FUZZY-BASED RANDOM PERTURBATION FOR REAL WORLD BANK DATASETS

Authors: Alok Singh Yadav I.Vikramadithya** Zeshan Haris MP****

Author Addresses :
*School of Computing Science and Engineering, VIT University, Vellore-632014, India
**School of Computing Science and Engineering, VIT University, Vellore-632014, India
***School of Computing Science and Engineering, VIT University, Vellore-632014, India
Emails: mpzeeshan@gmail.com vikramivaturi@gmail.com @gmail.com

Abstract: *Mining previously unknown patterns from enormous size of data is the main objective of any data mining algorithm. In current days there is a remarkable expansion in data collection due to the development in the field of information technology. The patterns revealed by data mining algorithm can be used in various domains like Image Analysis, Marketing, weather forecasting etc. As a side effect of the mining algorithm some sensitive data is also exposed. There is a need to preserve the privacy of individuals which can be achieved by using privacy preserving data mining. In this project, fuzzy based data transformation methods are proposed for privacy preservation in database environment. A fuzzy data transformation method is proposed using the Triangular fuzzy membership function to transform the original dataset.*

Keywords: *Privacy Preservation; Fuzzy Membership Function; Random Rotation Perturbation; Data Transformation*

Introduction:

Data mining is the process used to analyse large quantities of data and gather useful information from them. It extracts the hidden information from large heterogeneous databases in many different dimensions and finally summarizes it into categories and relations of data. In order to learn a system in detailed manner, we should be able to decrease the system complexity and increase our understanding about the system.

Privacy preservation is the major concern while real time datasets are handled. Privacy preserving data mining (PPDM), deals with data modification and also limits information loss. Data perturbation is one of the PPDM techniques which deals with numerical data and focuses on maintaining statistical properties of data. There are two types of perturbation, additive perturbation and multiplicative perturbation, where generated random value is either added or multiplied to the data, which results in a modified data.

In PPDM, fuzzy logic is one of the methods used for preservation of data. Fuzzy logic is a type of many-valued logic that deals with degree of truth of a statement or an argument, this truth value can be any real number in the range of 0 and 1, where 0 being that the argument is totally false, and 1 being argument is completely true.

The idea of using fuzzy logic is applied to preserve the individual information while revealing the details in public. This project mainly focuses on converting the sensitive data into modified data by using Triangular fuzzy membership function.

The term fuzzy logic was coined by Lotfi Zadeh in his explanations of fuzzy set theory. Fuzzy logic has since been adopted in many fields like control theory and artificial intelligence. Fuzzy models and sets are based on the way of human decision making which is generally imprecise and non-specific. These models have the capability to represent and utilize information that is vague and arrive at a definite conclusion.

In this project, a fuzzy methodology is considered for selection of customers whom the bank should target for deposit subscription. Fuzzy set theory aims at modelling imprecise and ambiguous information. Computers cannot still effectively handle such scenarios as general machine intelligence systems employ sequential (Boolean) logic. The human brain is exceptional and superior as it has the capacity to handle fuzzy statements and decisions. It works by filtering and selecting data that is relevant and has purpose.

So we propose a model in which the perturbation is done by randomization, where the data is generated in intervals based on the level of privacy generated from a fuzzy system based on various inputs.

Personal data is any information relating to you, whether it relates to your private, professional, or public life. In the online environment, where vast amounts of personal data are shared and transferred around the globe instantaneously, it is increasingly difficult for people to maintain control of their personal information. This is where data protection comes in. The primary goal of privacy preserving is to hide the sensitive data. The raw data contains some sensitive information of individuals, which should not be exposed to protect individual privacy.

System Model:

The participants considered for this banking system are,

- Bank Manager
- Assistant Manager
- Campaign Executive

- Customer Service Manager
- Account Operating Executive

Bank Manager is responsible for promoting and marketing the bank and its products, ensuring uncompromised levels of customer service. Keeping the staff fully trained and motivated.

