J. Basic. Appl. Sci. Res., 3(12)379-384, 2013

ISSN 2090-4304
Journal of Basic and Applied Scientific Research
www.textroad.com

Classification of Bank Customers for Granting Banking Facility Using Fuzzy Expert System Based on Rules Extracted from the Banking Data

Ali Bazmara[1], Soheila Sardar Donighi[2]

[1]Department of Management and Economics, Science and Research Branch Islamic Azad Universities, Tehran, Iran
[2]Department of Industrial managment, Islamic Azad University, Tehran North Branch, Iran
Received: October 4 2013
Accepted: November 192013

ABSTRACT

Banking facility in recent years is faced with problems such as customer credit risk. And more accurately predict of worthiness of customers is the overall purpose of granting credit. So far many techniques using advanced artificial intelligence systems have been used for this purpose. But they all had advantages and disadvantages. In this article moreover we try to classify customer credit using fuzzy expert system based on rule extracting from banking data, the weaknesses of previous methods, most notably the indescribable and the dependency of experts will address.
In the implemented system, the rules were extracted of data rather than expert. So there is no longer dependence on the expert knowledge and the banking rule. And also changes in banking regulations will not cause the system to lose its efficiency. Because of the fuzzy rules, the proposed system is describable and an expert who can accept the validity of the decision to approve or reject it. The results of comparing fuzzy expert systems with the existing methods show that the proposed method is more accurate and will not have the drawbacks of previous methods. The using data was collected from Ansar Bank in Iran.
KEYWORDS: expert system, fuzzy rule, banking facility, rule extracting, interpretability

I. INTRODUCTION

With increasing advances in science, the need for more efficient decision making seems to be more essential. This requirement due to its importance in the field of finance and economics seems to be more considerable. Using traditional methods of making decisions based on personal opinion, in some of which with very sensitive matters such as finance and economics, was not acceptable because of the large errors. So we use the decision models with high accuracy which are going to give us the science of artificial intelligence that allow to accurately determine patterns that are identical for users to make decisions. The most important such systems are fuzzy expert systems (Bazmara M., et al, 2012). and data mining algorithms that have many applications in different sciences (Bazmara M., et al, 2013, Jamali I., et al, 2012). artificial intelligence not only used in banking but also issues such as bioinformatics, geographic information systems and etc (Mohammadi F., et al, 2013, Ramadhani S., et al, 2012).

Financial market has the main role in the mobilization and allocation of investigation resources in the economic activities. Granting facility will form an important part of the operations financed by bank. Since the banks and financial institutions seeking to avoid the risk of not returning of principal and profit of in granting credit facilities, credits which are irreversible with high probability should be identified by decision-making system to perform appropriate actions. In (Lee T.-S., et al, 2006) the modern banking system, different decision-making systems are an alternative to traditional banking facilities. The advantages of these methods are removing personal judgments, and increasing the accuracy and confidence and also reducing the time and cost of the evaluating condition.

Currently intelligent systems are used to optimize decision-making process, and as one of the tools developed in the field of banking have many applications. Most of such systems are based on artificial intelligence algorithms and expert systems. The output of these algorithms are models generated based on the input information with high accuracy. Input data includes the type of facility (loan), amount and number of loans, type of repayment, applicants' income, the number and value of guarantee and etc. Finally, the proposed model divides applicants into groups (Thomas L.C., 2000). In this classification, the model should consider the amount of facility allocated to the applicant so that Less risk for the bank to be established.

