Assessing and Predicting Small Enterprises' Credit Ratings: a Multicriteria Approach

Baofeng Shi

May 15, 2022

Assessing and predicting small enterprises' credit ratings: A multicriteria approach

Baofeng Shi

1. College of Economics and Management, Northwest A&F University, Yangling, Shaanxi, 712100, China;
2. Research Center on Credit and Big Data Analytics, Northwest A&F University, Yangling, Shaanxi, 712100, China.

* Email: shibaofeng@nwsuaf.edu.cn.

ABSTRACT**:** Credit rating plays a crucial role in helping financial institutions make their lending decisions and in reducing the financial constraints of small enterprises. However, small enterprises have made it difficult for financial institutions such as commercial banks to determine their credit risk precisely, thus creating salient loan difficulties, because of the short duration, high frequency, urgent credit demand, and small amount of their loans. In an attempt to relieve the financing difficulty of small enterprises, this paper develops a new approach for small enterprises' credit risk assessment by combining high dimensional attribute reduction methods with fuzzy decision-making methods. Based on 687 small enterprises in a regional commercial bank of China, we find 17 indicators that have significant impact on the default risk of small enterprises. Then, it utilizes TOPSIS together with fuzzy C-means to grade the credit ratings of enterprises requesting loans. With the dual test of default discrimination and ROC curve, the prediction accuracy of the established indicator system has reached 85.40% and 90.09% respectively, indicating the strong default discrimination of this rating system and its practicability in commercial banks and other financial institutions.

Keywords**:** credit rating, default risk, fuzzy C-means, small enterprises

JEL: C61, E51, H81.

1 Introduction

China is the largest developing country in the world, and its small enterprises demonstrate great development momentum. According to statistics, in 2021, Chinese SMEs contribute more than 80% of national employment, 60% of the gross domestic product (GDP) (iResearch, 2021). Yet small enterprises have difficulty obtaining financing in general and loans in particular, severely restricting their development, because of unreliable financial information, loans of enormous volume but for low amounts, and diverse risks (Liang et al., 2007; Ciampi and Gordini, 2013; Shi et al., 2016; Chi and Zhang, 2017; Ruan et al., 2018; Sun et al., 2022). In an attempt to ease these financial difficulties, the State Council, the China Banking Regulatory Commission, and other agencies have required the banking industry to establish an “Inclusive Finance Business Division,” provide financial services to small and micro businesses, and address issues affecting agriculture, rural areas, and farmers, and strengthen the identification, monitoring, early warning, and assessment of borrowers’ credit risks (CBRC, 2015; SCPRC, 2016, 2017).

Many scholars have conducted useful studies on the best way to estimate the credit risk of loan-granting enterprises, in terms of the establishment of credit rating indicator systems, credit scoring, and other systems. Dimensionless processing of statistics is often needed before a rating indicator system can be established (Shi et al., 2015). In reality, the quantifiable financial data of small enterprises are less and more text-based non-financial data. Consequently, scholars often use subjective analytic hierarchy process (AHP) or the Delphi method to deal with data in a dimensionless fashion (Liang, 2007; Shi et al., 2018).

With regard to the construction of indicator system, Altman built Z-score and ZETA credit scoring models on the basis of indicators such as return on assets and pretax margins of asset interest to predict the possibility of lender default (Altaman, 1968; Altaman et al., 1977). Gu et al. (2017) combined (AHP) with data envelopment analysis (DEA), using indicators such as the cash ratio, inventory turnover, and accounts receivable turnover ratio from the perspective of financial status, credit status, enterprise development, and internet financial status to predict defaults by enterprises that take out loans. This research has great reference value for creating a credit rating indicator system for small enterprises, but little of it studies wholesale and retail enterprises and uses distinctive default variables to forecast the credit outlook of loan customers.

A credit scoring solution can be built using three methods: metrological statistics, artificial intelligence, and fuzzy evaluation. Metrological statistics include logistic regression, discriminant analysis, and linear regression (Reichert et al., 1983; Yurdakul and Iç, 2005; Iç and Yurdakul, 2010). Artificial intelligence consists of a support vector machine (Hens and Tiwari, 2012; Harris, 2015; Tomczak and Zieba, 2015; Abedin et al., 2018; Abedin et al., 2019), artificial neural nets (Marcano-Cedeño et al., 2011; Ala'Raj & Abbod, 2016; Rui & Mendes, 2017; Chi et al., 2017), a decision tree (Zhu and Hu, 2013; Florez-Lopez et al., 2015; Bahnsen et al., 2015; Zhang et al., 2017; Chern et al., 2021), ensemble learning (Abedin et al., 2022), and so forth. In recent years, some scholars have begun to combine fuzzy evaluation with these methods and subsequently devise a credit rating. Akkoç et al. (2012), combining artificial intelligence with fuzzy evaluation, built a credit rating model for a hybrid adaptive neuron fuzzy inference system in three stages and forecast the default risks of Turkish credit card holders. The empirical research shows that this model is better at correctly averaged classification and wrongly estimated classification cost than liner discriminant analysis, logistic regression, and artificial neural nets. Bai et al. (2019) calculate the default risks of farm lenders in a combined approach using fuzzy rough set and fuzzy C-means (FCM). This research focuses on the factors that influence loan customers' default, without grading their credit or including any decision function in their evaluation results.

