Granular Fuzzy Web Intelligence Techniques for Profitable Data Mining

Yan-Qing Zhang, M. Shteynberg, S. K. Prasad and R. Sunderraman

Department of Computer Science
Georgia State University
Atlanta, GA 30303 U.S.A.
yzhang@cs.gsu.edu

Abstract- **Data mining has a lot of e-Commerce applications. The key problem is how to find useful hidden patterns for better business applications. For these problems, granular fuzzy Web intelligence techniques are used to implement the granular fuzzy Web data mining system for available historical data of the credit company customers. Fuzzy computing and granular computing are used to design the Web fuzzy-interval data mining system that can do fuzzy-interval data clustering under uncertainty.**

Index Terms—Fuzzy Logic, Granular Computing, Computational Web Intelligence, Data Mining, e-Commerce

I. Introduction

Data mining has useful business applications such as finding useful hidden information from databases, predicting future trends, and making good business decisions [1][6][7]. Soft computing techniques such as fuzzy logic, rough sets, and neural networks are useful in data mining [1][10]. Granular computing techniques [3][4][9] can be used in data mining applications [3][4][8][11]. Using different techniques to design a hybrid system is a useful approach because different technical merits can be combined [12]. Now the interesting problem is how to use soft computing and granular computing to enhance the intelligent functionality of data mining systems. For the challenging problem, in this paper, both fuzzy computing and interval computing techniques are used to design the fuzzy-interval data mining system for credit card companies with actual large data sets. Fuzzy logic is used to cluster customers into several groups based on the customers' data. In addition, the interval computing technique is applied to do classification of customers in a different way.

Computational Web Intelligence (CWI) is a hybrid technology of Computational Intelligence (CI) and Web Technology (WT) dedicating to increasing quality of intelligence of e-Business applications on the Internet and wireless networks [13]. CWI uses Computational Intelligence (CI) and Web Technology (WT) to make intelligent e-Business applications on the Internet and wireless networks. Especially, how to effectively handle uncertainty of smart e-Business systems to make the Internet and wireless networks intelligent is the key thrust of CWI. Fuzzy logic, probabilistic methods, rough sets, neural networks, granular computing and evolutionary algorithms are major techniques for CWI. This paper focuses on using granular fuzzy Web intelligence techniques for a profitable data mining system for credit card users so as to do better marketing.

The rest of the paper is organized as follows. The profitable data mining using granular fuzzy Web intelligence techniques for credit card companies is discussed in Section II. The basic implementation of the Web-based granular fuzzy data mining software system is given in Section III. Finally, conclusions are given in Section IV.

II. Fuzzy Interval Web Intelligence Technicues for Profitable Data Mining

A. Overview

The goal of analyzing customers' historical data is to evaluate how possible customers may respond the offers. So, we need a powerful algorithm to calculate possibilities of the responses based the customers' historical data for the relevant periods of time, credit card data and mortgage rates.

If the data set is large and the number of input variables is large, fuzzy reasoning will be very hard to implement because the number of fuzzy rules is also large. In this case, interval computing would be a better choice.

B. Fuzzy Computing

A fuzzy rule is given:

IF X is A and Y is B and Z is C

THEN $P1$ is α, $P2$ is β and $P3$ is γ,

where X is the first input fuzzy linguistic variable for *Cash Balances Difference*, Y is the second input fuzzy linguistic variable for *Credit Card Balances Differences*, Z is the third input fuzzy linguistic variable for *Difference between an existing Mortgage Rate and current Mortgage Rate*, P1 is the first output fuzzy linguistic variable for the *probability for the first offer*, P2 is the second output fuzzy linguistic variable for *the probability for the second offer*, and P3 is the third output fuzzy linguistic variable for *the probability for the third offer*. A, B and C are fuzzy linguistic values of X, Y and Z, respectively, and α, β and γ are fuzzy linguistic values of $P1$, $P2$ and $P3$, respectively.

27 fuzzy rules with triangular fuzzy membership functions are used to map original customers' data in databases to the three output values. The outputs may show the probability of the customer response for all the three offers: first mortgage offer, refinancing an existing mortgage offer, and adding a second mortgage offer.

Each input value is divided into three fuzzy linguistic values Low, Medium and High, depending on the interval, where the value falls. (Note: Almost all input categories are cumulative, which means, for example, that a customer is the best candidate for the mortgage offer if his cash balances

have increased considering that he does not have any mortgage.)