In (Chen L.-H., et al, 1999) a fuzzy approach for credit rating of a bank's customers in Taiwan were studied. The problem with this approach is that the rules were based on expert knowledge. While different experts produce different rules and also with the change of the rules of this field, fuzzy rules lose their effectiveness. In (Van Gestel I.T., et al) credit scoring was done by using support vector machine (SVM) and least square error technique. Indescribably are the most disadvantages of this approach, because the expert

*Corresponding Author: Ali Bazmara, Department of Management and Economics, Science and Research Branch Islamic Azad Universities, Tehran, Iran

couldn't understand why such an output was determined. In (Lahsasna A., et al, 2010) several methods like neural network, support vector machine and evolutionary algorithms are used as the soft computing methods to resolve the credit scoring problem. As we know neural networks and support vector machine do like a black box and don't explain anything about the reasons of making decision to the users. Also evolutionary algorithms have a stochastic process which there is no insurance to reach the same answer in several executions. In (Chen F.-L., et al, 2010) combination of support vector machine and four different methods, decision tree, linear discriminant analysis, rough sets and F-score, are used to credit rating. High complexity of computation is the most noticeable problem in this combination. Also (Chen M.-C., et al, 2003) was studied, which used neural networks and genetic algorithm to classify customers into two groups, "good" and "bad". This method like the previous method has high computational complexity and also because of the stochastic computation, has no the fixed precision in its decision. In (Madani A.M., et al, 2012) a model was obtained for credit rating by using logistic regression for a bank in Iran. This paper used information of 310 people to determine the main factors in credit risk. This data is used to find the performance of the model. A genetic programming method was also used in (Abdou A. H., 2009). for credit scoring in Egyptian public sector banks and the results of the model were compared with probit analysis (an alternative to logistic regression) and weight of evidence measure. Two evaluation criteria used in the modeling were average correct classification and estimated misclassification cost.

We try to purpose an expert system so that, at first doesn't need to expert for determining rules, secondly is describable and has no computational complexity. Thirdly generates high accuracy output for users and finally dynamically acts so that doesn't lose its performance by changes in banking rules.

II. DATA COLLECTION

To dare be said that, in this study the most tedious and the longest time of work was obtaining and collecting data. As mentioned in the previous description, the advantage of fuzzy system compared to others is extracting rules from data. If data is changed in the future, these rules could extract in the line with them. Whereas other systems that laws are applied to, are not applicable for other banks or couldn't work with other information.

The data set used in this article includes customers who have account in a time interval on bank (Ansar Bank). And till now they may use of the bank facility. In the present study, the population have worked on was included of 2,100 clients of Ansar Bank.

Data preparation is one of the most important part of data mining process, which has several procedures such as data cleaning, data normalization and making required features. The raw data received from the database of Ansar Bank were in four separated tables, shown in tables 1 to 4, which were then aggregated at a common table.

Table 1: customers' personal information

Customer ID	Date of birth	Career	Sex	Marital status	Education
1	1966/4/5	Army	Male	Married	Bachelor
2	1970/12/1	Operator	Male	Single	Diploma
3	1958/3/9	Employment	Female	Married	Associate Degree
4	1961/16/1	Retired	Male	Married	Diploma

Table 2: customers' deposit information

Customer ID	No deposit	Deposit account($)	Opening deposit date
1	3612_854_1_1	219	2012/25/12
2	2076_801_2_17	12001	2009/13/1
3	4314_804_3_1	315525	2006/5/7
4	2030_705_4_17	72	2007/22/9

Table 3: number of returned check for each customer

Customer ID	No returned check
1	3
2	0
3	0
4	1

Also on the table, there is a set of data (Characteristics), that doesn't require for deducing model or extracting rules, like the description of facilities field, no deposit and no credit, such the features were removed from the final table. Moreover, some features are merged together and new features were created.

In most articles and jobs in this field a few selected features are used. Therefore, this article did not receive the output with high accuracy. Although we initially tried to aggregate all the necessary features into a complete

data set for validation. Features that were used in other papers was given to the experts to determine the label of data(class) for us. The vast majority of them were of the opinion that by using these features can not be sure of the tags for the sample. Then, each expert was asked to give the properties of data according to his opinion. So the number of features that the was increased from 6 to 27 characters. Hence the experts were able to label each data more confidence by using the 27 properties. A part of obtained data set is shown in table 5.