To address this problem, some scholars have begun to divide consideration of credit ratings of loan customers into three credit rating models: scoring intervals of customer credit, establishing the threshold of default probability, and the loss given default (LGD) of loan customers. The Industrial and Commercial Bank of China (ICBC) (2005) divided the credit scores of its loan customers among 10 credit ratings into AA, AA-, so forth. Florez-Lopez (2007) calculated the probability of default (PD) of loan customers using metrological statistics and artificial intelligence and divided customers into five credit ratings based on dummy variables. Chi and Zhang (2017) employ nonparametric approaches to establish a credit rating model for small enterprises. They grade loan customers' credit ratings according to their LGD. Therefore, credit rating models based on scoring intervals of customers' credit give different results from models based on the threshold of default probability, so different loan approvers may give different credit rating results to loan customers with those credit scores. The reason is that scoring intervals and the threshold of default probability are given ahead of time, and this increases the subjectivity of ratings. With regard to the credit rating method based on LGD, a prerequisite is that the default loss of each customer must be known. But default loss data are unavailable for some small enterprises that have only recently applied for loans, which makes this rating method infeasible.

Through our literature review, we find that no existing research has a suitable rating indicator system to measure credit risk based on the loan characteristics of small wholesale and retail enterprises. In fact, the industry differences among small enterprises lead to obvious heterogeneity in their estimation of loan and credit risks. For example, the statistics on credit at commercial banks show that the average maximum value of single loans to small enterprises in real estate development and operations is as much as 17 million Yuan (approximately USD 2.50 million or 6.8 Yuan for each U.S. Dollar) and that of small wholesale and retail enterprises is only 0.41 million Yuan (Bank of Dalian, 2014). If these two types of enterprises are compared in the same credit rating system, even though the default false positive of the model is very low, they will create completely different losses for a bank. As a result, it is necessary to establish different credit rating models for diverse industries, based on their being small enterprises so as to distinguish their credit risk from that of other kinds of enterprises.

In view of the foregoing, this paper makes three contributions to the literature. First, in the category of credit rating, it adds to the literature by focusing on Chinese small wholesale and retail enterprises. Second, by establishing suitable credit rating models for small wholesale and retail enterprises, it offers a decision-making reference for credit rating by commercial banks, microcredit organizations, and those enterprises. Third, we use triangular fuzzy numbers in a dimensionless scoring process for non-financial data at small wholesale and retail enterprises, which helps to avoid the subjectivity and randomness caused by expertise scoring and makes the quantified processed qualitative indicator more accurate.

The paper is organized as follows. Section 2 establishes credit rating models for small enterprises. Section 3 builds the rating system based on credit data for 687 small wholesale and retail enterprises seeking loans from an urban commercial bank in China. Section 4 offers our main conclusion and lists the innovative aspects of this paper.

2 Methodology

In this section, we create credit rating models for small enterprises. To begin with, we establish an evaluation system based on the characteristics of loans for small wholesale and retail enterprises. Second, on basis of the indicator weights calculated by entropy weight, TOPSIS is employed to calculate credit scores for loan customers. Finally, we use fuzzy C-means to grade loan customers' credit ratings according to their credit scores. The framework can be seen in Figure 1.

2.1 Establishment of a Credit Rating System

Setting up this credit rating system involves two steps. First, initial data need to be standardized in order to eliminate the incompatibilities among dimensional indicators that are measured differently. Second, partial correlation analysis and probit regression are combined to create quantitative screening so as to reduce the number of indicators.

