful to the best.

Data mining is the extraction of information from the set of dataset that is available to us. Here privacy of the data is maintained and so this technique is called as the privacy preservation of data sets. Data mining concept has been vast deliberately expanded, one part to it is the part called as the privacy data mining. The information that is much protected, and those, which are personal, are to be the data that are to be protected. Collection of this type of information all together is called as the privacy protecting mining of data. Nowadays there are vast and quite a lot of techniques that are done to maintain the data. However, the most acknowledging in recent years are these techniques because individual's information are considered and are preserved. Security issues can also be overcome in slicing techniques. Characteristics of data have to be maintained and preserved.

Some data can fall into both the columns and so steps have to take to overcome this issues.Correlation is the name given to these overlapping of data. In slicing concept the above said disadvantages are relatively arranged in order and steps are taken to overcome it. Data utility is maintained in the concept of slicing. Along with data utility the additional factor that is necessary in the modern recent times are the data security.in the data security the data along with the vertical and horizontal information are secured and their privacy is maintained. As a detailed study in made in the field of data mining for the security and efficient utility of data there are various techniques and their best part and their drawbacks are also dealt with. On reading these Slicing is proved to be the most efficient everlasting technique which employees use a bit of all these concepts but yet stands significantly in the own technique of maintaining the data which may be public or may be private too. Hope a thorough knowledge has been gained about the concept of slicing and the maintenance of data. The era of slicing has emerged.

3.1 Algorithm For Accessing The Database

- Create a Class for establishing the database connection.
- Create a Database Username
- Create a Database Password
- If the Connection is failed, return the Database connection.

3.2 Algorithm For Creating A Patient Class For Handling The Database Operation

- Create a Patient class that handles all the Database operation of the table Patients.
- Get the data of the patients to be outsourced
- Return ResultSet .The ResultSet of the patients data that has to be outsourced
- Shuffle the sliced data.
- The rows are fetched as bucket with each bucket containing 5 rows.
- Buckets are shuffled among each other.
- return Returns the ArrayList of ArrayList of PatientTraitSlice.
- Get the bucket of rows with upper limit and lower limit
- Begin Lower limit of the bucket
- End Upper limit of the bucket
- Return ArrayList of PatientTraitsSlice.
- PatientTraitsSliceList List of PatientTraitsSlice
- Return List of PatientTraitsSlice is shuffled
- Slices the PatientLocation and puts them in the bucket.
- Slices data are shuffled within the bucket
- Return Bucket containing the PatientLocations
- The buckets containing the shuffled PatientLocationSlices are shuffled.
- Begin The upper limit
- End The lower limit
- Return Collection of shuffled buckets.
- The List of PatientLocationSlice are shuffled.
- PatientLocationSliceList is the list to be shuffled

3.3 Algorithm For Outsourcing Of Patients Table

- Create a class DB handling for patients_outsource table.
- Slice the rows, Shuffles the bucket of Patient Location, Shuffles the Patient Traits
- Insert the data into the new table which would be outsourced. This makes the data secured.
- Slice and shuffles the patient location

- Slice and shuffles the patient traits
- Insert the table to be out sourced

4. PARTIAL DATABASE SHUFFLING

In Partial database applications, a malicious database server can give sensitive information for the user queries all of them are sensitive in nature, patterns access points are observed to simply cover these, e.g., the continuous access to frequent records. The fundamental need for security is partial database outsourcing is privacy and a major drawback is disclosure from database. The PIR protocol makes the user to retrieve the tuples without extracting the whole information. It is the technique where the client learns about the attributes in the query added. PIR is formulated for cyrptographic technqiues to avoid information language. The techniques are proposed to flatten the complexity of data transfer between the user and the server. The PIR schemes introduce reducible time than that of giving the data to the user. The problem is that it is a solution to user's privacy but not to the server's privacy in different areas. However, the query in encrypted database is a challenge The problem is the performance of the system is getting reduced in order to improve the privacy. When data is stored in encrypted form, decryption has to be implemented later. To protect the pattern of the database, the research is implemented which only shuffles a portion of the database.

5. PERFORMANCE, CONCLUSION AND FUTURE ENHANCEMENTS

The performance graph for slicing and shuffling is shown above. The graph explicitly shows that slicing and shuffling the data shows a higher efficiency that the existing bucketization. Medical data also holds information about the personal details of a person, which in collection elaborately called as medical report in regarding of the person. Therefore, these details have to maintain so carefully so that it can be preserved and maintained. The data collected should be accessible at a faster rate. The faster the accessing of data the faster the information is gathered. Many recent works have been done to generalize these collected data. But where the studies insists that on generalizations of these data many information are lost and therefore alternate methods have to carried out in order to overcome the lost information through generalization of data.