Table 4: customers' loan information (* This feature is used just in Iran)

Customer ID	No deposit	Deposit value($)	Deposit type	Deposit status	Description*	Guarantee value($)	No Guarantee
1	3611_102_1_13	14285	428	Active	Loan with no fees	9573	1
1	4311_101_1_1	8571	318	Delayed	Loan with no fees	29191	1
1	4310_101_1_1	14285	320	Active	Loan with no fees	26575	3
1	4310_101_1_2	14285	420	Active	Loan with no fees	28699	1
1	3610_101_1_1	2857	800	Expired	Loan with high fees	160844	1
2	4311_101_2_2	1428	320	Delayed	Loan with no fees	29914	1
2	4310_101_2_1	3428	421	Active	Loan with no fees	26121	2
2	4311_701_2_1	11428	497	Active	Loan with high fees	36711	3
3	4310_101_3_4	2857	421	Delayed	Loan with no fees	23121	1

Table 5: aggregate table

features Expert's opinion	0	1	3	2
Type1-active	1	1	7	3
Type1-expired	0	0	0	0
Type1-delayed	0	0	0	0
Type1-doubtful	0	0	0	0
Type1- settled	1	0	1	0
Type2-active	3	2	2	2
Type2-expired	0	0	0	0
Type2-delayed	0	0	0	0
Type2-doubtful	0	0	0	0
Type2- settled	0	0	1	0
Type3-active	0	0	0	0
Type3-expired	0	0	0	0
Type3-delayed	0	0	0	0
Type3-doubtful	0	0	0	0
Type3-settled	0	0	0	0
No guarantee	5	5	13	5
Guarantee value	103178	391714	126854	137057
Sex	1	0	0	0
Age	42	38	50	39
Education	7	5	9	5
Marital status	2	2	2	1
Career	8	22	5	22
No returned check	3	0	0	0
Deposit account ($)	1352	8147	82318	36899
Deposit period (month)	121	65	101	124
Average salary($)	1400	2628	2428	2742

III. METHODOLOGY

3.1. Rule extraction and representation of knowledge by the "if- then" rules of fuzzy

Fuzzy logic is capable of formulating linguistic rules for fuzzy models so that can be easily understood by experts. While, all mathematical details are concealed. To do this, knowledge was represented by rules of fuzzy if - then that is displayed on the below form:

If x_1 Is A_1 AND x_2 IsA_2 ... AND x_m Is A_m THEN y is B

That $x_1 \ldots x_m$ Are the linguistic input variables with value of A_1 To A_m And y is the linguistic output variables with value of B.

Systematic approach has been used in the past weren't completely intelligent and the main base of such systems was the rules obtained by the experts. If experts comment and give his opinion and determine rules, the intelligent system will be removed. In this article we try to do the entire process intelligently. Therefore, we would use data to get rules and not to use predefined rules in the system. In the rule extraction phase the famous method "Wang & Mendel" is used which has 5 phase to get final solution[15]:

Phase 1: divide the input and output space into fuzzy region

Phase 2: generate fuzzy rules from given data pairs

Phase 3: assign a degree to each rule
Phase 4: create a combined fuzzy rule base
Phase 5: determine a mapping based on the combined fuzzy rule base

3.2. From crisp to fuzzy sets

Let U be a collection of objects u which can be discrete or continuous. U is called the universe of discourse and u represents an element of U. A classical (crisp) subset C in a universe U can be denoted in several ways like, in the discrete case, by enumeration of its elements: $C = \{u_1, u_2,..., u_p\}$ with $\forall i$: $u_i \in u$. Another way to denote C (both in the discrete and the continuous case) is by using the characteristic function X_F:$U \rightarrow \{0, 1\}$ according to $X_F(u) = 1$ if u∈ C, and $X_F(u) = 0$ if u∉ C. The latter type of definition can be generalized in order to define fuzzy sets. A fuzzy set F in a universe of discourse U is characterized by a membership function μ_F Which takes values in the interval [0, 1] namely, μ_F: U→[0,1].

