The Credit Scoring System For Evaluating Personal Loans Based On The Fuzzy Sets Theory

Nikita S. Lukashevich, Dmitriy A. Garanin

St. Petersburg State Polytechnical University; St. Petersburg, Russia
mailto:lukashevich@kafedrapik.ru

Abstract: The implementation of the Third Basel Accord raises many technical and methodological questions regarding the development and validation of credit risk models and makes these questions much more important. The current article presents a model for creditworthiness analysis of the bank loan applicants. The fuzzy sets theory is approved as the proposed approach to the creditworthiness assessment using the factual data. An approach based on profiling of the distribution histograms was developed to find the nodal points of the membership functions and eliminate their subjectivity. The factors that determine the creditworthiness of the borrowers were selected and the fuzzy classifiers were simulated based on the obtained quasi-statistics to find the exact parameters of the membership functions.

[Lukashevich N.S., Garanin D.A. **The Credit Scoring System For Evaluating Personal Loans Based On The Fuzzy Sets Theory.** *Life Sci J* 2014;11(10):1118-1122]. (ISSN:1097-8135). http://www.lifesciencesite.com. 168

Keywords: creditworthiness; credit scoring; credit rating; fuzzy set; fuzzy classifier4 linguistic variable; aggregation.

1. Introduction

Lending to individuals is the promising financial activity in the world and domestic banking sector. The competition in the retail lending market forced lenders to conduct the aggressive credit policies in order to increase the loan portfolio by engaging in a short time a wide range of the borrowers. That problem was solved by simplifying the credit procedures. Attracting borrowers in a such way creditors take the additional credit risk, which led to an increase in the defaulted debt. The problem of reducing credit risk has become an urgent task. Between credit risk and the creditworthiness of the borrower is a logical association. From this, we can conclude that the correct credit policy in assessing the borrowers' creditworthiness will allow the credit institutions to carry less risk in the credit operations. An important condition for the effective credit operations on retail lending becomes an active investigation of the issues related to the credit scoring models that allow making a reasonable selection of the potential borrowers. In accordance with the latest Basel Capital Accord, known as Basel III, to assess borrowers when lending it is recommended to use an approach based on the internal credit ratings as follows: development of the internal credit ratings system; assigning the borrowers' credit ratings; estimating the risk parameters such as probability of default (*PD*), loss given default (*LGD*), exposure at default (*EAD*), maturity (*M*). This approach involves the development of the adequate mathematical models for all mentioned above stages.

2. Materials and methods

Each credit scoring model can be summarized as follows [1]:

$$< I_0 (G, L, \Phi, A) >,$$

where I_0 - credit rating as a measure of creditworthiness of the borrower; G - a set of factors of the borrower's creditworthiness; L – a set of estimates for each factor from the set G; Φ - a set of weights defining the significance of each factor from the set G; A – a method for calculation I_0.

In papers [2, 3, 4] authors examined several groups of methods that can be the basis of the credit scoring models. To develop models based on the statistical methods the analyst needs the historical data sample, which meets the stringent requirements on the homogeneity, its sufficient size and the invariance of the influence of the significant factors in the sample at a certain time interval. Neural networks, that are capable of adaptation, memorization and modeling the behavior of complex, multiply and nonlinear systems, have the disadvantage associated with the lack of the rigorous studies regarding the choice and the structure of the neural network and the practical impossibility of extracting acquired knowledge [3]. Limitations of the methods based on expert estimates is the subjectivity the possibility of an erroneous judgment. The fuzzy set descriptions have the main disadvantage consisting in the fact, that the parameters of the membership functions are chosen subjectively. In the context of the defined problem the main advantages of the fuzzy set approach are as follows: 1) an opportunity to describe the problem situation in a language close to natural one; 2) a possibility of solving the problems with unreliable data; 3) an opportunity to use the experience and intuition of the expert; 4) an opportunity to qualitatively and quantitatively assess the creditworthiness of the borrower.

The last assertion is very important because besides the quantitative values, it is necessary for the decision-maker to know whether the obtained values are acceptable and to what extent. The subjectivity when creating membership functions (classifier) can be partially eliminated by using the data described in terms of quasi-statistics (the sample which is not "classical", but which can be used to determine the model parameters) [5], which brings expert judgment to reality. The transition from the collected quasi-statistics to classifier is based on so-called linguistic analysis of distribution histograms for the factor values.