For future work, we have provided heuristic approach in the first phase of the partition step. This involves comparison of the dataset anonymized using different set of requirements. In the second step, we measure the utility loss rather measuring the utility gain but when we implement the privacy function using the FFD we could end up in utility loss. So our main motto in our future enhancement is that is to eliminate the utility loss by implementing full privacy too.

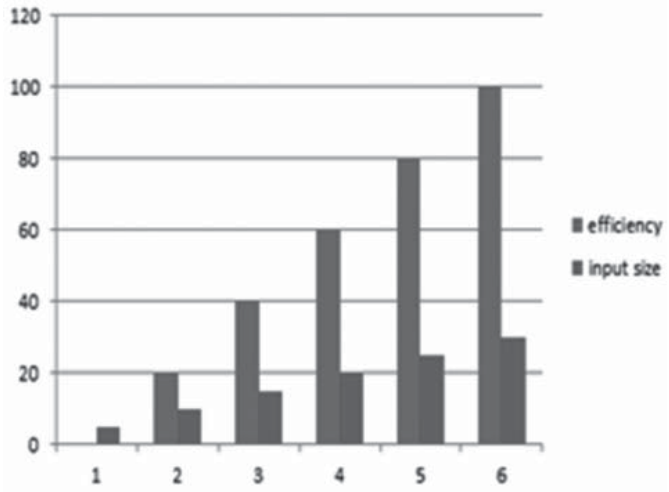

Figure 1. Performance Graph

This is mainly important because we need both the gain and the privacy, which could develop our data consistencies. From this we can say that we can construct an initial partition for bottom-up approach which has been leaded by frequency distributers. Likewise, we can also use ideas that can be applied to top – down approach too. Finally, we are in our development plan where we could enhance our privacy loss in worst-case scenario and measure the total utility loss that we will get from our system models. Thus, our main aim in this paper is to minimize the privacy loss for each individual system and to decrease the utility loss for all the pieces of useful and helpful knowledge.

Our research shows that the random grouping, which we have used in the previous model, is not useful and less effective. So we have dedicating ourselves to design a more efficient tuple grouping algorithms which we have already discussed. As a third research development, we are improving the slicing technique for handling the high dimensional data. By dividing the attributes into columns, were providing and protecting privacy by separating the uncorrelated attributes and is trying to preserve the connection between correlated attributes.

6. ACKNOWLEDGMENT

I thank my Supervisor Dr. K.Ramesh Kumar, Associate Professor, HITS,PADUR for his Guidance and support for the Research work.

References

1. Fudong Qiu; FanWu Guihai Chen“SLICER: A Slicing-Based K Anonymous Privacy Preserving Scheme” for Participatory Sensing 2013 IEEE 10th International Conference on Mobile Ad-Hoc and Sensor Systems.
2. Tiancheng Li; Ninghui Li; Jian Zhang; Ian Molloy “Slicing: A New Approach for Privacy Preserving Data Publishing”IEEE Transactions on Knowledge and Data Engineering.
3. C.Gokulnath; M. K. Priyan; E. Vishnu Balan; K. P. Rama Prabha; R. Jeyanthi“ Preservation of privacy in data mining by using PCA based perturbation technique” Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), 2015 International Conference.
4. Min Lu; Zhiguo Shi; Rongxing Lu; Ruixue Sun; Xuemin Sherman Shen “PPPA: A practical privacy-preserving aggregation scheme for smart grid communications”2013 IEEE/CIC International Conference on Communications in China (ICCC).
5. Mohamed M. E. A. Mahmoud; Sanaa Taha; Jelena Misic; Xuemin Shen“ Lightweight Privacy-Preserving and Secure Communication Protocol for Hybrid Ad Hoc Wireless Networks.”IEEE Transactions on Parallel and Distributed Systems.
6. Fudong Qiu; Fan Wu; Guihai Chen“Privacy and Quality Preserving Multimedia Data Aggregation for Participatory Sensing Systems”IEEE Transactions on Mobile Computing.
7. Khaled Alotaibi; V. J. Rayward-Smith; Wenjia Wang; Beatriz de la Iglesia“ Non-linear Dimensionality Reduction” for Privacy-Preserving Data Classification Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Conference on Social Computing (SocialCom).
8. Yuzhe Tang; Ling Liu; Arun Iyengar; Kisung Lee; Qi Zhang“e-PPI: Locator Service” in Information Networks with Personalized PrivacyPreservation Distributed Computing Systems (ICDCS), 2014 IEEE 34th International Conference.
9. Yilin Shen; Hongxia Jin“Privacy-Preserving Personalized Recommendation: An Instance-Based Approach via Differential Privacy”2014 IEEE International Conference on Data Mining.
10. M. Balamurugan; J. Bhuvana; S. ChenthurPandian“Shared and secured data partitioning for privacy preserving of collaborative file transfer in multi path computational mining 2012 Ninth International Conference on Wireless and Optical Communications Networks (WOCN).