3.3. Operators on fuzzy sets

Let A and B be two fuzzy sets in U with membership functions μ_AAnd μ_B, respectively. The fuzzy set resulting from operations of union, intersection, etc. Of fuzzy sets are defined using their membership functions. Generally, several choices are possible:

Union: The membership function $\mu_{A\cup B}$ Of the union $A \cup B$ can be defined by:

$\forall u$:$\mu_{A\cup B}$= $max\{\mu_A(u), \mu_B(u)\}$ or by $\forall u$:$\mu_{A\cup B}$= $\mu_A(u) + \mu_B(u) - \mu_A(u)\mu_B(u)$.

Intersection: The membership function $\mu_{A\cap B}$ of the union for all $A \cap B$ can be defined by:

$\forall u$:$\mu_{A\cap B}$= min$\{\mu_A(u), \mu_B(u)\}$ or by $\forall u$:$\mu_{A\cap B}$= $\mu_A(u)\mu_B(u)$

Complement: The membership function of the complementary fuzzy set A^C Of A is usually defined by $\forall u$: $\mu_{A^c} = 1 - \mu_A(u)$.

3.4. Linguistic variables

Fuzzy logic is able to model of expert knowledge. The key notion to do, is that of a linguistic variable (instead of a quantitative variable) which takes linguistic values (instead of numerical ones). In the used dataset there are 5 features of fuzzy properties. These features are: age, guarantee value, deposit account, deposit period and average salary. Seven levels are considered for each linguistic variable. In other words the value of the linguistic variable can be an interval of its range. For example, age as a linguistic variable has the range of 21 to 86 year olds (T(age)∈ U[21,86]), this range is divided into seven level. Each variable in T(age) is characterized by a fuzzy set in the universe of discourse, here, e.g., U = [21, 86]. Therefore each level has a sub function which shown the degree of dependency of each value to this level. Table 6 shows the range of each level for linguistic variable age.

Table 6: levels and their interval of linguistic variable age

Level	Its interval
L1	20 – 33
L2	30 – 44
L3	39 – 53
L4	50 – 65
L5	61 – 74
L6	70 – 83
L7	80 – 90

For each linguistic variable boundary of class will be fuzzy. Therefore, these linguistic values are characterized by fuzzy sets described by a membership function as shown in Fig.1 designed by Matlab. Each level has a triangular membership function (trimf) that is shown in figure by its name.

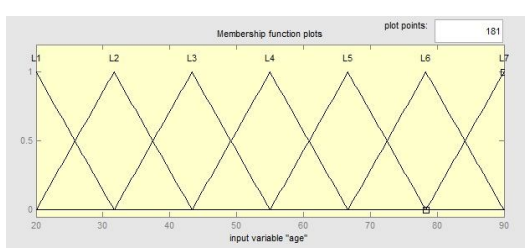

Fig. 1 linguistic variable for feature "age"

Also fig 2 shows the membership functions for the linguistic variable "average salary". For this variable the membership function is triangular. Average salary has the range of 657$ to 17028$ $(T(Salary) \in U[2300000,59600000])$, levels and their values are shown in table 7.

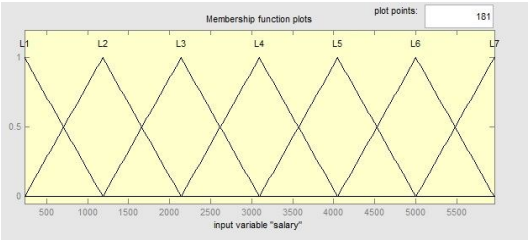

Fig. 2 linguistic variable for feature "average salary"

Table 7: levels and their interval of linguistic variable average salary

Level	Its interval
L1	571 – 3714
L2	2571 – 6857
L3	5714 – 8571
L4	7714 – 10857
L5	10000 – 13142
L6	12571 – 16258
L7	15428 – 17142

IV. CONSTRUCTING THE FUZZY MODEL AND EVALUATION

The fuzzy expert system was executed as a independent process on the Ansar Bank dataset to determine whether or not the loan is paid. Then the model was reviewed by experts and university's professors.