To assess the creditworthiness, based on the systemic principle (the borrower is considered as a system of interrelated characteristics), on the one hand, and fuzzy-set approach, on the other hand, the credit analyst has to implement the following steps: 1) identify the factors affecting the creditworthiness of the borrower; 2) determine the significance of the factors influence on creditworthiness; 3) build fuzzy-set classifiers for all factors; 4) determine the credit rating of the borrower by aggregating model factors.

Credit rating in this context is a qualitative or quantitative value of the creditworthiness of the borrower and its willingness and ability to repay loan resulting from credit application assessment.

Determination of the model factors: Let us propose the model factors, which are combined into meaningful groups. The integrated and private indicators (factors) are represented as an inference tree in Fig. 1.

Fig. 1 has the following notation: I_0 - credit rating; X_1 – the socio-demographic indicator; X_2 – the financial condition indicator; X_3 – the current liabilities indicator; X_4 - the special indicator of the borrower's credit history; X_5 – the credit security indicator; $X_{1.1}$ - the number of dependants/children, persons; $X_{1.2}$ - seniority, years; $X_{2.1}$ - the borrower's income, rub.; $X_{2.2}$ - the total income of the borrower's family, rub.; $X_{2.3}$ – the borrower's expenses, rub.; $X_{2.4}$ – the volatility of the borrower's income, %; $X_{3.1}$ – the current credit payments, rub./month; $X_{3.2}$ - other liabilities payment, rub./month; $X_{4.1}$ - term of the use of the consumer lending, years ; $X_{4.2}$ - a number of delays in a month payment; $X_{5.1}$ - the value of borrower's assets with high liquidity, rub. ; $X_{5.2}$ - the estimated value of the borrower's car, rub.; d_n – a variant of the credit decision based on I_0.

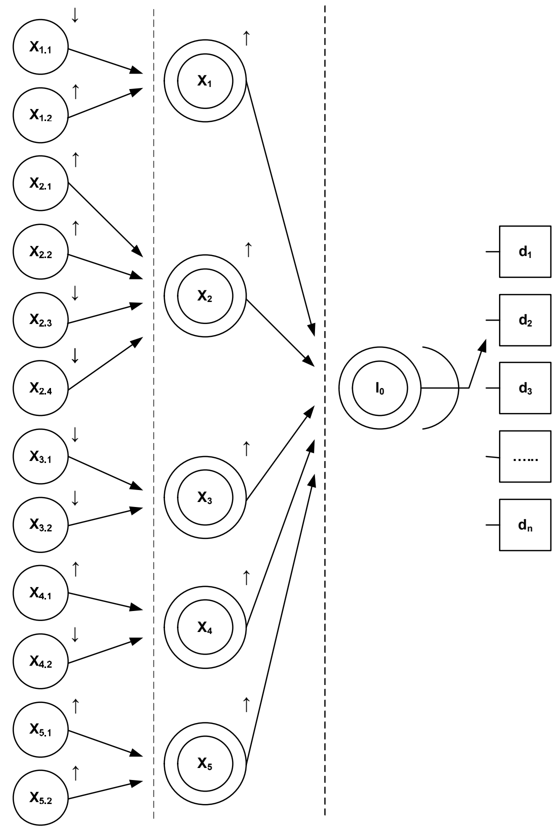

Figure. 1. Inference tree for factors of creditworthiness

Determination of the significance of the influence of factors: Let us form the system of weights for each level of factors. The system is made in a such way that

$$\sum_{i=1}^{I} p_i = 1, \sum_{j=1}^{J_i} p_{ij} = 1, \; i = 1 \ldots 5,$$

where p_i - the weight of X_i indicator; p_{ij} - the weight of X_{ij} indicator;

To build a system of weights were interviewed five experts who ranked the first- and second-level factors in order of importance. The coherence of factors ranking by experts was tested using the coefficient of variation. To determine the weights the scale of Fishburne was applied, moving from the ranks of the factors to the system of weights [6]. The final weight is formed starting from the weights determined by the experts, according to the method of Kemeny's median with Euclidean distance.

