



ISSN: 0975-766X  
CODEN: IJPTFI  
Research Article

Available Online through  
[www.ijptonline.com](http://www.ijptonline.com)

## ENHANCEMENTS OF SLICING TECHNIQUES USING IMBRICATING SLICING

Ishwarya M.V<sup>1\*</sup>, Dr.K.Rameshkumar<sup>2</sup>

Research Scholar, HITS, Sri Sai Ram Engineering College, West Tambaram, Tamil Nadu, India.

Assistant Professor, CSE Dept, Professor, IT Dept, HITS, Tamil Nadu, India.

Received on: 03-02-2017

Accepted on: 12-03-2017

### Abstract

**Objectives:** 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.

**Methods/Analysis:** 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. Unsanctioned disclosure is the major threat in data collection. So these can be overcome by the privacy preservation techniques. Nowadays there are vast and quite a lot of techniques that are done to maintain the data. **Findings:** 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.

**Keywords:** Slicing, Bucketisation, Generalisation, Imbricating Technique, Shuffling, Black Record, White Record.

### 1. Introduction

When particular information regarding a topic is collected, it is then processed until the knowledge about that data is gained. Knowledge extraction is the key concept under data mining. In data mining the input is known to us but the output at stage and at the final level are unknown and at times it remain hidden always. The gaining of output always have some hindrances and levels of stopping factors and here these factors are called as the scalability, data quality, complexity and many more. A good technique should overcome all these difficulties and produce the needful results. In slicing technique multiple correlations are maintained compared to the other attributes, which have their own features. Data are made into vertical and horizontal data and they the correlation between are data are preserved. Bucketisation, a

known technique where the data are resolved into buckets is the final technique which is done at the last[1]. Before using the anonymized data for workload task, it has to be processed. The anonymized table that is computed by the bucketisation. Slicing, overlapping slicing contains multiple columns that does not have any links between them. These data must be processed before the workload experiments are applied on it. The data are portioned into two or more column in bucketisation, slicing and overlapping slicing. We generate k tuples for a bucket that contains k tuples and 3 columns as follows: values in each column are randomly permuted. This procedure is applied to all the tuples from all the anonymized tables as in [10]. In random fashion this procedure generates link between two random columns. We apply this method five times and the average is reported.

## **2. Materials and Methods**

### **2.1 Imbricating Slicing**

The overlapping slicing is compared with the slicing and bucketisation based the data utility of the anonymized data. Mondrian algorithm is applied for all three techniques to compute the diverse table. 5, 8, and 10 are the values taken by 'l'. {place, occupation} are the sensitive column. For classifier learning we evaluate the quality of anonymized data. For decision tree and naïve bayes software package is applied to find the classifier accuracy. One attribute is chosen as the target attributes and all the other attributes are chosen as the predictor attribute. For performance evaluation we learn both the sensitive and QI attribute. To learn the sensitive attribute we build a classifier on sensitive attributes. Here, Occupation is the target attribute and all the other attributes are predictor attributes. By applying the classification process on the 3 anonymization technique we find that overlapping slicing is better. Weka tool is used to measure the accuracy of the database[2]. Accuracy of bucketisation, slicing and overlapping slicing is as follows. From the graph we came to know that the accuracy has been increased from the previous techniques and the newly proposed technique. The slicing and overlapping slicing may seem to have the same accuracy but the overlapping slicing uses the less number of attributes than slicing and produces high accuracy. The performance of overlapping slicing is better than all the other techniques, which means that slicing preserves correlations between the sensitive attributes and QI attributes. This may be due to that our processing randomly associates the sensitive and QI attributes values in each bucket. False associations may be introduced, but in generalization only the correct attributes are associated and the exact associations are hidden. When the value of 'l' increases the performance of slicing and bucketisation decreases much faster than the overlapping

slicing. This confirms that overlapping slicing contains better data utility. Nowadays collection of data is indeed a great task. Data along with the information needed exist in such a way that they are vast and varied. Since the data available and in abundance, there 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 [7]. 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 be 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. When the data are wide, generalization occurs and so therefore considerable amount of wastage of data occurs. Recent techniques also viewed that once generalizations are carried out information are lost. Bucketisation is a concept developed towards preventing of lost information but disclosure of data cannot be matched. These techniques have some disadvantages too. An alternative method has to be established to overcome the wastage of lost information. Data may be classified as high-dimensional, low-dimensional and medium-dimensional data. A new technique called the slicing of data or overlapping of slicing of data is a method proposed for high-dimensional data [1]. In normal, data are divided into rows and columns. Each row and column consists the needed information. Attributes are present in both rows and columns. Here attributes are partitioned into more than one column. Data present may be correlated or uncorrelated. In overlapping of slicing method, privacy is attained by breaking the privacy of uncorrelated attributes. Data utility is preserved between the correlated attributes. Here the attribute correlation is prevented. Therefore, the data is highly preserved in this technique when compared to generalization concept. There is also another concept called bucketisation where the disadvantages are overcome in overlapping of slicing method. On comparison between these three techniques overlapping of slicing, generalization and bucketisation the data utility is highly preserved in overlapping methods which is more effective too [6]. Here data utility is the main concept that is taken into concern and on making an analysis and comparative study the overlapping of slicing has more advantages and data utility and so will be used in data collection too.

