**PRIVACY PRESERVATION USING BANK DATA PERTURBATION**

***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*

**LITERATURE SURVEY:**

Fuzzy Based Decision Making for Promotional Marketing Campaigns

The use of a fuzzy based expert system for targeting specific customer is taken into consideration by Nikhat Khan and Dr Fozia Khan. This paper 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.

A Study on Data Perturbation Techniques in Privacy Preservation Data Mining

In this paper written by Nimpal Patel, Shreya Patel, they have discussed about different types of data perturbation and popular technique for privacy preserving data mining. The major challenge of data perturbation is balancing privacy protection and data quality, which normally considered as a pair of contradictive factors. Geometric data perturbation technique is a combination of Rotation, Translation and Noise addition perturbation technique. It is especially useful for data owners to publish data while preserving privacy –sensitive information. Typical examples include publishing micro data for research purpose or outsourcing the data to the third party that provides data mining services. In this paper they try to explore the latest trends in Geometric data perturbation technique.

Fuzzy-Based Loan Evaluation System

In this paper written by Sadig Mammadlia, a fuzzy logic approach to evaluating retail loans that can be used to describe imprecise knowledge or human subjective judgment by linguistic terms is proposed. A fuzzy model has been created which is based on the information inputs mostly used by Azerbaijani banks when evaluating a retail loan. The fuzzy information inputs are the loan applicant’s income level, credit history, character, collateral and employment with linguistic terms such as “low”, “medium”, “high” and etc. The model’s knowledge base consists of rule block of “IF. Then...” rules. The output of the constructed fuzzy model is credit standing which is also a fuzzy variable with linguistic terms.

Fuzzy-Based Random Perturbation for Real World Medical Datasets

Mary A. Geetha has mentioned from experimental analysis that we could understand the accuracy of data is independent of the size of the dataset and also the number of attributes being perturbed. This paper also cited that the method that was specifically implemented need not have its values to be reconstructed like other general methods as the accuracy values are maintained more or less similar to the measures of loss due to perturbation, classification results.

Bank Customer Credit Scoring by Using Fuzzy Expert System

In this paper by Ali Bazmara and Soheila Sardar Donighi shows that the fuzzy system could get good outcome for solving the bank customer classification problem for several datasets. There have been remarkable improvements in the efficiency as of extracting the rules and using the data. Furthermore, the dependence on expert was removed. So, the proportion of error that the expert has on this model had no effect on creating it. Due to the nature of fuzzy rules, they are describable and experts could confirm or reject the validity of its decision. This method has another advantage that it could adapt by altering the rules, because the rule is applied on the data and rule is extracted based on data. So, this expert system is not designed only for a bank or particular database, and the rules re not completely rigid.

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.

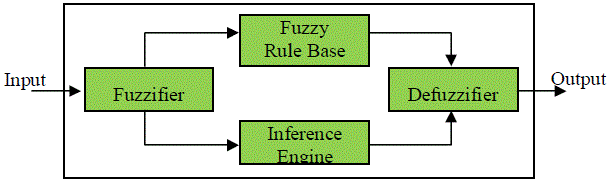
In PPDM, fuzzy logic is one of the methods used for preservation of data. Fuzzy logic is a type of many-valued logic that deals with degree of truth of a statement or an argument, this truth value can be any real number in the range of 0 and 1, where 0 being that the argument is totally false, and 1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

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.

**PROPOSED METHOD:**



The role of Fuzzifier is to convert a crisp input variable into linguistic variables. That is ready to be processed by the inference engine. The inference engine using the fuzzified inputs and the rules stored in the rule base processes the incoming data and produces linguistic output. Once the output linguistic values are available, the Defuzzifier produces the final crisp values from the output linguistic values.