Development of fuzzy classifiers: For any second-level factor X_{ij} let us specify the linguistic variable B_{ij}=*{«level of factor X_{ij}»}* with the following term-set: B^1_{ij} - a subset of *«low level of factor X_{ij}»*; B^2_{ij} - a subset of the *«average level of factor* X_{ij}*»*; B^3_{ij} - a subset of «*high level of factor* X_{ij}». The next step is for each value of the linguistic variable B_{ij} to associate membership function values μ^k_{ij} of the factor X_{ij} with a particular fuzzy subset. Mathematically, it can be written as follows:

$$B^k{}_{ij} = \left\{x_{ij} \Big/ \mu_{ij}{}^k(x_{ij})\right\},\ \mu_{ij}{}^k(x_{ij}) \to [0,1],$$

$$x_{ij} \in [a_{ij}, b_{ij}]$$

$$i = 1..I,\ j = 1..J_i,\ k = \begin{Bmatrix} 1-\text{"}low\text{"} \\ 2-\text{"}average\text{"} \\ 3-\text{"}high\text{"} \end{Bmatrix}.$$

Scale gradation over three levels are not common, because the process of refining the scale becomes an "infinite" logic. Commonly used functions are the trapezoidal membership functions [7]. The upper base of the trapezoidal membership functions corresponds to the full confidence in the correctness of classification, and the lower one corresponds to the absolute unconfidence. Uncertainty in expert classification decreases (increases) linearly. For the purposes of describing the trapezoidal membership function compactly, it is convenient to define the trapezoidal numbers as follows:

$$(a_1, a_2, a_3, a_4),$$

where a_1 and a_4 - x-coordinate of the lower base of the trapezoid, a_2 and a_3 - x-coordinate of the upper base of the trapezoid.

3. Results

Let us find the exact classifiers parameters, i.e. for each fuzzy subset within the model factors find the trapezoidal numbers describing the corresponding membership functions. The collected quasi-statistics will enhance the objectivity of building the membership functions.

To calculate the trapezoidal numbers let us define the nodal point for the subset B^2_{ij} as the median of the distribution histogram. For calculating, the nodal points of subsets B^1_{ij} and B^3_{ij} let us apply the weighted average rule taking into account the histogram profile.

Intervals between the nodal points are divided into three zones of equal length: the absolute confidence zone, the uncertainty zone and absolute unconfidence zone. To illustrate this approach to the simulation of the fuzzy classifier let us consider quasi-statistics for the factor $X_{1.2}$, the distribution histogram of which is shown in Fig. 2.

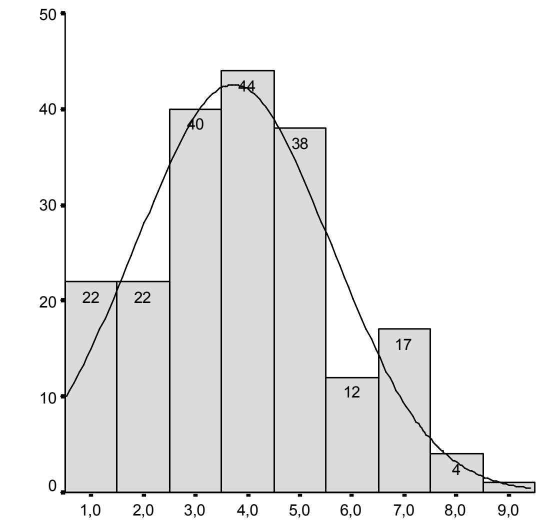

Figure. 2. The distribution bar chart for the factor $X_{1.2}$, years

Taking into consideration the profile of the distribution histogram for the factor $X_{1.2}$, it can be written:

$$\eta_{(12)C} = Me\{x_{12}\},$$

$$\eta_{(12)H} = {(1\cdot Z_1 + 2\cdot Z_2 + 3\cdot Z_3 + 4\cdot Z_4)} \Big/ {(Z_1 + Z_2 + Z_3 + Z_4)},$$

$$\eta_{(12)B} = {(4\cdot Z_4 + 5\cdot Z_5 + 6\cdot Z_6 + 7\cdot Z_7 + 8\cdot Z_8 + 9\cdot Z_9)} \Big/ {(Z_4 + Z_5 + Z_6 + Z_7 + Z_8 + Z_9)},$$

$$Z_l = \{22;22;40;44;38;12;17;4;1\}.$$

The result of the calculation are the following values of the nodal points:

$$\eta_{(12)LOW} = 2.8;\ \eta_{(12)AVERAGE} = 4;\ \eta_{(12)HIGH} = 5.$$

Knowing the values of the nodal points, the intervals can be formed by calculating the trapezoidal numbers, which are presented in Table. 1.

