Indian Journal of Science and Technology, *Vol 10(24), DOI: 10.17485/ijst/2017/v10i24/114039, June 2017*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Privacy Preserving Data using Fuzzy Hybrid Data Transformation Technique

K. Abrar Ahmed[1]* and H. Abdul Rauf[2]

[1]Department of Computer Science and Engineering, Manonmaniam Sundaranar University, Chennai – 600017, Tamil Nadu, India; kaa406@yahoo.co.in
[2]Sree Sastha Institute of Engineering and Technology, Chennai – 600113, Tamil Nadu, India; harauf@yahoo.com

Abstract

Objective: To provide improved privacy for dataset using fuzzy set properties with optimal Information loss, thereby achieving better utility. **Methods:** We propose Fuzzy Hybrid data Transformation method by combining fuzzy data modification method and Random Rotation Perturbation techniques (RRP). Fuzzy data modification method contains fuzzy K-member clustering along with membership function to be executed to distorted data. RRP preserves the geometric structure on dataset. **Findings:** Experimentation proves that our method gives least information loss when compared with existing method along with fuzzy membership function individually for different values of k. Thereby, we achieve dual goal of privacy and utility. **Applications:** The experiment has been done over the Adult dataset derived from UCL Repository and used for numerous applications such as analysis, mining, forecasting and prediction etc.

Keywords: Fuzzy Data Modification, PPDM, Privacy Threats, RRP

1. Introduction

In recent years, there was rapid increase in accessing data about individuals for research purpose. These data on the other hand contain sensitive information which leads to more concerns about privacy. Today many researcher uses privacy preserving data mining (PPDM) algorithm to attain privacy of individual's data[1,2]. The main idea of PPDM is to develop a method that should change or modify original data in such a way that no one can identify the sensitive information about individuals[3,4]. Many techniques of PPDM exist and some of them are K-anonymity, Perturbation, Bucketization etc. But these methods have side effect in that sense it reveals some sensitive information about individuals.

To overcome this drawback, we use the properties of fuzzy sets whose idea is to make the attribute values to distort by performing partition using fuzzy membership function[5,6]. The main idea of Fuzzy membership function is to derive the pattern of data that is present. The value of each QI attribute is modified on a range of 0-1 by mapping them to membership value.

In this paper we propose Fuzzy Hybrid data Transformation method, which is a combination of Fuzzy data modification (FDM) and Random rotation perturbation method (RRP). FDM contains Fuzzy K-member clustering along with membership function to be executed to distorted data. RRP preserves the geometric structure on dataset. In this paper we experimentally prove that hybrid method gives optimal information loss for different values of k along with different membership function.

The remainder of this paper is organized as follows: Section-II is the related work. Section-III explains proposed methodologies. Section-IV describes about experiments and results. Section-V summarizes the conclusion of the paper.

2. Background

In[7] found the limitation of existing data mining tools and also discuss about the problem associated with privacy preserving data mining techniques. Authors propose a privacy framework which not only improves perfor-

*Author for correspondence

mance but also preserves privacy of sensitive information and also gives distorted dataset used for various analysis purposes.

In[8] takes important issues occurring in preserving privacy of data for clustering. Author pointed out that existing privacy preserving techniques only concentrate on data property. Authors propose hybrid method (HDTTR and HDTSR) which gives solution to the problem of addressing privacy of confidential categorical data in clustering. Also they made a complete analysis and show effectiveness of clustering of sensitive categorical data before and after transformation.

In[9] describes about the problem of K-Anonymity such as attribute disclosure, Identity disclosure etc. The method increases computational complexity to achieve privacy. The authors propose a technique using fuzzy set approach to achieve maximum overhead. This technique can be useful for both numerical and categorical attributes.

In[10] states that it is difficult to resolve conflict between privacy of data and correct mining result at once. Authors proposed a method called Random Response which is a combination of Random Response Technology. The main idea of this algorithm is to solve conflict between privacy and mining result. And the authors experimentally proved the same.