The crisp output can be generated by the traditional center-of-gravity fuzzy reasoning method [2]:

$$output = \frac{\sum_{i=1}^{27} w_i \mu_{A_i}(x)\mu_{B_i}(y)\mu_{C_i}(z)}{\sum_{i=1}^{27} \mu_{A_i}(x)\mu_{B_i}(y)\mu_{C_i}(z)} \quad (1)$$

where μ_{Ai}, μ_{Bi}, and μ_{Ci} are the membership functions of input values x, y and z, respectively, and w_i is the center of gravity of the relevant fuzzy output of the fired fuzzy rule i.

C. Interval Computing

The interval computing rule is

IF X is a and Y is b and Z is c

THEN $P1$ is α, $P2$ is β and $P3$ is γ,

where X is *Cash Balances Difference*, Y is *Credit Card Balances Differences*, Z is *Difference between an existing Mortgage Rate and current Mortgage Rate*, P1 is the *probability for the first offer*, P2 is *the probability for the second offer*, and P3 is *the probability for the third offer*. a, b and c are interval linguistic values of X, Y and Z, respectively, and α, β and γ are fuzzy linguistic values of $P1$, $P2$ and $P3$, respectively.

Basic steps of the interval computing algorithm:

(1) Assign each input value to a particular interval linguistic value: Low, Medium or High. So a, b and $c \in$ {Low, Medium, High}.

(2) Determine the output category based on the combination of the input values using the interval IF-THEN rules.

(3) Assign a crisp value to the output score based on the following classification: Low is in [0, 0.3), Medium is in [0.3, 0.7), and High is in [0.7, 1.0].

III. System Implementation and Performace

The Web-based fuzzy interval data mining system using the fuzzy computing algorithm and the interval computing algorithm is implemented, and introduced in this Section [5].

A. System Configuration

As it was mentioned above, implementation of fuzzy rule base gets very complicated with every new input variable we need to consider in the process of data mining. So, the following decision was made: consider only three major influence factors in this system implementation, treat all other input variables as a constant factor. Basically, one rule base consisting of twenty-seven rules is implemented. The results of the data mining of the credit company customers data using only one major rule base came out precise and reliable.

Very important part in the successful system modeling is to have the actual set of data similar to what the system will have to deal in the real world. The data for processing were found in the open source database on the Web. For the security purposes, last names of the bank customers have been modified. Also, accounts balances were rounded off for easier processing.

The application is a web-based program, written using Java Servlets, with database on Oracle 8i server. All information processing is dynamic. The speed of data processing highly depends on Internet connection speed, speed of the processor of the local system and amount of RAM of the local system. System's input is the information about the credit company customers, which is available to the credit company managers – targeted users of the application. System's goal is to "mine" the input data in order to produce bank customers' classification based on their eligibility for a particular bank offer. Three possible credit company offers are considered: opening first mortgage, adding second mortgage and refinancing of the existent mortgage loan. So, the output of my system is a list of customers eligible for each of those offers with the scores representing probability of response.

The application produces a dynamic output with the results of data processing: all credit company customers are divided into three categories, depending on the highest response probability for credit company offers.

The database with the credit company customers' information resides on Oracle 8i server. The application is written using Java dynamic HTML pages – servlets. Each servlet either updates database or just requests all necessary data. Users will be able to use their customers' database on their own server, without using World Wide Web. Since the engine of the software is written in Java, the only constraint for the server will be a support for Java.

B. System Functionalities

The application is implemented as an on-line single user application. GUI is a very important software component, in order for the software to survive on the market it has to have an attractive, user-friendly Graphical User Interface, or GUI. The application is easily customizable, so as a part of my future users requirement GUI could be made more sophisticated. The application start up page is login screen, where the existing customer will only type in a unique user number, which my system assigned to every user, and a password. New users of the application are taken to another screen, where they fill out a simple application with the information about themselves and their company.

The application start up page is login screen, where the existing customer will only type in a unique user number, which the system creates automatically for each new user, and a password. New users of the application are taken to another screen, where they fill out a simple application with the information about themselves and their company.

