



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 4, April 2017

Privacy Preserving Microdata Publishing using Slicing Technique

Sangeeta Rani¹, Deepika Goyal²

M. Tech, Dept. of CSE, Advanced Institute of Technology and Management, Palwal, India¹

Assistant Professor, Dept. of CSE, Advanced Institute of Technology and Management, Palwal, India²

ABSTRACT: In the era of digital data/information, storing data and mining that data for further analysis become a common and important task. Because every industry wants to grow by doing analysis of its previous data. For this purpose, data mining is the basic step. But we also need to secure the information before performing any operation on that info for e.g. medical records are very important to get secured. SO, we have proposed an innovative approach i.e. slicing to boost the present state of the art. Whereas slicing partition the information into horizontal and vertical groups. Which reduces the spatiality of the info and preserves higher utility that generalization and bucketization.

KEYWORDS: Data Mining, Bucketization, Privacy, Knowledge Discovery

I. INTRODUCTION

The collections of digital data by governments, firms and people have created tremendous opportunities for knowledge-based higher cognitive process. Driven by mutual edges or by laws that need bound knowledge to be revealed there is a requirement for exchange and publication of knowledge among varied parties. The limitation of that approach is that it either distorts knowledge overly or needs a trust level that is impractically high in several data-sharing situations. For instance, contacts and agreements cannot guarantee that sensitive knowledge would not be carelessly misplaced and find you within the wrong hands. Privacy suggests that however a private control Worlds Health Organization has access to his personal data [7]. From another purpose of read, privacy could also be however the info is collected, shared and utilized by the shoppers. Thus, definition of privacy varies from one atmosphere to the opposite. Some of the definitions are: Privacy because the right of an individual to work out that personal info regarding him/her is also communicated to others [13]. Privacy as restricted access to an individual and to any or all the options associated with the person.

A task of the utmost importance is to develop strategies and tools for small information business information in an exceedingly additional hostile atmosphere, in order that the printed information remains much helpful whereas individual privacy is preserved. The enterprise is named privacy-preserving information business (PPDP) [1]. Within the past few years, analysis communities have versed this challenge and planned several approaches.

As data processing matures as a field and develops a lot of powerful algorithms for locating and exploiting patterns in knowledge, the number of knowledge regarding people that is collected and hold on continues to quickly increase. The increase in knowledge heightens issue that data processing violates individual privacy [2]. The goal of knowledge mining to derive mixture conclusions, that mustn't reveal sensitive data. However, the data-mining algorithms run on databases containing data regarding people which can be sensitive. The goal of privacy-preserving data processing is to supply high quality mixture conclusion whereas protect the privacy of the constituent people [10]. Many anonymization techniques like generalization and bucketization are designed for privacy conserving micro data publication.

Many government departments and corporations use advance data processing techniques to realize insights. into the behaviours and characteristics of their voters and customers. On the opposite hand, many polls associated surveys indicate that the general public has an augmented sense of privacy intrusion as a result of the augmented level of security when the terrorist attack terrorist attacks. Since data processing is usually a key part of the many Office of

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 4, April 2017

Home and Security systems, observance and police investigation systems, and enterprise info systems, the general public has non-heritable a negative impression that data processing may be a technique for privacy intrusion. The dearth of trust in data processing has become an obstacle to the advancement of the technology. For instance, a probably helpful data processing scientific research known as Terrorist Act into Awareness (TIA) was terminated by the U.S. Congress principally as a result of its polemical varieties of aggregation, tracking and analysing knowledge trails left by people.

II. RELATED WORK

In [18], Tiancheng Li, Ninghui Li, Senior Member, IEEE, Jian Zhang, Member, IEEE, and Ian Molloy have introduced Slicing: a replacement Approach for Privacy conserving knowledge business enterprise, 2012 IEEE. They planned a replacement approach slicing that partitions the information horizontally and vertically. We have tendency to show that slicing preserves higher knowledge utility that generalization and may be used for membership speech act protection. Their employment experiments ensure that slicing preserves higher utility that generalization and are simpler than bucketization in workloads involving the sensitive attribute. Privacy-preserving business enterprise of microdata has been studied extensively in recent years. Microdata contain records regarding a personal entity like someone, a home or a corporation.

In [16], Alexandre Evfimievski and Tyrone Grandison have introduced privacy-preserving data processing (PPDM) to the worlds of data mining that seeks to safeguard sensitive information from uninvited or unofficial disclosure. Most ancient data processing techniques analyse and model the information set statistically in aggregation whereas privacy preservation is primarily involved with protective against revelation individual knowledge records. This domain separation points to the technical feasibility of PPDM.

