ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Survey on the use of Fuzzy Membership Functions to Ensure Data Privacy.

Manikandan G*, Sairam N, Harish V, and Nooka Saikumar.

ICT Department, School of Computing, SASTRA University, Tamil Nadu, India.

ABSTRACT

The digital era has empowered us to store huge volume of data in various storage devices and also allows for access to extensive data collections. Data mining has appeared as a noteworthy skill for ascertaining facts from these data gathering. A notable limitation in data mining is that it reveals some private information during the mining process. Consequently, privacy-preserving data mining has obtained a raising awareness in the recent past. Fuzzy logic can be used to preserve privacy by distorting the original data. In this paper, we present a detailed analysis on the use of various fuzzy membership functions in achieving privacy. Experiments were performed using the adult data set in the UCI machine repository.

Key words: Clustering, Data Utility, Data Privacy, Misclassification error, Data Perturbation.

Corresponding author

ISSN: 0975-8585

INTRODUCTION

With the advancement in Information technology many organizations have started storing information about their customers for efficient decision making. There is a need to collaborate with other organizations for obtaining more precise results. For making accurate predictions organizations rely on various data mining tools. Usually data mining algorithms extract unknown interesting patterns from the data warehouse. Apart from being so helpful, some susceptible information is also exposed by mining algorithm. In order to ensure privacy the extracted data needs to be modified before releasing it to the public and researchers. Many privacy preserving techniques are available in the literature that ensures privacy as well as accuracy. This paper presents a detailed survey on the use of various fuzzy membership functions for achieving Privacy. Each membership function is evaluated in two dimensions namely data utility and data privacy. The merits and demerits of each membership function are tabulated in the experimental results section.

Fuzzy Membership Functions:

The concepts of fuzzy logic can be implemented for ensuring privacy conservation in data mining. The fuzzy logic can be implemented for the whole group of data. The appraisal of various kinds of data can be achieved by applying various Membership Functions.

The data from a dataset can be modified on a scale of 0-1, by mapping them to a membership value. The curve which defines this methodology is said to be a membership function. Fuzzy membership functions are used to extract the pattern of the data that is present. The next section elaborates the role of various membership functions.

Triangle Membership Function:

The Triangular Membership Function can be achieved using the relation

$$\text{TriMF}(D: X, Y, Z) = \begin{cases} 0 & \text{when } D < X \\ (D - X) / (Y - D) & \text{when } X \le D \le Y \\ (Z - D) / (Z - Y) & \text{when } Y \le D \le Z \\ 0 & \text{when } D > Z \end{cases}$$

In the above relation TriMF means Triangle Membership Function. D is the value in the Dataset. X, Y, Z are the three boundary points and their values are 20, 50 and 80 respectively.

Table 1: Output of Triangular Membership Function.

S.No	1	2	3	4	5	6	7	8	9	10
Original Data	39	50	38	53	28	37	49	52	31	42
Modified Data	0.633	1.0	0.6	0.9	0.266	0.566	0.966	0.933	0.366	0.733

Trapezoidal Membership Function:

The Trapezoidal Membership function can be achieved using the relation

$$\text{TrapMF}(D: W, X, Y, Z) = \begin{cases} 0 & \text{when } D < W \\ (D - W) / (X - W) & \text{when } W \le D < X \\ 1 & \text{when } X \le D < Y \\ (Z - D) / (Z - Y) & \text{when } Y \le D < Z \\ 0 & \text{when } D \ge Z \end{cases}$$

In the above relation TrapMF means Trapezoidal Membership Function. D is the value in the Dataset. W, X, Y, Z are the four boundary points and their values are 10, 20, 60 and 100 respectively.

ISSN: 0975-8585

Table 2: Output of Trapezoidal Membership Function.