All the information is displayed on the widest part of the screen – right frame. If the user clicks on the link "Add a new customer", an application appears on the right frame to be filled out for the new customers. Upon a successful login, a three-framed screen appears. The upper frame is just a banner welcoming a user to the application, which contains the user's name and a greeting message. The left part of the

screen is the menu with the list of commands – active links and buttons. The results of the user action appear on the right frame. Here the user will update all the input information and will see the results of the data processing.

The next step is to process all the available information to classify the customers by eligibility for one of three available offers. By clicking on "Process Clients Data" link, the warehouse information is processed on the server in a real time, and the system user is shown the results of the processing for all the clients. The table displayed on the right hand side screen contains information with client name, ID, first offer score, second offer score and third offer score. The value of a score is an equivalent of the probability of the response for a particular offer. The closer this value to 1.0, the more possible, that the given customer is a good candidate for the offer.

If the user wants to see a list of the customers qualified for the particular offer, he needs to click on any of the three links depending on which offer he is interested in: "View Candidates for First Mortgage Offer", "View Candidates for Refinancing Offer", or "View Candidates for Second Mortgage Offer". The user may click on the "e-mail" link and the customer will be notified by an instant email, that this company approves his or her offer.

IV. Conclusions

Both the fuzzy computing algorithm and the interval computing algorithm are implemented. The benefits of using the granular fuzzy Web intelligence technique here are: (1) a relatively robust classification of the input data sets with uncertainty, (2) the granular fuzzy reasoning algorithm is easy to implement, and (3) fuzzy interval reasoning algorithm can use a small number of rules to do data classification. In the future, real data sets will help to verify the fuzzy interval data mining method, and improve it.

References

[1] K. Cios, W. Pedrycz, and R. Swiniarski, *Data Miming methods for Knowledge Discovery*, Kluwer Academic Publishers, 1998.

[2] G.J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic Theory and Applications*, Prentice Hall P T R, 1995.

[3] T.Y. Lin, "Granular Computing on Binary Relations I: Data Mining and Neighborhood Systems," *Rough Sets in Knowledge Discovery*, pp. 107-121, A. Skowom and L. Polkowski (eds), Springer-Verlag, 1998.

[4] T.Y. Lin, "Granular Computing on Binary Relations II: Rough Set Representations and Belief Functions," *Rough Sets in Knowledge Discovery*, pp. 121-140, A. Skowom and L. Polkowski (eds), Springer-Verlag, 1998.

[5] M. Shteynberg, "Mining Large Data Sets with Fuzzy Reasoning Methods", thesis, 2000.

[6] K. Thearling, "Increasing customer value by integrating Data Mining and Campaign Management Software," Direct Marketing Magazine, February 1999.

[7] C. Westphal and T. Blaxton, *Data Mining Solutions - Methods and Tools for Solving Real-World Problems*, John Wiley \& Sons, 1998.

[8] Y.Y. Yao and N. Zhong, "Potential applications of granular computing in knowledge discovery and data mining," *Proceedings of World Multiconference on Systemics, Cybernetics and Informatics*, Vol. 5, pp. 573-580, 1999.

[9] L.A. Zadeh, "Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic," *Fuzzy Sets and Systems*, Vol. 19, pp. 111-127, 1997.

[10] Y.-Q. Zhang, M. D. Fraser, R. A. Gagliano and A. Kandel, "Granular Neural Networks for Numerical-Linguistic Data Fusion and Knowledge Discovery," Special Issue on Neural Networks for Data Mining and Knowledge Discovery, *IEEE Transactions on Neural Networks*, Vol. 11, No. 3, pp.658-667, May, 2000.

[11] Y.-Q. Zhang, G. S. Owen and R. Sunderraman, "Quick Granular Rule Generation from a Database," *Proc. of the 5th Joint Conf. on Information Sciences*, Vol.1, pp.194-197, Feb. 27 - March 3, 2000.

[12] Y.-Q. Zhang and A. Kandel, "Compensatory Genetic Fuzzy Neural Networks and Their Applications," *Series in Machine Perception Artificial Intelligence*, vol. 30, World Scientific, 1998.

[13] Y.-Q. Zhang and T.Y. Lin, "Computational Web Intelligence (CWI): Synergy of Computational Intelligence and Web Technology," Proc. of FUZZ-IEEE2002 of World Congress on Computational Intelligence 2002: Special Session on Computational Web Intelligence, pp.1104-1107, May 2002.