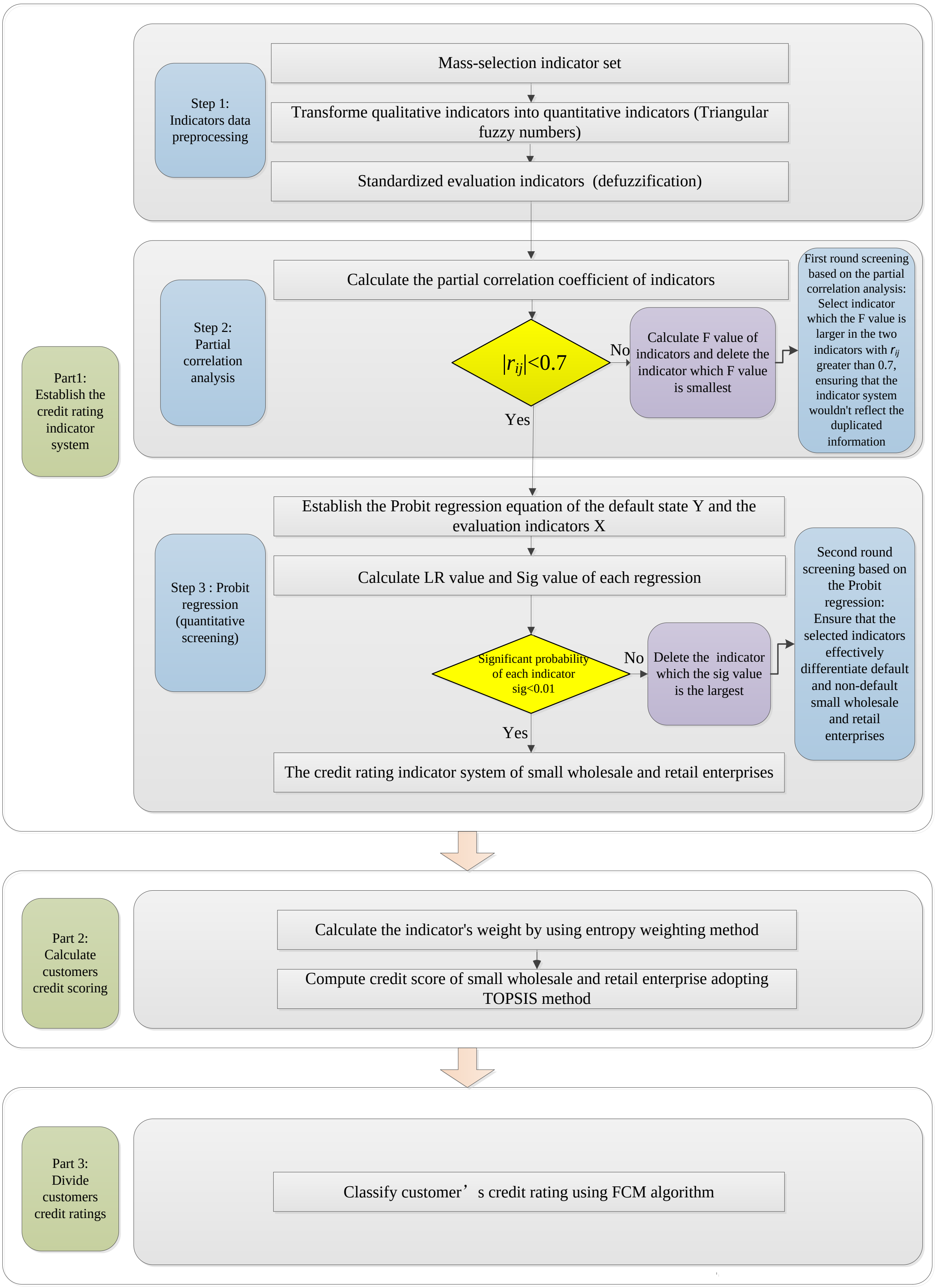

Figure 1. Framework of the credit rating model

2.1.1 Pre-Processing of Indicator Data

(1) Pre-Processing of Qualitative Indicator

Qualitative indicators cannot be directly quantified but, rather, are described narratively. For instance, the indicator for education background has five possible values: "Primary school diploma," "junior high school diploma," "senior high school diploma," "junior college diploma," and "bachelor's degree or above." Qualitative indicators have an advantage similar to that of triangular fuzzy numbers in how they process data with diverse characteristics. To quantify qualitative indicators, we need to turn them into triangular fuzzy numbers according to their semantics; then, we use defuzzification, that is, turning triangular fuzzy numbers into fixed values.

Suppose that A is a fuzzy set in a given domain; and for any $x \in U$, if there always exists a corresponding number $\mu_A(x) \in [0, 1]$, $\mu(x)$ is the membership of x to U and μ_A will act as the membership function of x. Let l and u be the lower limit and upper limit of the fuzzy numbers, and m be the most frequent value, then the array composed of the fuzzy numbers (l,m,u) can be depicted as in Figure 2, and its membership function μ_A is shown as equation (1) (Promentilla et al., 2008). Three-classification, five-classification, and seven-classification triangular fuzzy numbers are three commonly seen methods (Cheng et al., 2008; Khalili-Damghani et al., 2013; Wang et al., 2016), whose corresponding classification functions are illustrated in Figures 3 to 5 (Chai et al., 2019).

$$\mu_A(x) = \left\{ \begin{array}{ll} 0 & x < l, \\ \dfrac{x-l}{m-l} & l < x < m, \\ \dfrac{u-x}{u-m} & m \le x \le u, \\ 0 & x > u, \end{array} \right\} \tag{1}$$

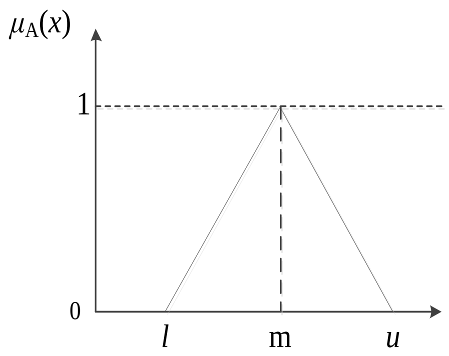

Figure 2. Triangular fuzzy numbers (TFNs)

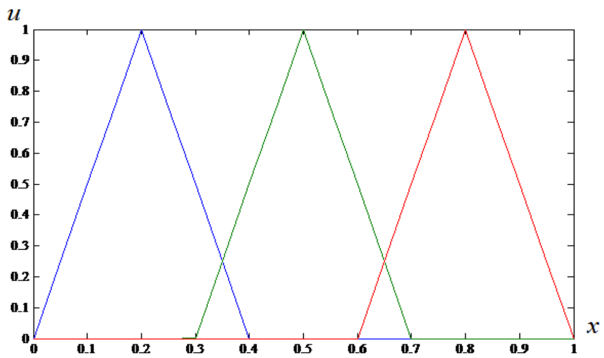

Figure 3. TFNs with three classifications

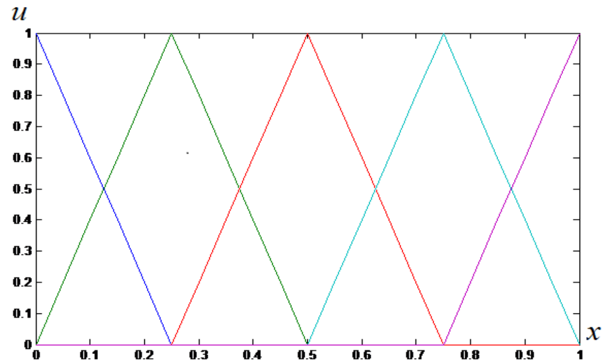

Figure 4. TFNs with five classifications

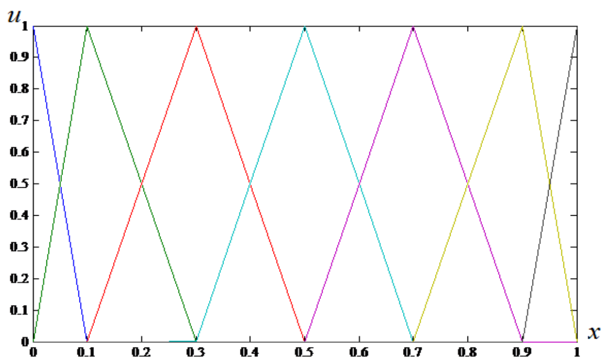

Figure 5. TFNs with seven classifications

Let A_{max} be the defuzzified value, and then with the combination of equation (1), we can obtain the result of A_{max} (Wu et al., 2016):

$$A_{max}=(l+m+u)/3 \tag{2}$$

(2) Pre-Processing of Quantitative Indicator

Quantitative indicators are usually divided into four categories: positive indicators, negative indicators, moderating indicators and interval indicators. We can use max-min standardization to process these indicators (Chi and Zhang, 2017; Shi et al., 2018; Abedin et al., 2019); to avoid repetition, it is not described here.