S.No	1	2	3	4	5	6	7	8	9	10
Original Data	39	50	38	53	28	37	49	52	31	42
Modified Data	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

BellShaped Membership Function:

The BellShaped Membership function can be achieved using the relation

$$\text{BellMF}(D: X, Y, Z) = \frac{1}{1 + |(D - Z) / X|^{2Y}}$$

In the above relation, BellMF means BellShaped Membership Function. D is the value in the Dataset. X, Y, Z are the three boundary points and their values are 17, 90 and 38 respectively.

Table 3: Output of BellShaped Membership Function.

S.No	1	2	3	4	5	6	7	8	9	10
Original Data	39	50	38	53	28	37	49	52	31	42
ModifiedData	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Gaussian Membership Function:

The Gaussian Membership function can be done using the relation

$$\text{GaussMF}(D: C, W) = \text{EXP}\{-(D - C)^2 / W^2\}$$

In the above relation, GaussMF means Gaussian Membership Function. D is the value in the Dataset. C is the center and W is the width of the function and their values are 50 and 40 respectively.

Table 4: Output of Gaussian Membership Function.

S.No	1	2	3	4	5	6	7	8	9	10
Original Data	39	50	38	53	28	37	49	52	31	42
Modified Data	0.927	1.0	0.913	0.994	0.738	0.899	0.999	0.997	0.798	0.960

S- Shaped Membership Function:

The S-Shaped Membership function can be achieved using the relation

$$\text{SMF}(D: X, Y) = \begin{cases} 0 & D \le X \\ 2 * [(D-X)/(Y-X)]^2 & X \le D \le (X+Y)/2 \\ 1 - 2 * [(D-Y)/(Y-X)]^2 & (X+Y)/2 \le D \le Y \\ 1 & D \ge Y \end{cases}$$

In the above relation, SMF means S-Shaped Membership Function. D is the value in Dataset. X is the minimum and Y is the maximum value in the Dataset and their values are 17 and 90 respectively.

Table 5: Output of S-Shaped Membership Function.

S.No	1	2	3	4	5	6	7	8	9	10
Original Data	39	50	38	53	28	37	49	52	31	42
Modified Data	0.181	0.408	0.165	0.486	0.045	0.15	0.384	0.459	0.073	0.234

ISSN: 0975-8585

EXPERIMENTAL STUDY

For Experimental study, we have utilized the Adult Dataset, which is available at UCI Repository. The Dataset is comprised of 32561 records. Each record comprises of 14 attribute values as age, work class, fnlwgt, education-num, education, marital-status, occupation, relationship, race, sex, capital-gain and capital-loss, hours-per-week and native country. Here we make use of Age attribute for implementing Fuzzy Membership Functions. Here, all the tables consist of first 10 elements for comparison purpose. Table 1 and Table 2 represent the output for first 10 values in the dataset after applying Triangular and Trapezoidal membership functions respectively. Table 3 and Table 4 shows the output for Bellman and Gaussian membership functions and finally in Table 5, modified data obtained using S-Shaped membership function is shown. The graph generated for the original data is shown in Fig.1 and Fig.2 shows a comparison graph for the above membership functions.