Table 1. X-coordinates for the trapezoidal membership function

X-coordinate	Fuzzy subset		
	"Low"	"Average"	"High"
a_1	0	3,2	4,33
a_2	0	3,6	4,67
a_3	3,2	4,33	∞
a_4	3,6	4,67	∞

The analytical and graphical view of the fuzzy classifier is presented in Fig. 3.

$$\mu^{1}_{12}(x) = \begin{cases} 1, 0 \le x \le 3.2 \\ -2.5x + 9, 3.2 \le x \le 3.6 \\ 0, x \ge 3.6 \end{cases}$$

$$\mu^{2}_{12}(x) = \begin{cases} 0, x \le 3.2 \\ 2.5x - 8, 3.2 \le x \le 3.6 \\ 1, 3.6 \le x \le 4.33 \\ -3.03x + 14.15, 4.33 \le x \le 4.67 \\ 0, x \ge 4.67 \end{cases}$$

$$\mu^{3}_{12}(x) = \begin{cases} 1, x \ge 4.67 \\ 3.03x - 13.12, 4.3 \le x \le 4.67 \\ 0, x \le 4.33 \end{cases}$$

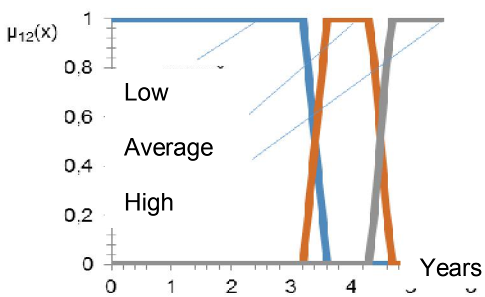

Figure. 3. Fuzzy classifier for the factor X_{12}

It should be noted that the second-level factors have a different impact on the credit rating of the borrower. For example, the increase in borrower's spending leads to a reduction of the borrower's credit rating. To eliminate this problem the axis inversion for the factor values was used, and then the high values of the factor will correspond to the low level, and vice versa. The results based on the axis inversion are shown in Table 2.

Similarly, as mentioned above, let us define the linguistic variable for the first-level indicators

$$B^{k}{}_{i} = \{x_i / \mu_*{}^{k}(x_i)\},\ \mu_*{}^{k}(x_i) \to [0,1],\ x_i \in [0,1]$$

$$i = 1..I,\ k = \begin{Bmatrix} 1 - \text{"}low\text{"} \\ 2 - \text{"}average\text{"} \\ 3 - \text{"}high\text{"} \end{Bmatrix}.$$

The transition from a quantitative description of the first level indicators to their qualitative description will be based on a standard three-level classifier (standard membership functions $\mu_*(x_i)$), proposed by the author.

Let us define the linguistic variable B_0 = *{«Credit Rating»}* with a term-set - *{"low", "medium", "high"}*. The indicator I_0 is also recognized by using the standard three-level classifier.

Determination of the borrower's credit rating: To assess the creditworthiness of the borrower quantitatively and qualitatively, it is necessary to produce aggregation of the data collected within the inference tree shown in Fig. 1. For the aggregation, Yager's OWA-operator can be successfully used [8, 9]. The quantitative value of the credit rating is determined by the following formula:

$$I_0 = \sum_{k=1}^{K} a_k \sum_{i=1}^{I} p_i \cdot \mu_{ik},$$

where μ_{ik} – the grade of membership of i-th factor value to the fuzzy subset B_{ik}; p_i - the weight of the i-th factor in aggregation; a_k – the nodal points of the standard three-level fuzzy classifier.