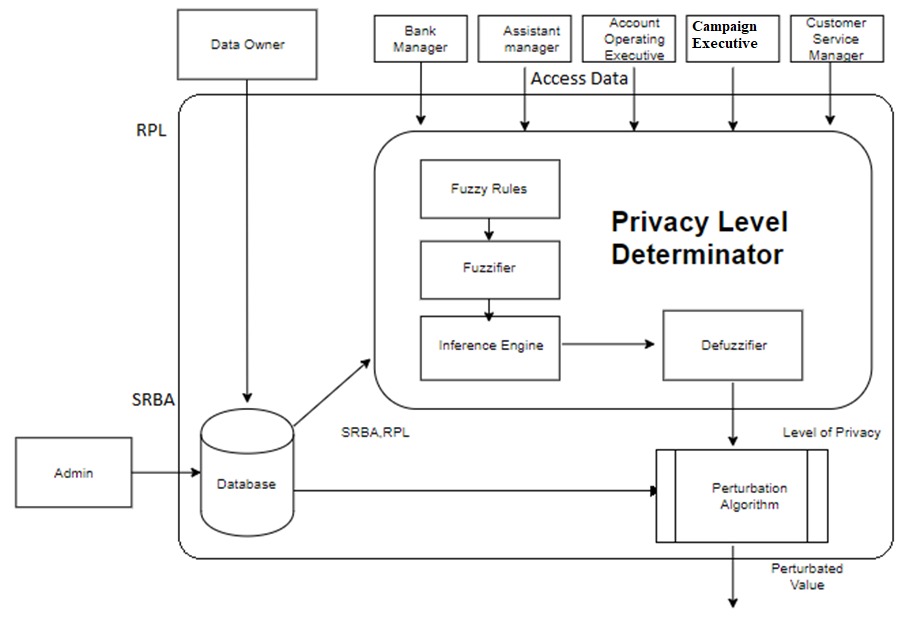
**System Model:**

The participants considered for this banking system are,

* Bank Manager
* Assistant Manager
* Campaign Executive
* Customer Service Manager
* Account Operating Executive

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:**

1. **Input**. The crisp value of various input parameters like age, marital status, average yearly balance, existence of loan, prev0ious contact outcome, duration of contact is obtained in the normalized form.

2. **Evaluate the main parameter**. Determine the SRBA, RPL crisp values.

3. **Fuzzify the crisp values of inputs**. Through the use of membership functions defined in fuzzy set for each linguistic variable, determine the degree of membership of every crisp value in the fuzzy set.

4. **Fire the rule bases that correspond to these inputs**. All systems which are based on fuzzy logic uses IF-THEN rules. The “IF” part is known as antecedent or premise, whereas the “THEN” part is termed as a consequence or conclusion. There are three outputs: Low, Medium and High.

5. **Execute the inference engine**. After the fuzzification of all crisp input values into their respective linguistic values, the inference engine will use the fuzzy rule base of the fuzzy expert system for deriving linguistic values for the intermediate and the output linguistic variables.

**6. Defuzzification.** Now the defuzzification of the linguistic values of the output linguistic variables into crisp values is the final step of the fuzzy system.

**Member Functions:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **S.no** | **Type** | **Variable Name** | **Full Form** | **Range** | **Membership**  **Function** |
| **1** | Input | RPL | Required Privacy Level | 0-10 | Low  Medium  High |
| **2** | Input | SRBA | Specific Role Based Access | 0-10 | Allow  Not Allow |
| **3** | Output | LOP | Level Of Privacy | 0-12 | Low  Medium  High |

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

**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.

1. Hüllermeier, E.: Fuzzy methods in machine learning and data mining: Status and prospects. Fuzzy Sets and Systems. 156, 387-406 (2005).
2. W. Pedrycz, “Fuzzy set technology in knowledge discovery,” Fuzzy Sets Syst., vol. 98, pp. 279–290, 1998.
3. 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
4. Nimpal Patel, Shreya Patel “A study on data perturbation techniques in privacy preserving data mining” International Research Journal of Engineering and Technology (IRJET) Volume: 02 Issue: 09 | Dec-2015