2.1.2 Reduction of Attributes

(1) The First Indicator Screening Based on Partial Correlation Analysis

In the same standard layer, partial correlation analysis (PCA) is employed to eliminate redundant indicators, guaranteeing that no information in indicators is repeated. Suppose that x_{ij} is the value of enterprise j in indicator i, x_{kj} is the value of enterprise j in indicator k, r_{ik} is the correlation coefficient between indicator i and indicator k, and r_{ik} is:

$$r_{ik}=\frac{\sum_{j=1}^{n}\left(x_{ij}-\bar{x}_i\right)(x_{ij}-\bar{x}_k)}{\sqrt{\sum_{j=1}^{n}\left(x_{ij}-\bar{x}_i\right)^2}\sqrt{\sum_{j=1}^{n}\left(x_{ij}-\bar{x}_k\right)^2}} \tag{3}$$

In this equation, n is the total number of small enterprises; $\overline{x}_i$ is the average value of indicator i; $\overline{x}_k$ is the average value of indicator k.

Let R be the correlation matrix composed of r_{ik}, the simple correlation coefficient of indicator i and indicator k, and m be the number of variables at the criterion level. The correlation matrix R is:

$$R=\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mm} \end{bmatrix} \tag{4}$$

The inverse matrix C of the correlation matrix R is:

$$C=R^{-1}=\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1m} \\ c_{21} & c_{22} & \cdots & c_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mm} \end{bmatrix} \tag{5}$$

According to the calculation equation for partial correlation coefficients, we obtain the partial correlation coefficient of indicator i and indicator k:

$$r'_{ik}=\frac{-c_{ik}}{\sqrt{c_{ii}c_{kk}}} \tag{6}$$

The larger the partial correlation coefficient r'_{ik} becomes, the stronger the relativity between indicator i and indicator k will be. When the absolute value of the partial correlation coefficient between x_i and x_k, i.e. $|r'_{ik}|>0.7$, F test (Nami and Shajari, 2018) is used to measure the evaluation of the two indicators on small enterprises' defaults respectively. Subsequently, the indicator with a smaller F value will be eliminated, which presents a weaker evaluation of defaults.

(2) The Second indicator Screening Based on Probit Regression

In the same standard layer, the maximum likelihood function is used to calculate the probit regression coefficients between m indicators and default y_j, and determine the LR statistics of each indicator. Using χ^2, we eliminate the indicator with the largest sig but that shows the least remarkable effects on defaults among indicator with a significance probability (Sig>0.01), and complete the screening of the first indicator. The remaining m-1 indicators, will be screened in the same manner as above until the corresponding significance probability of each indicator fails to exceed 0.01, i.e., Sig≤0.01. Then the indicator screening is done. Now, the remaining indicator can all significantly distinguish the defaults of small enterprises. The specific resolution equation is as follows.

Suppose that $\boldsymbol{X_j}=(x_{1j},x_{2j},\ldots,x_{mj})$ is the row vector of enterprise j made up of its all types of indicators; $\boldsymbol{\beta}=(\beta_0, \beta_1,\cdots, \beta_m)^T$ is the regression coefficient vector of indicators; m is the number of indicators; $P(Y_j=1)$ indicates the probability of default; $\varphi(z_j)$ is the standardized normal cumulative distribution function, and $z_j=\alpha+\boldsymbol{X_j\beta}$. Then,

$$P(Y_j=1)=\phi(z_j)=\int_{-\infty}^{z_j}\frac{1}{\sqrt{2\pi}}\ell^{(\frac{-s^2}{2})}ds \tag{7}$$

The maximum likelihood method is used to predict the indicators in the probit model, and its log-likelihood function is:

$$\max lnL=\sum_{j=1}^{n}\left[y_j\ln(\phi(z_j)+(1-y_j)\ln(1-\phi(z_j))\right] \tag{8}$$

In equation (8), the larger the log-likelihood function LnL appears, the more accurate the evaluation of default Y_j will be.

Let LR_k be the LR statistic value of indicator k, $\sigma_{\beta k}$ the standard error of regression coefficient β_k, $\tilde{\beta}_k$ the estimated parameter value of indicator k which is within constraints, $\hat{\sigma}_{\beta k}$ the standard error of the estimated parameter value of qualified indicator k, and $\hat{\beta}_k$ as well as $\hat{\sigma}_{\beta k}$ independently the estimated value and standard error beyond constraints. Then:

$$LR_k=-2[\log L(\overset{\sqcup}{\beta}_k,\overset{\sqcup}{\sigma}{}^2_{\beta k})-\log L(\hat{\beta}_k,\hat{\sigma}^2_{\beta k})] \tag{9}$$

2.2 Solution to Credit Scoring

Entropy weight is a method of describing information differences between indicators based on the information entropy of evaluated statistics; it has been widely applied in the determination of indicators in a complex systems evaluation (Chi and Zhang, 2017; Bai and Zhao, 2022). In this section, we use entropy to calculate the evaluation indicator weight $W = (w_i)$ in the first place; then we employ TOPSIS to calculate the credit scores of enterprises that wish to obtain loans (Yurdakul and Iç, 2005; Iç and Yurdakul, 2010; Wang and Leng, 2021). The specific procedure is as follows:

Step 1: Solve for the optimal and the worst scores of these indicators.