Table 2. Classification of the factors levels taking into account the inversion

X_{ij}	Fuzzy numbers B_{ij}:		
	"low"	"average"	"high"
X_{11}	(∞,∞,4,3)	(4,3,3,2)	(3,2,0,0)
X_{12}	(0,0,3.2,3.6)	(3.2,3.6,4.3,4.6)	(4.3,4.6,∞,∞)
X_{21}	(0,0,6000,9000)	(6000,9000, 13000,15000)	(13000,15000,∞,∞)
X_{22}	(0,0,7000,10000)	(7000,10000, 12000,15000)	(12000,15000,∞,∞)
X_{23}	(∞,∞,14000,10000)	(14000,10000, 8000,6000)	(8000,6000, 0,0)
X_{24}	(∞,∞,25,15)	(25,15,12,5)	(12,5,0,0)
X_{31}	(∞,∞,6500,5000)	(6500,5000, 4000,3000)	(4000,3000, 0,0)
X_{32}	(∞,∞,12000,8000)	(12000,8000, 7000,3500)	(7000,3500, 0,0)
X_{41}	(0,0,1,5,2)	(1,5,2,2,3)	(2,3,∞,∞)
X_{42}	(∞,∞,4,3)	(4,3,3,2)	(3,2,0,0)
X_{51}	(0,0,40000,75000)	(40000,75000, 80000,150000)	(80000,150000,∞,∞)
X_{52}	(0,0,90000,20000)	(90000,200000,250000,500000)	(250000,500000, ∞,∞)

In turn, to determine μ_{ik} let us aggregate factors within the second level

$$\mu_{ik} = \sum_{j=1}^{J} p_{ij} \cdot \mu_{(ij)k} ,$$

$$\mu_{(ij)k} = \mu_{ij}^{k}(x_{ij}),\ x_{ij} \in \left[a_{ij}, b_{ij}\right],$$

where $\mu_{(ij)k}$ – the grade of membership of ij-th factor value to the fuzzy subset $B^{k}_{(ij)}$; p_{ij} – the weight of the ij-th factor in aggregation.

The value of the private indicator X_i is defined as follows:

$$x_i = \sum_{k=1}^{K} a_k \sum_{j=1}^{J} p_{ij} \cdot \mu_{(ij)k} ,$$

$$\mu_{(ij)k} = \mu_{ij}^{k}(x_{ij}),\ x_{ij} \in \left[a_{ij}, b_{ij}\right].$$

It should be noted that the application of the mentioned above method of aggregation reduces the range of values of I_0 to the interval [0.2; 0.8]. This is due to the use of the nodal points in aggregation. For the further use of credit rating value for credit decisions, it is necessary to conduct the procedure of normalization.

4. Discussion

The proposed approach generates the quantitative and qualitative credit rating values. The qualitive one is used for the decision about whether to grant the loan. The decision may be based on the following principle: «the low level of credit rating - denial of credit; «the average credit – the further study of the borrower»; «the high level of creditworthiness - granting a credit». The quantitative value may be used for determining the credit conditions.

Thus, the article proposed the fuzzy model for the creditworthiness assessment. Model provides a qualitative and quantitative assessment of the creditworthiness of the borrower and can be a tool to support the credit management decisions in the area of the consumer lending. The problem of determining the credit conditions by using credit rating will be the further research area.

Corresponding Author:
Lukashevich Nikita
Business and Commerce Department
St. Petersburg State Polytechnical University
St. Petersburg, Russia
E-mail: mailto:lukashevich@kafedrapik.ru

References

1. Garanin D.A., Lukashevich N.S., Salkutsan S.V., 2014. The Evaluation of Credit Scoring Models Parameters Using Roc Curve Analysis. World Applied Sciences Journal, 30(8): 938-942. DOI: 10.5829/idosi.wasj.2014.30.08.265
2. Cole, R.H., 1988. Consumer and Commercial Credit Management, Boston: Irwin: 468.
3. Yap, B.W., S.H. Ong, and N.H.M. Husain, 2011. Using data mining to improve assessment of credit worthiness via credit scoring models. Expert Systems with Applications, 38(10): 13274-13283.
4. Hand, D.J. and W.E. Henley, 1997. Statistical classification methods in consumer credit scoring: a review. Journal of the Royal Statistical Society: Series A (Statistics in Society). 160(3): 523-541.
5. Batyrshin I., Wagenknecht M., 2002. Towards a Linguistic Description of Dependencies in Data. Int. J. Appl. Comput. Sci., 12(3).
6. Fishburn, P.C., 1986. The axioms of subjective probability. Stat. Sci., 1(3): 295-416.
7. Kaufmann A., Gupta M., 1991. Introduction to Fuzzy Arithmetic: Theory and Applications. Van Nostrand Reinhold.
8. Yager, R. R., 1996. Quantifier guided aggregation using OWA operators. International Journal of Intelligent Systems 11: 49-73.
9. Yager, R. R., 1993. Families of OWA operators. Fuzzy Sets and Systems, 59: 125-148.

10/2/2014