In [17], Smita D Patel, Sanjay Tiwari (2013) have introduced data system should persuade one amongst the foremost necessary properties as privacy. For this basis, many efforts are dedicated to incorporating privacy conserving techniques with data processing algorithms so as to stop the revelation of sensitive info throughout the information finding

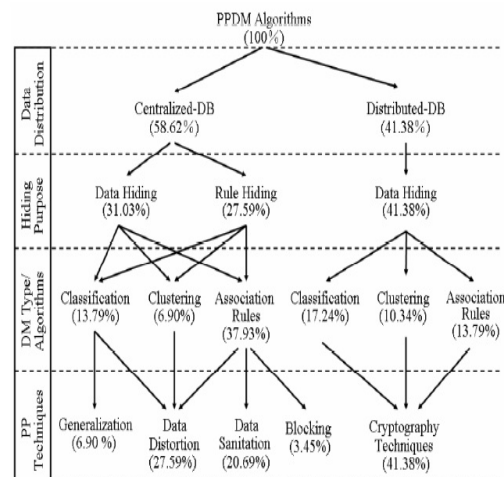


Figure 2.1 node.

III. PROPOSED ALGORITHM

Data mining is that the exploration and analysis of enormous quantities of knowledge to get valid, probably method that plays a significant role in extraction of helpful info. Now days' large database exists in numerous application i.e. medical information, census information, communication and media-related information, shopper purchase info and information gathered by government agencies. Thus, information sharing is required for full utilization of

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

collected information as a result of pooling of medical information will improve the standard of medical analysis additionally the info gathered by the govt. ought to be created.

We introduce a unique information anonymization technique referred to as slicing to boost the present state of the art. Slicing partitions the info set each vertically and horizontally. Vertically partitioning is completed by grouping attributed into columns supported the correlations among the attributes. Horizontal partitioning is completed by buckets. Basic plan of slicing is to interrupt the association cross columns, however to preserve the association inside every column. This reduces the spatiality of the info and preserves higher utility that generalization and bucketization [12]. Slicing protects privacy as a result of it breaks the association between unrelated attributes that square measure occasional and therefore distinguishing. eq. (3)

IV. METHODOLOGY

The planned system develops a technique that involves slicing for information privacy preservation in business of dataset for any statistics or survey.

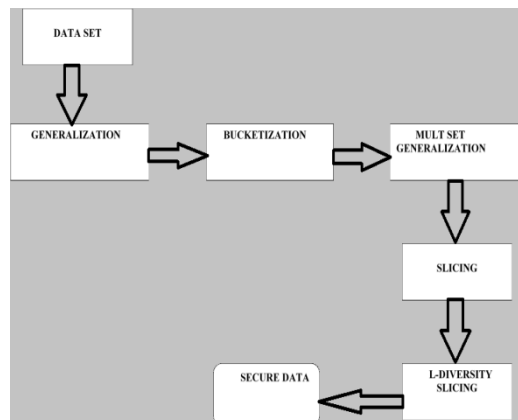


Figure 4.1 System Architecture

Additionally, we tend to add protection against identity revealing within the sliced information by applying partial generalization i.e. go of the quasi-identifiers. The subsequent area unit the list of steps or operations concerned within the implementation of our work;

1. Information processing
2. Anonymity
3. Multi set based generalization
4. One-attribute-per-column slicing
5. Slicing method

Information Processing: - It is associated degree usually neglected however necessary step within the method as process. This is basically parsing the data to extract useful information from large amount of unstructured information.

K-anonymity: - The k-anonymity model is finished so as to stop the re-identification of people within the free information set. However, is does not take account into abstract thought relationship

From the quasi-identifier to some sensitive attribute [4,5, 14, and 15].

Multi set based generalization: - We tend to bring home the bacon this by showing that slicing is healthier than the native coding approach delineated as follows

Instead of employing a generalized price to interchange additional specific attribute values, one uses the multiset of actual values in every bucket.

One attribute per column slicing: -Our works observe that while one-attribute-per-column slicing preserves attribute distributional information, it does not preserve attribute correlation, because each attribute is in its own column.

Slicing Method: - Slicing partitions the data set both vertically and horizontally. Vertical partitioning is done by grouping attributes into columns based on the correlations among the attributes [3,6].



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

V. SIMULATION RESULTS

In this planned work, presents a brand-new approach known as slicing to privacy-preserving microdata distributing. It overcomes the restriction of generalization and bucketization and preserves higher utility whereas protective to privacy threats. We have tendency to illustrate a way to apply slicing to stop attribute revealing and membership revealing. Our experiments show that slicing preserves higher information utility than generalization and are simpler than bucketization.