Suppose that b_i^+ is the optimal score of indicator i, b_i^- is the worst score of indicator i, and b_{ij} is the score of enterprise j in indicator i; so

$$b_i^+ = \begin{cases} \max(b_{ij}) & , i \text{ denote the i-th positive indicator} \\ \min(b_{ij}) & , i \text{ denote the i-th negative indicator} \end{cases}$$
$$b_i^- = \begin{cases} \min(b_{ij}) & , i \text{ denote the i-th positive indicator} \\ \max(b_{ij}) & , i \text{ denote the i-th negative indicator} \end{cases} \tag{10}$$

Step 2: Determine the standardized score and the difference between the optimal score and the worst score of these indicators. Let d_j^+ be the difference between the optimal score and the score of enterprise j, d_j^- be the difference between the worst score and the score enterprise j, and m be the number of small enterprise indicators. Then,

$$d_j^+ = \sqrt{\sum_{i=1}^{m} (w_i b_{ij} - w_i b_i^+)^2}$$
$$d_j^- = \sqrt{\sum_{i=1}^{m} (w_i b_{ij} - w_i b_i^-)^2} \tag{11}$$

Step 3: Independently solve for the relative closeness of the credit scores of small enterprises and that of the difference between the best and worst scores. Let c_j be the relative closeness of the score of enterprise j, and P_j be the score of enterprise j.

$$P_j = c_j = \frac{d_j^-}{d_j^- + d_j^+} \tag{12}$$

Step 4: The credit scores P_j of small enterprises in equation (12) are between 0 and 1, which are not consistent with the customary scoring regulations on a scale of 100. In view of this, we standardize P_j to render it in a period from 0 to 100.

$$S_j = \frac{P_j - \min(P_j)}{\max(P_j) - \min(P_j)} \times 100 \tag{13}$$

In this equation, S_j is the standardized credit score of enterprise j.

This paper employs default discrimination and an ROC curve to evaluate the predictive ability of the credit rating system for small enterprises as follows: if the credit score of a rating system meets the requirement that "all the credit scores of non-defaulting small enterprises are higher than those of small defaulting enterprises" the stronger the evaluation ability of the indicator system on the defaults of loan enterprises becomes, the fewer the losses of financial institutions such as banks. Accordingly, we follow Chi and Zhang (2017) in their method for evaluating the defaults of small enterprises and then determine the rationality of the indicator system in this paper.

$$S_c^1 = \frac{1}{m}\sum_{j=1}^{m} S_j^1 \tag{14}$$

$$S_c^0 = \frac{1}{n}\sum_{j=1}^{n} S_j^0 \tag{15}$$

$$s_c = \frac{\frac{1}{m}\sum_{j=1}^{m} S_j^1 + \frac{1}{n}\sum_{j=1}^{n} S_j^0}{2} \tag{16}$$

In equations (14)-(16), S_c^1 indicates the average value of the credit scores of defaulting samples; S_c^0 is the average value of the credit scores of non-defaulting samples; S_c=($S_c^1 + S_c^0$)/2.

ROC is a scientific quantized method of calibration first applied in signal detection, first used by Sobehart and Keenan (2001) to assess credit rating accuracy. First, the sensitivity and specificity of the credit rating indicator system for small enterprises are calculated. Given that the defaulting number correctly determined by a defaulting sample (y_j=1) is TP (true positive); the number erroneously determined by a defaulting sample is FN (false negative); the number correctly determined by a non-defaulting sample (y_j=0) is TN (true negative); and the number erroneously determined by a non-defaulting sample is FP (false positive). Equations for sensitivity and specificity are as follows:

$$\text{Sensitivity}=\text{TP}/(\text{TP}+\text{FN}) \tag{17}$$

$$\text{Specificity}=\text{TN}/(\text{FP}+\text{TN}) \tag{18}$$

Second, we use the two indicators of sensitivity and specificity to draw the ROC curve of the indicator system. The larger the area of AUC under the ROC curve appears, the stronger the selected indicator system's ability to recognize defaults will be.

2.3 Dividing Credit Ratings of Loan Customers

Compared with traditional clustering algorithms, fuzzy clustering algorithms do not strictly require each object to be identified as belonging to a certain class, demonstrating flexible attribute requirements. Thus it fits the special requirement that the initial indicator information not be a specific value but a value range in a triangular fuzzy function. Therefore, this paper follows Bai et al.'s (2019) fuzzy C-means (FCM) clustering algorithm, in rating the credit of small wholesale and retail enterprises. The principle is shown as Figure 6.

FCM associates each sample with all clusters through a real value. u_{ij}, the value of this vector, ranges from 0 to 1; it reflects the degree of membership of indicator *j* in category *i*. If the value of a given sample is close to 1, it means there is a strong correlation between this sample and a certain cluster; conversely, if its value is close to 0, it means there is a weak correlation between this sample and its corresponding cluster.