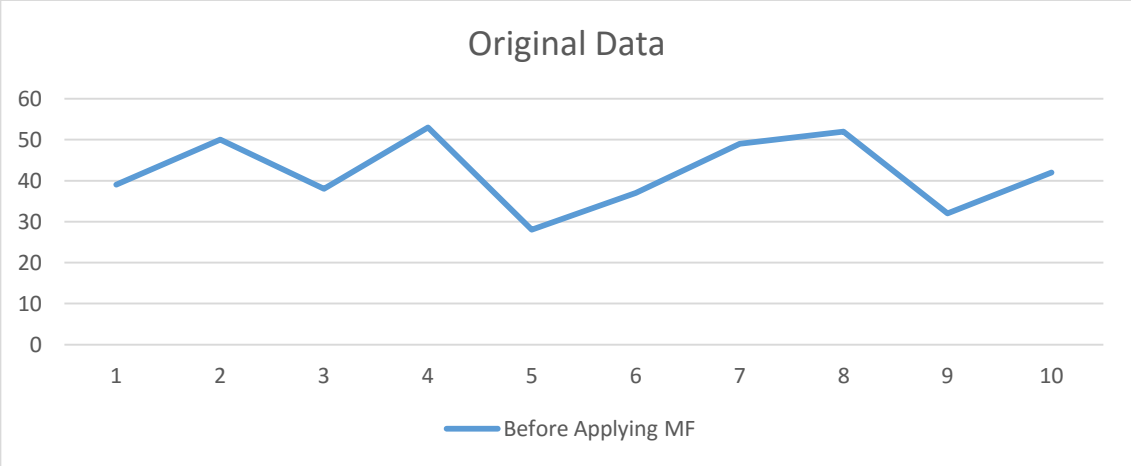
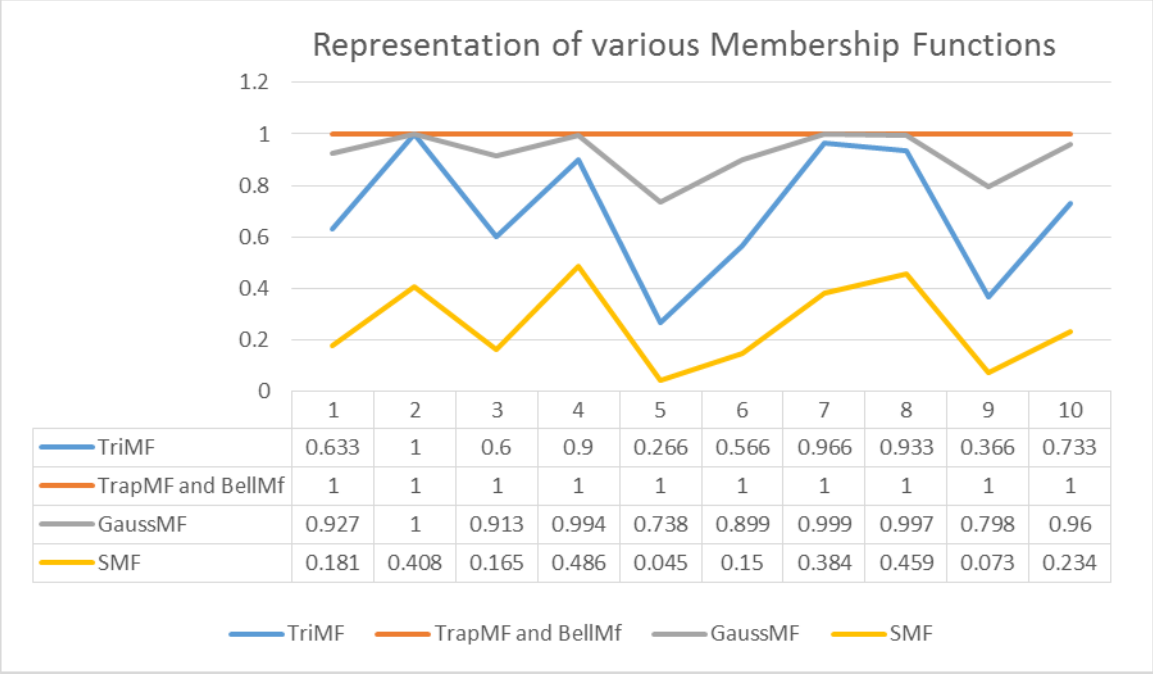


Figure 1: Graph representing original data



	1	2	3	4	5	6	7	8	9	10
TriMF	0.633	1	0.6	0.9	0.266	0.566	0.966	0.933	0.366	0.733
TrapMF and BellMf	1	1	1	1	1	1	1	1	1	1
GaussMF	0.927	1	0.913	0.994	0.738	0.899	0.999	0.997	0.798	0.96
SMF	0.181	0.408	0.165	0.486	0.045	0.15	0.384	0.459	0.073	0.234

Figure 2: Comparison Graph representing various Membership Functions.