The overall methodology planned by this work is that before anonymizing the info, one will investigate the information characteristics and use this individually in data anonymization. The premise is that one will style higher information anonymization techniques once we grasp that the info is healthier. The objective of my planned work is to judge a comparative study of varied techniques used for maintaining privacy in data processing and additionally evaluates those techniques has highest accuracy in data processing and sensible privacy preservation capability. Information sharing has become a part of the routine activity of the many people, companies, organizations, and government agencies. Privacy protective knowledge business enterprise could be a promising approach to data sharing whereas protective individual privacy and protective sensitive data. During this survey, we have a tendency to review the recent developments within the field. The final objective is to remodel the first knowledge into some anonymous type to stop from inferring its record owners' sensitive data. We have a tendency to best of our views on The distinction between privacy preserving knowledge business enterprise and privacy-preserving data processing, and gave an inventory of fascinating properties of a privacy-preserving knowledge business enterprise technique.

VI. CONCLUSION AND FUTURE WORK

Where variety of anonymization techniques is designed, it remains associate open downside on the way to use the anonymized information. In our experiments, we tend to at random generate the associations between column values of a bucket. This could lose information utility. Another direction is to style data processing tasks exploitation anonymized information computed by varied anonymization techniques.

REFERENCES

1. A. Inan, M. Kantarcioglu, and E. Bertino, "Using Anonymized Data for Classification," Proc. IEEE 25th Int'l Conf. Data Eng. (ICDE), pp. 429-440, 2009.
2. A. K. Ilavarasi, S. Poorani, "A Survey on Privacy Preserving Data Mining Techniques", International Journal of Computer Science and Business Informatics (IJCSBI), vol. 7, no. 1. November 2013.
3. Amar Paul Singh, "An Efficient Sliced Data Algorithm Design for Personalized Data Protection to Prevent Generalized Losses", International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE), Volume 4, Issue 3, March 2014.
4. A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam, "ℓ²-Diversity: Privacy Beyond k-Anonymity," Proc. Int'l Conf. Data Eng. (ICDE), p. 24, 2006.
5. Aristides Gionis and Tamir Tassa, "k-Anonymization with Minimal Loss of Information", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 21, NO. 2, FEBRUARY 2009.
6. Alphonsa Vedangi V. anandam, "Data slicing technique to privacy preserving and data publishing", International Journal of Research in Engineering and Technology (IJRET), Volume: 02 Issue 10, Oct-2013.
7. Benjamin c. m. fung, Kewang, Rui chen, Philip s. yu, "Privacy-preserving data publishing: a survey of recent developments", ACM Computing Surveys, Vol. 42, No. 4, Article 14, Publication date: June 2010.
8. 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.
9. D. Kifer and J. Gehrke, "Injecting Utility into Anonymized Data Sets," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD), pp. 217-228, 2006.
10. D. Mohanapriya , Dr. T. Meyyappan, "Slicing Technique For Privacy Preserving Data Publishing", International Journal of Computer Trends and Technology (IJCTT), volume 4 Issue 5, May 2013.
11. G. Ghinita, Y. Tao, and P. Kalnis, "On the Anonymization of "Sparse High Dimensional Data," Proc. IEEE 24th Int'l Conf. Data Eng. (ICDE), pp. 715-724, 2008.
12. G. Kesavaraj, S. Sukumaran Ph.D, Bucketization based Flow Classification Algorithm for Data Stream Privacy Mining International Journal of Computer Applications (0975 – 8887) Volume 81 – No.12, November 2013.
13. J. Brickell and V. Shmatikov, "The Cost of Privacy: Destruction of Data-Mining Utility in Anonymized Data Publishing," Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD), pp. 70-78, 2008.
14. K. LeFevre, D. DeWitt, and R. Ramakrishnan, "Mondrian Multidimensional k-Anonymity," Proc. Int'l Conf. Data Eng. (ICDE), p. 25, 2006.



ISSN(Online): 2320-9801
ISSN(Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

15. K. LeFevre, D. DeWitt, and R. Ramakrishnan, "Incognito: Efficient Full-Domain k-Anonymity," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD), pp. 49-60, 2005.
16. Privacy-Preserving Data Mining Alexandre Evfimievski IBM, Almaden Research Center, USA Tyrone Grandison, IBM Almaden Research Center, USA, 2009, IGI Global.
17. Smita D Patel, Sanjay Tiwari, Privacy Preserving Data Mining, Smita D Patel, International Journal of Computer Science and Information Technologies(IJCSIT), Vol. 4 (1) , 2013,pp. 139 – 141.
18. Tiancheng Li, Ninghui Li, Jian Zhang, and Ian Molloy, "Slicing: A New Approach for Privacy Preserving Data Publishing", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 24, NO. 3 MARCH 2012.