**Table 1: Imbricating Slicing.**

AGE,GENDER	PINCODE,DISEASE
(22,M)	(14589, cancer) <b>WHITE RECORD</b>
(22,F)	(14586,flu ) <b>BLACK RECORD</b>
(33,M)	(14587,polycystic) <b>BLACK RECORD</b>
(45,M)	(14588, flu) <b>WHITE RECORD</b>
(56,M)	(14558, cancer) <b>BLACK RECORD</b>
(60,F)	(14577,bronchitis) <b>BLACK RECORD</b>
(60,M)	(14788,bronchitis) <b>WHITE RECORD</b>
(64,M)	(14566, flu) <b>BLACK RECORD</b>

The Table shown above is the Imbricating Slicing . From all of the inferences made it is observed that overlapping of slices is the method of handling and maintaining the high dimensional data. As discussed earlier, by partitioning the data attributes column and row wise overlap slicing reduces the dimensionality of data. The major advantage behind this technique is that each table can be viewed as the lower table, which can also be treated as the sub table. So the contents specified within that particular column along with the attributes can be viewed individually where the dimensionality of the data is reduced and the utility has been increased by comparative studies made so far have already a thorough knowledge about the concept of bucketisation and nowhere we include this concept as the additional methodology .Here the data within the table are categorized and arranged in the form so that they can be viewed as sub tables too. These sub tables can be elaborate content consisting of multiple and various data together as multiple independent tables. But here in the method called overlapping slicing concept is that the unique feature that is so called added is the linking those tables nor merely sub tables through their contents by the buckets[9]. Earlier techniques used various concepts for partitioning of the attributes called the QI attribute where based upon the results inferred from the tabular column the attributes are classified and another technique called the sensitive attribute also merely that concept too. One of the significance in this overlapping slicing technique is that these two techniques are completely avoided and the methods are neglected too. So overlap slicing works out independently without using the concepts of those that of earlier mentioned techniques.

In the bucketisation method, the data are separated into multiple buckets, which individually hold upon the data. The buckets, and their number can be varied. Tuples are the keyword, which we have already come across in various concepts related to our topic, and here too these tuple hold the data contents. Another amazing surprise is that these tuples can hold the data together or any number of buckets together. Here the bucketization concept of holding of data in

the buckets in further enhanced that these data are connected together by the tuples. Overlapping slicing concept includes multiple properties, where one of the unique property is that tuple has ability to match to the buckets. These buckets hold many data together. A tuple hold the multiple buckets together. Here each table can hold multiple Data and attributes together. The important keyword which adds up to the concept are data publishing which are arranged in the form of rows and columns and then deliberately visible to the users. The next keyword is that of the technique known as bucketisation which we are already familiar with. Another concept called as the generalization where the data are arranged in generalize order .another key concept is the micro data. Atomization techniques are also one of the key concepts that are involved in our data concept. In addition, other techniques that are involved add up their property to it[2]. Once the data collected, organized, and then released many agencies and organizations come forward to release it. This 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 micro data. 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.

### **3. Future Enhancements**

In future, this work can be extended where a design of more tuple grouping algorithm can be developed using overlapping shuffling scheme. Apart from this, there are a lot of anonymization techniques which can be used but the problem of how to use those techniques continue to be unknown. Therefore, in this research, it is clear that CCP is the most effective technique to preserve the privacy. So further research may include studying CCP more in detail and determine what are the ways the hackers may exploit this hotspot.

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[3]. Then in order to overcome these difficulties a new concept called slicing comes into play..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.