Assistant Manager maintains and balances vaults, taxes, fica. Further maintains quarterly branch efficiency. Keeps workforce informed of pertinent modifications in operational coverage and strategies.

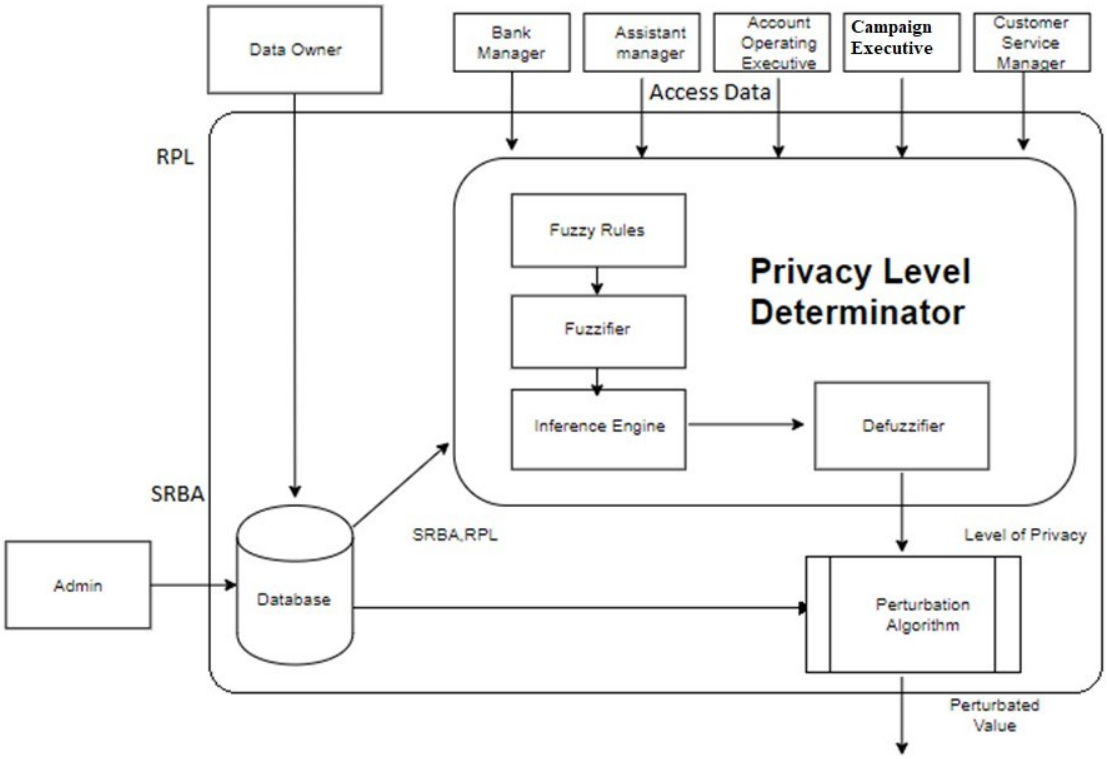
Campaign Executive Keep abreast of modern developments in marketing Work within a price range and record any overspend Brief and oversee the paintings of internal and outside agencies, which includes designers, animators and printers, who will deliver tons of the campaign activity Manage the implementation, tracking and size of advertising campaigns

Customer Service Manager officials (also known as customer service representatives or specialists) deal with all factors of the customer-enterprise relationship. Usually, they work by themselves at desks, speaking with clients either on the telephone or through the Internet. They often cope with calls and emails from the consumer to remedy troubles or direct them to a branch or individual that focuses on their problem.

Account Operating Executive entertains both existing and capacity customers, answering questions on various banking merchandise and explaining the sorts of debts the financial institution offers. Should a brand new customer make a decision to enrol in the financial institution, the financial institution account operating executive can be the one to installation new bills, solution any questions with the account, and clear up possible problems related to person money owed.