FCM divides *m* vectors $S_j(j=1,2,\cdots;m)$ into *c* fuzzy groups, and calculates the clustering center of each group, so as to minimize the non-similarity objective function. Its objective function $J(U,c_1,\cdots,c_c)$ (Yu et al., 2010) is:

$$J(U,c_1,\cdots,c_c)=\sum_{i=1}^{c}\sum_{j=1}^{m}(u_{ij})^n d^2(x_j,c_i) \tag{19}$$

In equation (19), c_i is the clustering center of category i; $d(S_j,c_i)$ is the Euclidean distance of the clustering center c_i in sample S_j; $n \in [1,\infty)$ is the weighting indicator, which controls the shared degree of categorized objects in the fuzzy category. This objective function refers to the weighted distance sum of squares from the sample points to the clustering centers in all categories.

Its structure is shown as the following objective function $\bar{J}(U,c_1,c_2\cdots,c_c,\lambda_1,\cdots\lambda_m)$ (Sun et al., 2022), which can help to calculate the necessary condition in which equation (19) is minimized.

$$\begin{aligned}\bar{J}(U,c_1,c_2\cdots,c_c,\lambda_1,\cdots\lambda_m) &= J(U,c_1,c_2\cdots,c_c)+\sum_{j=1}^{m}\lambda_j(\sum_{i=1}^{c}u_{ij}-1)\\ &=\sum_{i=1}^{c}\sum_{j=1}^{m}(u_{ij})^n d_{ij}^2+\sum_{j=1}^{m}\lambda_j(\sum_{i=1}^{c}u_{ij}-1)\end{aligned} \tag{20}$$

In this equation, λ_j is the Lagrange multiplier; c_i and u_{ij} are as follows (Demircan and Kahramanli, 2016):

$$c_i=\frac{\sum_{j=1}^{m}(u_{ij})^n S_j}{\sum_{j=1}^{m}(u_{ij})^n} \tag{21}$$

$$u_{ij}=\frac{1}{\sum_{k=1}^{c}(\frac{d_{ij}}{d_{kj}})^{\frac{2}{n-1}}} \tag{22}$$

Based on these two conditions, the basic steps in an FCM clustering algorithm are as follows:

(1) The number of clusters c is given, $1<c\le m$, and m is the number of samples. Given the maximum number of iterations is T,the threshold is ε, and the fuzzy number is ω; the indicator setting iterative counter t=0;

(2) Rectify partition matrix $U^{(t)}$ according to equation (20);

(3) Calculate the new clustering center $c_c(t)$ according to equation (19);

(4) $t\leftarrow t+1$; repeat steps 2 and 3 until $t\ge T$ or $\left|U^{(t)}-U^{(t-1)}\right|\le\varepsilon$.

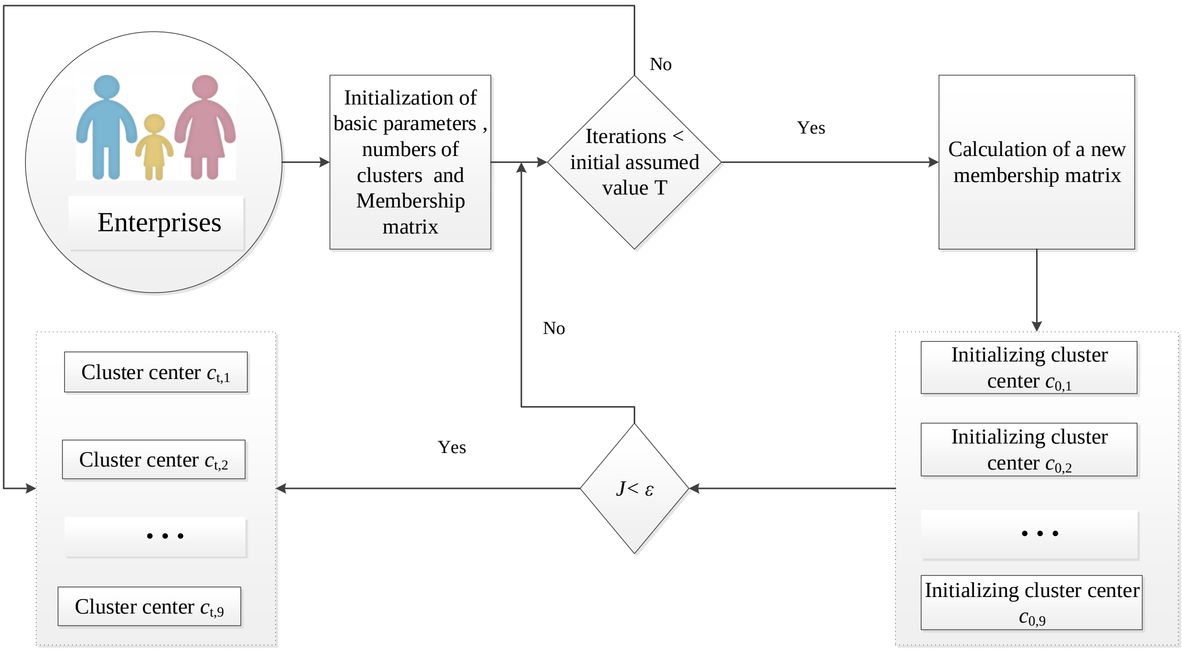

Figure 6. The framework for dividing credit ratings using the FCM method

3 Empirical Analysis

3.1 Sample Selection and Data Sources

This paper uses credit statistics on 687 small wholesale and retail enterprises that are customers of a commercial bank in a Chinese city, to test the model constructed in Section 2. Further details about the credit rating indicators and default status of these 687 small wholesale and retail enterprises are as follows. We select credit rating indicators for small wholesale and retail enterprises first by using the standard variables of ratings agencies such as Moody, Standard & Poor, and Fitch (Standard and Pool's Services, 2011; Fitch Rating, 2013; Dagong, 2010), and second from papers on credit rating (Mijid and Bernasek, 2013; Hai et al., 2013; Shi and Chi, 2014; Shi et al., 2016; Abedin et al., 2018, 2019; Sun et al., 2022). In this way, we select a total of 107 indicators on repayment ability and willingness to repay, and so forth. These indicators cover seven secondary standard layers such as financial factors, non-financial factors, and the personal situation of the legal representative of small wholesale and retail enterprises. Furthermore, we eliminate 26 indicators for which statistics are unavailable, leaving 81 indicators, as shown in Table 1.