4.1. Reviewed no 1

A group of 21 to 86 year old customers were selected. The data set included of 2026 rows of the information of this group. Each selected customer allocated some values for all features. The fuzzy system could get the accuracy of 97%. According to the experts' viewpoint this system had High precision. Respect to its interpretability, this method could be used in banks practically.

4.2. Reviewed no 2

18 experts of the best are requested to explain about the output of fuzzy system. Each expert could choose one of the three situations “perfectly agree", "good" and "disagree". The result was shown in table 8.

Table 8: experts, viewpoints for fuzzy system

View point	No
Perfectly agree	11
Good	5
Disagree	2

V. CONCLUSION

As results and evaluation were shown, fuzzy system could get good outcome for solving the bank customer classification problem. The remarkable improvement in the efficiency is to extract rules using the dataset. Furthermore, the dependence of the expert was removed. So the percentage error would 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 changing 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.

Acknowledgment

The authors declare that they have no conflicts of interest in this research.

REFERENCES

Abdou A. H., (2009). Genetic programming for credit scoring: the case of Egyptian public sector banks. *Expert Systems with Applications*, 36, 11402-11417.

Bazmara M., Jafari S., & Pasan F. (2012). A Fuzzy expert system for goalkeeper quality recognition. *International Journal of Computer Science Issues (IJCSI).* 9(5), 318.

Bazmara M., & Jafari S. (2013). K Nearest Neighbor Algorithm for Finding Soccer Talent. *Journal of basic and applied scientific research*. 3(4).

Chen F.-L., & Li F.-C, (2010). Combination of feature selection approaches with SVM in credit scoring. *Expert Systems with Applications*, 37(7), 4902-4909.

Chen L.-H., & Chiou T.-W., (1999). A fuzzy credit-rating approach for commercial loans: a Taiwan case. *Omega*, 27(4), 407-419.

Chen M.-C., & Huang S.-H., (2003). Credit scoring and rejected instances reassigning through evolutionary computation techniques. *Expert Systems with Applications*, 24(4), 433-441.

homas L.C. (2000). A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. *International Journal of Forecasting*, 16(2), 149-172.

Jamali I., Bazmara M., & Jafari S. (2012). Feature Selection in Imbalance data sets. *International Journal of Computer Science Issues*, 9(3), 42-45.

Lahsasna A., Ainon R.N., & Wah T.Y., (2010). Credit Scoring Models Using Soft Computing Methods: A Survey. *Int. Arab J. Inf. Technol*., 7(2),115-123.

Lee C.-S., & Wang M.-H., (2011). A fuzzy expert system for diabetes decision support application. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 41(1), 139-153.

Lee T.-S., Chiu C.-C., Chou Y.-C., & Lu C.-J. (2006). Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Computational Statistics & Data Analysis*, 50(4), 1113-1130.

Madani A. M., Madani Y., EbrahimZadeh M., & Shahmorad G. M., 2012. Modeling credit Rating for bank of Eghtesade Novin in Iran. *Journal of basic and applied scientific research*, 2(5), 4423- 4432.

Mohammadi F., & Bazmara M. (2013). A New Approach of Fuzzy Theory with Uncertainties in Geographic Information Systems. *Journal of Mechatronics, Electrical and Computer Technology*, 3(6), 1001-1014.

Ramadhani S., Mousavi S.R., & Talebi M.(2012). An improved heuristic for haplotype inference. *Gene*, 507(2), 177-182.

Van Gestel I.T., Baesens B., Garcia I.J., & Van Dijcke P. A support vector machine approach to credit scoring, Forum Financier-Revue Bancaire Et Financiaire Bank En Financiewezen, *Citeseer*, 73-82.