Now consider a scenario where one of the parties mentioned above is trying to access information about the customers of the bank, it’s not necessary for all the parties to access all the data except of which is of work for them. This is when SRBA (specific role based access) comes into effect. SRBA is set according to each user of the system, for instance if the SRBA has been set to degree for any party that is in the range (0-10) or any other range as the system requires, so in this case less noise will be added for the respective party than any other party whose SRBA value is comparatively low.

Another attribute that should be considered is RPL (required privacy level) this again is set as a numeric value in a certain range, so if the RPL is high much noise is to be added to the data. SRBA and RPL are together are both to be considered in deciding the LOP (level of privacy) of a particular user.



Working:

The dataset is preloaded into the local database. The software into which the user has to login, has the SRBA value that is used to indicate if the party has to be given the access or not and also RPL value that is required privacy level are predefined by the owner. These values are passed into the membership function [figure 1.0] of the fuzzification module. In this function, the attribute values are mapped to a value between 0 and 1 which is the degree of membership, where the value '1' represents full membership and the value '0' represents no membership. The membership function is a graphical representation of the level of participation of each value that is given in the input. The crisp output values are then interpreted into a linguistic value of membership. These linguistic values of membership are then sent into the inference engine with the predefined fuzzy rules and the processed to get LOP (Level of privacy) in linguistic format. The LOP is then converted into crisp value in the de-fuzzification module. This value of LOP is passed into the function that gives a random value as output. This random value is then passed into the perturbation function that multiplies the required data with value.

Data description:

This dataset is publicly available for research. citation of the database: [Moro et al., 2014] S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing.The binary classification goal is to predict if the client will subscribe a bank term deposit (variable y).

Input variables:

bank client data:

1 - age

2 - job : type of job

3 - marital : marital status

4 - education

5 - default: has credit in default?

6 - housing: has housing loan?

7 - loan: has personal loan?

related with the last contact of the current campaign:

8 - contact: contact communication type

9 - month: last contact month of year

10 - day_of_week: last contact day of the week

11 - duration: last contact duration, in seconds

12 - campaign: number of contacts performed during this campaign and for this client

13 - pdays: number of days that passed by after the client was last contacted from a previous campaign

14 - previous: number of contacts performed before this campaign and for this client

15 - poutcome: outcome of the previous marketing campaign

16 - emp.var.rate: employment variation rate - quarterly indicator

17 - cons.price.idx: consumer price index - monthly indicator

18 - cons.conf.idx: consumer confidence index - monthly indicator

19 - euribor3m: euribor 3 month rate - daily indicator

20 - nr.employed: number of employees - quarterly indicator

output variable:

21 - y - has the client subscribed a term deposit?

Membership functions:

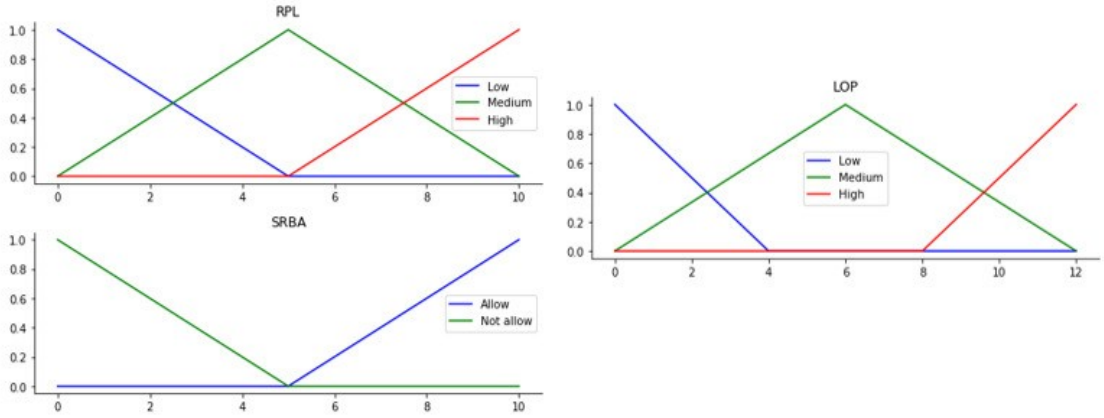


Figure 1.0 – Membership functions of (a) RPL (b) SRBA (c) LOP

Fuzzy rule base

If the SRBA is not allow OR the RPL is low, then the lop will be low

If the RPL is medium, then the lop will be medium

If the SRBA is allow OR the RPL is high, then the lop will be high

Discussion:

The test is done on a customer and her marital status is married and she doesn't have any loans, then it is very likely that the customer to go for fixed deposit scheme. If the customer is a person around 50 years of age, approximate saving around 1 lakh per year, married, no previous loan, never contacted earlier by marketing person, these inputs represent the degree of uncertainty and doubt in the information furnished during various time periods. The degree of uncertainty and doubt in the information and the level of judgment used by the marketing person in deciding to approach this customer for subscribing the deposit scheme is always a challenge. In this type of situation fuzzy based expert system is a very good tool for decision making with respect to both customers and marketing managers to invest on directed campaigns with a strict and rigorous selection of contacts.

Conclusion:

The use of a fuzzy based expert system for targeting specific customer is taken into consideration by the author. This report describes design of fuzzy decision support system for identification of prospective customer in situations of data diversity and imprecision, which can be used by specialized banking experts for improving their marketing campaign. Our future efforts will be to use association rules that can produce an optimum surface representing all the combination points from a few of the tested combinations. Also, further we would like to work on other areas for applying fuzzy logic.

References:

1. A. Geetha Mary, D. P. Acharjya, N. Ch. S. N. Iyengar. “Privacy Preservation in Fuzzy Association Rules Using Rough Computing and DSR”.” Int. J. Autonomous and Adaptive Communications Systems, Vol. 10, No. 1, 2017”.
2. Mary A. Geetha. “Fuzzy-based random perturbation for real world medical datasets”.” Int. J. Telemedicine and Clinical Practices, Vol. 1, No. 2, 2015”.
3. Ali Bazmara, Soheila, Sardar Donighi “Bank Customer Credit Scoring by Using Fuzzy Expert System”.” I.J. Intelligent Systems and Applications, 2014, 11, 29-35”.
4. Ahmed A. Mohamed, Ahmed S. Salama “A Fuzzy Logic based Model for Predicting Commercial Banks Financial Failure”. “International Journal of Computer Applications (0975 – 8887) Volume 79 – No 11, October 2013”.

5. M.Sridhar, B. Raveendra Babu, PhD “A Fuzzy Approach for Privacy Preserving in Data Mining”.” International Journal of Computer Applications (0975 – 8887) Volume 57– No.18, November 2012”.
6. Shrikant Zade, Dr. Pradeep Chouksey, Dr.R.S. Thakur “Fuzzy Based Approach for Privacy Preserving in Data Mining”.” IJCTT Journal Volume-45 Number-1 Year of Publication: 2017”.
7. Omar Adil M. Ali, Aous Y. Ali, Balasem Salem Sumait “Comparison between the Effects of Different Types of Membership Functions on Fuzzy Logic Controller Performance”.” International Journal of Emerging Engineering Research and Technology Volume 3, Issue 3, March 2015, PP 76-83”.
8. Derrig, R. and Ostaszewski, K., Fuzzy techniques of pattern recognition in risk and claim classification. In, The Journal of Risk and Insurance. Volume 62(3), (1995) 447-82.
9. Hüllermeier, E.: Fuzzy methods in machine learning and data mining: Status and prospects. Fuzzy Sets and Systems. 156, 387-406 (2005).
10. W. Pedrycz, “Fuzzy set technology in knowledge discovery,” Fuzzy Sets Syst., vol. 98, pp. 279–290, 1998.

11. Deshmukh, A. and Lakshminarayana, T.: A rule-based fuzzy reasoning system for assessing the risk of management fraud. In International Journal of Intelligent Systems in Accounting, Finance & Management. Volume 4 (1998) 231-41.