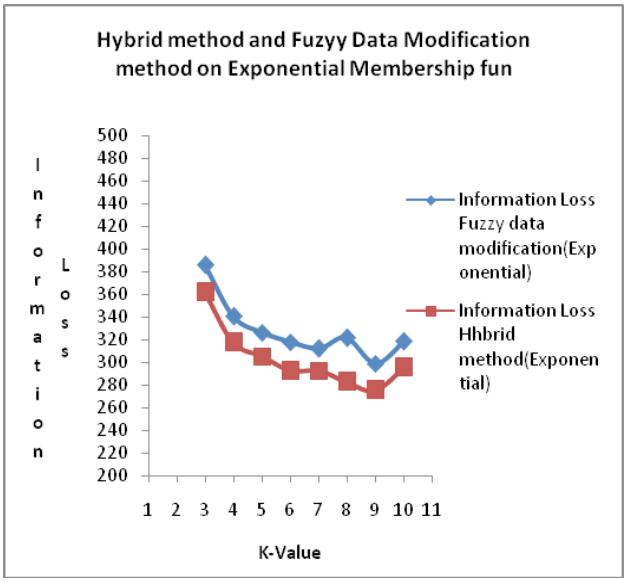

Figure 1. Comparison of Hybrid method with Fuzzy data modification method on exponential membership Function.

In[11] finds some important issues from an existing data mining techniques such as balancing privacy and accuracy. Authors propose a method which is built upon the application of Double-Reflecting Data Perturbation (DRDP) and Rotation based Translation (RBT) in order to provide secrecy of data confidential numerical attribute without losing accuracy.

In[12] studied four clustering algorithms in order to observe the performance of clustering algorithm in detecting outlier while performing clustering. The objective of this research is that clustering algorithm is used in much application such as fraud detection, network intrusion detection and clinical diagnosis. So Algorithm should be designed in such a way that it should detect outlier efficiently if it exists.

The survey discusses above suggested that different techniques have been proposed for achieving privacy on data for different purpose such as analysis, mining, forecasting and prediction etc. But no one use concept of fuzzy set properties. This is the first time we use fuzzy properties combined with secure computation technique to achieve privacy and utility.

3. Proposed Methodology

This section gives a detail description of our hybrid method used to anonymize original dataset and Information loss metric used to calculate data loss. Fuzzy Hybrid data Transformation method is a combination of fuzzy data modification with various membership function and Random Rotation Perturbation method (RRP) to Anonymized original data.

3.1 Fuzzy Data Modification Method (FDM)

FDM is used to Anonymize (distorts) the original data to its relevant data using built in fuzzy membership namely exponential membership function, bell shaped membership function, Triangular membership function[13–15]. Fuzzy data modification consists of fuzzy K-member clustering to be executed along with built in fuzzy membership such as exponential, bell shaped, Triangular Member function to anonymized data. The significant steps in fuzzy data modification method are shown.

Step 1 Load the dataset S which contain collection of QI attribute.

Step1.1 Initialize number of equivalence class 'EC'.

Step 1.2 Select the fuzzy membership function and initialize the membership boundary values.

Step 2 for each value of QI attribute, i.e., QI attribute>0. Repeat the process:

2.1 Modify the data of each QI attribute by mapping them to a corresponding membership value. And remove it from S.

2.2 t=t+1. Generate cluster G_t for the corresponding values of QI attribute.

Step 3 Repeat the following process till $|G_t|>k$.

Identify the best cluster G_t that suits the QI attribute value.

Include QI attribute value to cluster G_t. Remove from S.

3.3 If there is no remaining QI attribute stop the process.

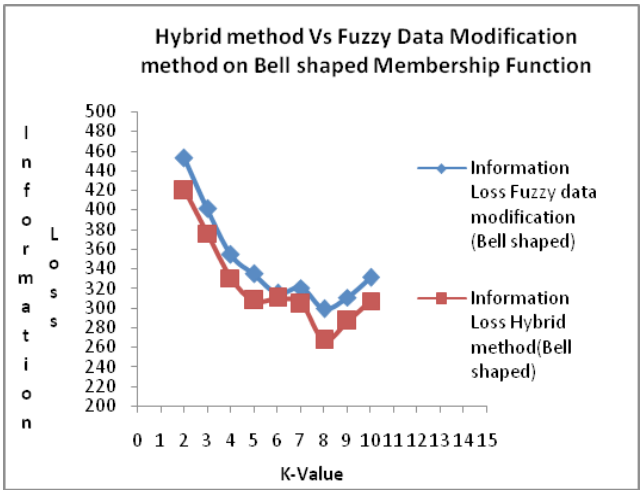

Figure 2. Comparison of hybrid method with fuzzy data modification method on bell shaped membership function.