Table 1. Screening criteria of credit rating indicators for small wholesale and retail enterprise

(1) No.	(2) 1st Criterion level	(3) 2nd Criterion level	(4) 3rd Criterion level	(5) Indicators	(6) Type	(7) Screening Result
1	Repayment Ability	Financial Factors	Solvency	Debt Asset ratio	Negative	Probit Delete
…				…	…	…
28				Source of repayment	Qualitative	Unobservable
…			…	…	…	…
55			Growth Capacity	Revenue growth	Positive	Pass
…				…	…	…
63				Wages, welfare growth rate	Positive	Unobservable
64		External Macroeconomic Conditions		Industry sentiment index	Positive	Pass
…				…	…	…
72				Economic environment	Qualitative	Unobservable
73		Internal Non-financial Factors		Years of relevant industry	Qualitative	Probit Delete
…				…	…	…
86		Legal Person Situation		Education background	Qualitative	Pass
…				…	…	…
98				Owner qualities	Qualitative	Unobservable
99	Willingness to repay	Enterprise Credit Situation		Registered capital classification	Qualitative	Partial Correlation Analysis Delete
…				…	…	…
103		Commercial Reputation		Tax records	Qualitative	Partial Correlation Analysis Delete
104				Legal disputes	Qualitative	Probit Delete
…				…	…	…
106				No. of breaches of contract	Qualitative	Probit Delete
107		Pledge guarantee Factor		Mortgage / pledge / guarantee	Qualitative	Probit Delete

3.2 Credit Rating of Small Wholesale and Retail Enterprises

(1) Establishment of Credit Risk Evaluation Indicator System

The pre-processing method for indicator information in section 2.1.1 is employed here to standardize the original data on the 687 small wholesale and retail enterprises in Table 2. The results are shown in Table 3.

Taking C1 enterprise's internal non-financial factors as an example, this paper illustrate the process of partial deleting correlation indicator (see Table 3). We put data on nine indicators related to “internal non-financial factors at enterprise C_1” in Table 3 into equations (3)-(6), so as to calculate r_{kj}, the partial correlation coefficient of the indicators. We respectively calculate the F-statistic of the indicator pairs whose partial correlation coefficients are over 0.7. Then we delete an indicator with a smaller F-statistic and retain the other one. The result is shown in Table 4. The rest can be done in the same manner. Using PCA, this paper deletes 14 indicators with redundant information.

Table 2. Original data for sample of small wholesale and retail enterprises

(a) No.	(b) Criterion level	(c) Indicators	Original Data 681 non-defaulting enterprises (1) C001	…	C681	6 defaulting enterprises (682) C682	…	(687) C687
1	C_1 Internal Non-financial Factors	X_1 Years of relevant industry	8	…	10	8	…	10
…		…	…	…	…	…	…	…
10		X_{10} Education background	Junior diploma	…	Bachelor's degree	N/A	…	Bachelor's degree
…	C_2 Legal Person Situation	…	…		…	…	…	…
20		X_{20} The value of car and real estate of legal representatives	1,000	…	1,000	N/A	…	100
21	C_3 Enterprise Credit Situation	X_{21} Registered capital classification	Found	…	Found	0.917	…	0.917
…	...	…	…	…	…	…	…	…
27	C_5 Operating Capacity	X_{27} Accounts receivable turnover rate	5.00	…	13.19	0	…	9.17
…		…	…	…	…	…	…	…
36		X_{36} Cash conversion cycle	-3,973.69	…	7.50	N/A	…	2.72
37	C_6 Profitability	X_{37} Rate of Return on Common Stockholders' Equity	0.078	…	0.003	0.000	…	0.280
…		…	…	…	…	…	…	…
49		X_{49} Operating activities generate cash inflows	112,458,001	…	625,800,630	0.000	…	26,139,847.75
50	C_7 Growth Capacity	X_{50} Operating income growth rate	0.000	…	0.023	0.00	…	1.36
…		…	…	…	…	…	…	…
54		X_{54} Retained revenue growth rate	0.076	…	1.251	0.510	…	0.507
55	C_8 Solvency	X_{55} Debt Asset ratio	6.84	…	0.56	0	…	0.604
…		…	…	…	…	…	…	…
74		X_{74} EBITDA / total debt ratio	0.043	…	0.003	-0.04	…	0.49
75	C_9 External Macroeconomic Conditions	X_{75} Industry sentiment index	137.45	…	139.50	137.45	…	127.20
…		…	…	…	…	…	…	…
80		X_{80} Engel coefficient	39.4	…	37.0	39.40	…	37.90
81	C_{10} Pledge guarantee factor	X_{81} Mortgage / pledge / guarantee	The guarantee amount is 5 million yuan	…	No guarantee	The guarantee amount is 18.9 million yuan	…	The guarantee amount is 3 million yuan
82	——	Default	0	…	0	1	…	1

Table 3. Standardized data on a sample of small wholesale and retail enterprise

(*a*) No.	(*b*) Criterion level	(*c*) Indicator	Standardized Data 681 non-default enterprises			6 default enterprises		
			C001	…	C681	C682	…	C687
1	C_1 Internal Non-financial Factors	X_1 Years of relevant industry	0.917	…	0.917	0.917	…	0.083
…		…	…	…	…	…	…	…
9		X_9 The proportion of total amount of money returned by enterprises through the bank	0.667	…	1.000	0.000	…	0.000
…	…	…	…	…	…	…	…	…
81	C_{10} Pledge guarantee factor	X_{81} Mortgage / pledge / guarantee	0.650	…	0.000	0.000	…	0.700
82	——	Default	0	…	0	1	…	1