ISSN: 0975-8585

CONCLUSION

In this paper we have analyzed the use of various fuzzy membership functions for ensuring data privacy in data mining. From the various plots and observed values, it makes clear that the S-Shaped Membership Function is an efficient procedure that can be used to obtain the sanitized data which makes an illusion as the original data to the users.

REFERENCES

[1] Jiawei Han and Micheline Kamber, Data Mining: Concepts and Techniques, Morgan Kauffman Publishers, 2006.

[2] Benjamin C.M. Fung, Ke Wang, Ada Wai-Chee Fu; Philip S. Yu, Introduction to Privacy-Preserving Data Publishing: Concepts and Techniques, Chapman and Hall, 2010

[3] B.Karthikeyan,G.Manikandan,Dr.V.Vaithiyanathan, "A Fuzzy Based Approach for Privacy Preserving Clustering", Journal of Theoretical and applied information Technology , Vol 32(2), 118-122,2011.

[4] G.Manikandan, N.Sairam, S.Sharmili, S.Venkatakrishnan, "Achieving Privacy in Data Mining Using Normalization", Indian Journal of Science and Technology, Vol 6(4) , 2013,4268-4272.

[5] G.Manikandan, N.Sairam, S.Jayashree, C.Saranya, "Achieving Data Privacy in a Distributed Environment Using Geometrical Transformation", Middle East Journal Of Scientific Research, Vol 14(1) , 2013,107-111.

[6] N.Abitha,G.Sarada,G.Manikandan,N.Sairam, " A Cryptographic Approach for Achieving Privacy in Data Mining", In Proceedings of 3rd IEEE International Conference on Circuit, Power and Computing Technologies (ICCPCT–2015), 2015. DOI:10.1109/ICCPCT.2015.7159300

[7] G.Manikandan, N.Sairam, N.Abitha, G.Sarada, R.Pranesh, M.Vigneshwaran, "A Random Noise Based Perturbation Approach For Achieving Privacy In Data Mining", Global Journal of Pure and Applied Mathematics, Vol 11(3), 2015, 1635-1639.

[8] G.Manikandan,N.Sairam,M.Sathiya Priya,Sri Radha Madhuri , "A General Critical Review on Privacy Preserving Data Mining Techniques", Global Journal of Pure and Applied Mathematics, Vol 11(4) , 2015,1899-1906.

[9] UCI Data Repository http://archive.ics.uci.edu/ml/datasets.html

[10] Jiang, J. and Umano, M. 2014, Privacy preserving extraction of fuzzy rules from distributed data with different attributes, Joint 7th International Conference on and Advanced Intelligent Systems (ISIS), 15th International Symposium on Soft Computing and Intelligent Systems (SCIS), 2014, pp : 1180-1185

[11] Torra, V. ; Miyamoto, S. ; Endo, Y. ; Domingo-Ferrer, 2008, On intuitionistic fuzzy clustering for its application to privacy, J. FUZZ-IEEE (IEEE World Congress on Computational Intelligence). IEEE International Conference on Fuzzy System DOI : 10.1109/FUZZY.2010.5584186, pp 1042 –1048

[12] Kasugai, H. ; Kawano, A. ; Honda, K. ; Notsu, A., 2013, A study on applicability of fuzzy k-member clustering to privacy preserving pattern recognition, IEEE International Conference on Fuzzy Systems (FUZZ), 2013, pp:1-6

[13] Honda, K. ; Kawano, A. ; Notsu, A. ; Ichihashi, H., 2012, A fuzzy variant of k-member clustering for collaborative filtering with data anonymization, Fuzzy Systems (FUZZ-IEEE), 2012 IEEE International Conference on, pp: 1-6

[14] Tanaka, D.; Oda, T.;Honda, K.;Notsu, A., 2014, Privacy preserving fuzzy co-clustering with distributed cooccurrence matrices Joint 7th International Conference on Soft Computing and Intelligent Systems (SCIS), 2014 and 15th International Symposium on Advanced Intelligent Systems (ISIS), pp: 700-705