3.2 Membership Function

Membership function is a curvature that shows how each point in the input space is mapped to a membership value between 0 and 1[16]. Input space is sometimes referred to as the universe of disclosure. I is represented as *uf(x)* , where f is fuzzy set and x is defined as set of ordered pairs.

3.2.1 Bell-Shaped Membership Function

[17] The Bell-Shaped membership function can be obtained from the following equation:

$$u_{tr}=\exp(-d_{tr}^2/\sigma) * u_r \quad (1)$$

Where d_{tr} is the distance between the core object of cluster 't' and its neighbour object 'r'.u_r is the residual membership of 'r' at the current iteration.

3.2.2 Exponential Membership Function

[17]The Exponential member function can be obtained from the following equation:

$$u_{tr}= \exp(-d_{tr}/\sigma) * u_r \quad (2)$$

Where d_{tr} is the distance between the core object of cluster 't' and its neighbor object 'r'.u_r is the residual membership of 'r' at the current iteration.

3.2.3 Triangular Membership Function

[17]The triangular membership function can be obtained using the relation:

$$u_{tr}= (\max \{ -1/\alpha\, d_{tr}+ 1,0 \}) * u_r \quad (3)$$

Where d_{tr} is the distance between the core object of cluster 't' and its neighbor object 'r'.u_r is the residual membership of 'r' at the current iteration.

3.3 Fuzzy Hybrid Data Transformation Method

Fuzzy Hybrid data Transformation method is combination of two techniques namely fuzzy data modification and Random rotation perturbation method (RRP).In this method, original dataset is distorted using fuzzy data modification method and will be given as input to RRP method to obtain final distorted data.

3.3.1 Random Rotation Perturbation

The idea of RRP is to preserve the geometric properties of the dataset. Let us consider a dataset D consists of 'n' records with 'm' attributes called as n*m matrix M, contain numeric values in all attributes. We can represent the matrix M as 'n' point in an 'm' dimensional space. We can then generate an **m*m** rotation matrix **R** and multiply matrix M with matrix R to obtain perturbed matrix P, which is defined as

$$P=M*R. \quad (4)$$

To generate Rotation matrix R, the following generalized matrix form has to be followed by selecting the angle randomly from 0 to 360.

For Two Dimensional matrix i.e. m=2

$$R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

For Three Dimensional matrix i.e. m=3

$$R = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

This perturbed matrix is use to preserved original dataset as it look quite different from it. A Rotation matrix R is a matrix which has to be satisfies the following property[18].

$R * R^T = R^T * R = I$

Here R^T denotes the transpose of R and I is the identity Matrix. This property implies that both the rows and columns

of the matrix are orthonormal, that is, for any row i,

$$\sum_{i=1}^{m} r_{il}^2 = 1 \quad (5)$$

and for any two different rows i; j,

$$\sum_{l=1}^{m} r_{il} * r_{jl} = 0 \quad (6)$$

Furthermore, for any column i,

$$\sum_{i=1}^{m} r_{il}^2 = 1 \quad (7)$$

and for any two different columns i; j,

$$\sum_{l=1}^{m} r_{il} * r_{jl} = 0 \quad (9)$$

The Significant step of Fuzzy Hybrid data Transformation Method is shown below

Step:1 Load the dataset D with 'n' records and 'm' attributes.

Step:2 Generate matrix 'M' of size n*m

Step:3 Generate matrix 'R' of size m*m form the generalize from give above by selecting an angle randomly.

Step:4 Multiply matrix M with Rotation matrix R to generate perturbed matrix P and save it in a file.

Step:5 Give this file as input to fuzzy data modification method.

Step 6: For each fuzzy membership function

Step 7: For each QI attribute do

Step 8: Modify QI attributes values using fuzzy membership function.