Overlapping slicing is an important method in protection the information about an individual without revealing much of the data about them when it comes to a publication or submission. This is generally called as data animalization, where It is a technique which is used to remove the identifiers form the database. This removal is mainly important while identifying the details because the identifier will give you the key to attain the required information . We have already seen three methods called generalization, bucketisation and slicing. This involve partition of data both horizontally and vertically, this can be even explained in three forms i.e. Attribute partitioning, tuple partitioning and column generalization. This technique is mainly useful for total portioning too which permute the sensitive attribute values randomly. But this has main disadvantage, because while we are using slicing we will generally face data utility .So the proposed model of ours will help you in better privacy threats. So we recommend overlap slicing than to slicing which could lead to loss in data utility. This involves a better attribute called chi value which could give you the information about highly correlated attributes as in [9]. This value even lead us to chi matrix which is used to find the correlated attributes in the form of columns. These are mainly used to deliver us out ultimate goal which is data privatization. But in this method we have more efficient way of disclosure of attributes and membership attributes disclosure

#### **4. Conclusion**

The technique is mainly useful for total portioning too which permuted the sensitive attribute values randomly. However, this has main disadvantage, because while we are using slicing we will generally face data utility .So the proposed model of ours will help you in better privacy threats. So we recommend overlap slicing than to slicing which could lead to loss in data utility. This involves a better attribute called chi\_value, which could give you the information about highly correlated attributes[10]. This value even leads us to chi matrix, which is used to find the correlated attributes in the form of columns. These are mainly used to deliver us out ultimate goal which is data privatization. But in this method we have more efficient way of disclosure of attributes and membership attributes disclosure. However, the

overall methodology is that before anonymizing the data, person or a record has to analyses the data characteristics in data anonymization. This could rationalize the design to get a better data. We also include the experimental results in this conclusion, which is being retrieved from the database. In this the preprocessing steps must be applied on the table before the experiments on workload is done on the data. After computation, the pre-processor data the sensitive attribute and quasi identifiers are examined. After this step of identifying the pre-processor the modified technologies such as Mondrian diversity functions are employed for the overlapped sliced table. In order to measure the performance level of overlapping slicing technique against several privacy threats such as identity, membership and attribute disclosure the accuracy of the method can be identified/measured.

## **5. Acknowledgment**

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

## **6. References**

1. Xuhua Ding, Yanjiang Yang, and Robert H. Deng ,” Database Access Pattern Protection Without Full-Shuffles”, IEEE Transactions On Information Forensics And Security, Vol. 6, No. 1, March 2011.
2. Avinash Kumar Singh, Narayan P Keer, Anand Motwani, ”A Review Of Privacy Preservation Technique”, NRI Institute of Research And Technology ,Bhopal, India. International Journal Of Computer Applications (0075-8887), Volume 90-No.3, March 2014.
3. Amar Paul Singh , Mr. Dhanshri Parihar “A Review of Privacy Preserving Data Publishing Technique“, School of CSE Bahra University Shimla Hills, India. Research Article June 2013 Amar Paul Singh Page 32 International Journal of Emerging Research in Management & Technology ISSN: 2278-9359 (Volume-2, Issue-6)
4. J. Xu, W. Wang, J. Pei, X. Wang, B. Shi, and A.W.-C. Fu, “Utility- Based Anonymization Using Local Recoding,” Proc. 12th ACM SIGKDD Int’l Conf. Knowledge Discovery and Data Mining (KDD), pp. 785-790, 2006
5. Ke Wang (Simon Fraser University), Philip S. Yu (IBM T. J. Watson Research Center), Sourav Chakraborty (Simon Fraser University), ” Bottom-Up Generalization: A Data Mining Solution to Privacy Protection “.

6. D.J. Martin, D. Kifer, A. Machanavajjhala, J. Gehrke, and J.Y. Halpern, “Worst-Case Background Knowledge for Privacy- Preserving Data Publishing,” Proc. IEEE 23rd Int’l Conf. Data Eng. (ICDE), pp. 126-135, 2007.
7. Y. He and J. Naughton, “Anonymization of Set-Valued Data via Top-Down, Local Generalization”, Proc. Int Conf. Very Large DataBases (VLDB), pp-934-939, 2009.
8. Priti V. Bhagat and Rohit Singhal, “An Overview of sectional Shuffle for Database Access Pattern Protection Using Reverse Encryption Algorithm”, 1M.Tech student, Department of Computer Science & Engineering, I.E.T. Alwar, Rajasthan, India, 2Asth. Professor, Department of Computer Science & Engineering, I.E.T. Alwar, Rajasthan, India. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 2, No 5, May 2013.
9. X. Xiao and Y. Tao, “Anatomy: Simple and Effective Privacy Preservation,” Proc. Int’l Conf. Very Large Data Bases (VLDB), pp. 139-150, 2006.
10. D. Aruna Kumari \*, Y. Vineela , T. Mohan Krishna , B. Sai Kumar , “Analyzing and Performing Privacy Preserving Data Mining on Medical Databases”, Indian Journal Of Science and Technology, Volume9, Issue17, May2016.