Step 9: End For

Step 10: Save the distorted data and calculate Information Loss for different values of K.

3.4 Information Loss [(Tarique Ahmed, Haque, Thauhid, 2014)[19]

The information loss in cluster Gc is measured as follows [19]:

$$ILc = |G_c| \left(\sum_i [max_{ci} - min_{ci} / size_{ci} + \sum_j \frac{[H_{ci}^U)}{H(T_{cj})} \right] \quad (5)$$

[1] [19] where $size_{ci}$ is the size of numeric domain of numeric attribute i, and max_{ci} and min_{ci} are the maximum and minimum values in Gc. T_{cj} is the taxonomy tree defined for the domain of categorical attribute cj and H(T) is the height of taxonomy tree T. ($^{\uparrow}U_{\downarrow}ci$) measures the deviation in Gc. If H(T) is 1, the second term is reduced to the number of different categories.

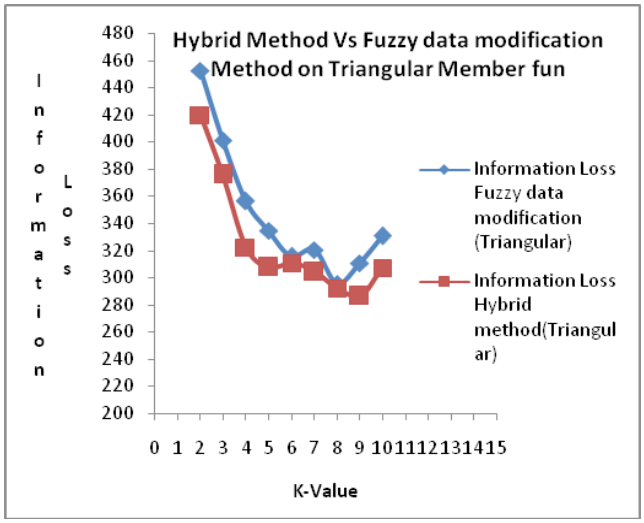

Figure 3. Comparison of hybrid method with fuzzy data modification method on triangular membership function.

4. Experimental Results

In this section the performance of proposed method is evaluated in terms of information loss. The experiment is conducted using JDK7 in Eclipse tool and MatLab. Fuzzy Hybrid data Transformation method is a combination of two techniques namely Fuzzy Data Modification and Random Rotation Perturbation (RRP). It was implemented in JDK 7.0 IN Eclipse Tools and Information loss is calculated in MatLab script. Both methods are experimented using Adult Dataset obtained from UCI Repository[20]. Adult dataset consists of 14 attributes with 10,000 records. From 14 attributes, 6 attributes are QI attributes and rest of them is identifier and sensitive attributes.

In our first experiment, a Fuzzy data modification is conducted to convert original dataset to k-anonymity dataset using three fuzzy membership function namely exponential membership function, Bell-Shaped mem-

bership function and Triangular membership function. Initially, Exponential membership function is used with Fuzzy data modification and for different k-value and information loss has been calculated. Similarly for Bell-Shaped and Triangular membership function, information loss has been calculated for different k-values.

In second experiment, Fuzzy Hybrid data Transformation method using same three membership function have been conducted for different k-value and information loss is obtained accordingly.

Figures1, 2 and 3, it is clear that, for different value of k, hybrid methods give optimal information loss when compared to fuzzy modification method. For k=2 in hybrid method using Exponential membership function, information loss is 403.67. Whereas in fuzzy data modification method using Exponential membership function, Information loss is 435.276. Similarly Fuzzy Hybrid data Transformation method using Bell-Shaped membership function for k=2, information loss is 419.82.Whereas in fuzzy data modification method using Bell-Shaped membership function, information loss is 452.69. From these results it is proved that information loss obtained for different k-values in hybrid method is optimal than fuzzy data modification. So it is experimentally proved that Fuzzy Hybrid data Transformation method have higher utility and good privacy than fuzzy data modification method.

5. Conclusion and Future Work

In this paper, we analyze Fuzzy Hybrid data Transformation method using different membership functions in terms of information loss with fuzzy data modification method (FDM). From the obtained result we observe that hybrid method gives led information loss for different k values along different membership functions .So thereby hybrid method achieves good privacy and utility. But the hybrid method work only for numerical dataset. The future enhancement is to search some other techniques like Random Rotation Perturbation (RRP) to be combined with FDM to distort numerical and text data.

6. References

1. Ahmed KA, Rauf HA, Rajesh A. Study of K-Anonymization Level with Respect to Information Loss AJBAS. 2016; 10(2):1–8.
2. Aggarwal G, Feder T, Kenthapadi K, Motwani R, Panigrahy R, Thomas D, Zhu A. Approximation Algorithm for K-Anonymity. Journal of Privacy Technology. 2005; 4(2):11–8.
3. Sweeney L, Anonymity K. A Model for Protecting Privacy. International Journal of Uncertainity Fuzziness and Knowledge-Based System. 2002; 10:557–70. Crossref
4. Samarati P. Protecting Respondent's Identities in Microdata Release. IEEE Transaction on Knowledge and Data Engineering. 2001; 13(6):1010–27. Crossref
5. Ichilhashi H. A Fuzzy Variant of K-member clustering for Collaborative Filtering with Data Anonymization. Proceedings of International conference on IEEE Word Congress on Computational Intelligence, 2012. p. 1–6.
6. Kawano A, Honda K, Notsu A, Ichihashi H. Comparison of membership function in K- member clustering data Anonymization. Proceeding of IEEE International Conference on SCIS-ISIS, 2012. p. 2004–8. Crossref
7. Kamakshi P, Babu AV. A Novel Framework to improve the quality of Additive Perturbation Technique. International Journal of Computer Application. 2011; 30(6):11–6.
8. Rajalaxmi RR, Natarajan AM. An Effective Data Transformation Approach for Privacy Preserving Clustering. Journal of Computer Science. 2008; 4(4):10–7. Crossref
9. Kumari VV, Rao SS. Fuzzy based approach for privacy preserving publication of data. IJCSNS. 2008; 8(4):1–7.
10. Liu J, Yifeng. Privacy preserving clustering by Random Response method of Geometric Transformation. Procedings of 4th International Conference on Internet Computing for Science and Engineering, 2009. p. 181–8.
11. Li L, Zhang Q. A Privacy preserving clustering Techniques using Hybrid data Transformation Method. Proceedings of IEEE International conference, 2009. p. 1502–6. Crossref
12. Poonam, Dutta M. Performance analysis of clustering method for outlier Detection. Proceedings of International conference on Advance Computing and Communication Technology, 2012. p. 89–95. Crossref
13. Karthikeyan B, Manikandan G, Vaithiyanathan V. A Fuzzy Based Approach for Privacy Preserving Clustering. Journal of Theoretical and Applied Information Technology. 2011; 32(2):118–22.
14. Kumar P, Verma KI, Sureka A. Fuzzy Based Clustering Algorithm for Privacy Preserving Data Mining. International Journal Business Information System. 2011; 7(1):27–40. Crossref
15. Syed MD, Ahmed T, Haque S, Tauhid FSM. A Fuzzy Based Approach for Privacy Preserving Clustering. International Journal of Scientific and Engineering Research. 2014; 5(2):1067–71.
16. Manikandan G, Sairam N, Harish V, Saikumar N. Survey on the Use of Fuzzy Membership Function to Ensure Data Privacy. Journal of Pharmaceutical Biological and Chemical Sciences. 2016; 7:344–8.

17. Kamakshi P, Babu AV. A Novel Framework to improve the quality of Additive Perturbation Technique. International Journal of Computer Application. 2011; 30(6):1–6.
18. Lin Z, Wang J, Liu L, Zhang J. Generalized Random Rotation Perturbation for vertically pertitioned Data sets. Proceedings of IEEE International conference, 2009; 4(7):1–4.
19. Syed MD, Ahmed T, Haque S, Thauhid SMF. A Fuzzy based approach for Privacy Preserving clustering. International Journal of Scientific and Engineering. 2014 Feb; 5(2):1–5.
20. UCI Data Repository. Available from http://archive.ics.uci.edu/ml/datasets/Adult Datasets. Accessed on 01/05/1996.