Table 4. Partial correlation deletion indicator related to "Internal non-financial factors"

(1) No.	Indicators with a partial correlation coefficient greater than 0.7 (2) Indicator 1	(3) F statistic of indicator 1	(4) Indicator 2	(5) F statistic of indicator 2	(6) Partial correlation coefficient	(7) Deleted indicator
1	X_{55} Debt Asset ratio	2.370	X_{63} Shareholder equity ratio	2.392	0.993	X_{55} Debt Asset ratio
2	X_{56} Current liabilities operating ratio	1.284	X_{73} Total debt operating activity net cash flow ratio	0.907	0.967	X_{73} Total debt operating activity net cash flow ratio
3	X_{57} Quick Ratio	0.079	X_{68} Cash ratio	0.753	0.809	X_{68} Cash ratio

After deleting some indicators s with PCA, we screen the remaining indicators in all standard layers through probit regression, and select the indicators with remarkable discriminatory power on defaulting status. Then we put the 67 remaining indicator data screened by partial correlation in Table 3 into equations (7)-(9) and screen them using Stata. The 17 remaining screened indicators are in Table 5.

Table 5. Credit indicators weights for small wholesale and retail enterprises

(a) No.	(b) Indicators	(*c*) Weight	Standardized data (1) C001	…	(687) C687
1	X_{10} Education background	0.025	0.500	…	0.700
2	X_{13} Gender	0.003	1.000	…	1.000
3	X_{14} Age	0.006	0.970	…	0.848
4	X_{18} Family monthly income	0.172	0.071	…	0.071
5	X_{19} Time in Current position	0.047	0.250	…	0.250
6	X_{20} The value of car and real estate of legal representatives	0.095	0.917	…	0.917
7	X_{31} Fix capital ratio	0.197	0.003	…	0.029
8	X_{50} Operating income growth rate	0.033	0.197	…	0.201
9	X_{51} profit growth rate	0.001	0.494	…	0.530
10	X_{52} Total asset growth rate	0.027	0.271	…	0.298
11	X_{53} Capital accumulation rate	0.001	0.496	…	0.496
12	X_{54} Retained revenue growth rate	0.017	0.510	…	0.518
13	X_{75}Industry sentiment index	0.001	0.633	…	0.833
14	X_{77} Per capita disposable income of urban and rural residents at the end of the year	0.001	0.300	…	0.002
15	X_{78} Residential price index	0.000	0.817	…	0.988
16	X_{79} Per capita disposable income of urban residents	0.007	0.155	…	1.000
17	X_{80} Engel coefficient	0.001	0.576	…	0.821

(2) Solution to Credit Scoring of Small Wholesale and Retail Enterprises

The weight of 17 variables is calculated by entropy weight in Table 5. With equations (10)-(13), it is easy to obtain the credit scores of 687 small wholesale and retail enterprises. The result is shown in Table 6.

Then, we put the credit scores of these enterprises into equations (14) to (16) and subsequently find that the prediction accuracy of this credit rating system is 85.40%. The result of the model classification by equations (17) and (18) is shown in Table 7, and its corresponding ROC curve is shown in Figure 7, where area under curve (AUC) is 0.909, indicating the strong predictive accuracy of the defaulting status of small wholesale and retail enterprises by the screened 17 indicators.

Table 6. Credit scoring of small wholesale and retail enterprises

(1) No.	(2) Loan No.	(3) Original credit score P_j	(4) Standardized credit score S_j
1	200410270004	0.391	48.846
2	200412150123	0.243	0.759
…	…	…	…
687	X2012060800099	0.453	89.149

Table 7. Discriminant results of credit rating indicator system

Actual default status	Model prediction result 1 (Default)	0 (Non-default)	Sum
1 (Default)	4	2	6
0 (Non-default)	96	585	681
Sum	100	587	687

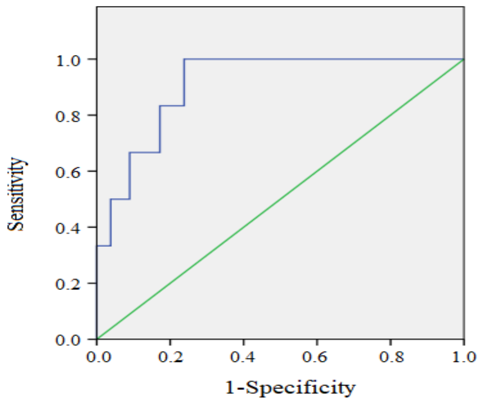

Figure 7. ROC curve (AUC=0.909)

(3) Crediting Rating of Small Wholesale and Retail Enterprises

According to the credit rating procedure, first, we program the clustering number c of the credit rating as 9; the largest iteration T = 1,000; the fuzzy number ω=2 (Zhong et al., 2014); the threshold ε=1E-5 (Robillard et al. 2014). Then, we import vector S_j, the credit scores of our sample of small wholesale and retail enterprises, into MATLAB, and obtain the corresponding data distribution and the classification of nine cluster centers, shown in Figures 8 and 9; the changing trend in their corresponding objective function is illustrated in Figure 10. Finally, we put the credit scores of cluster center in Table 8 in order from high to low and obtain nine corresponding ratings (AAA, AA, …to C). Using the upper and lower limits of the credit scores of samples in different clusters, it is easy to obtain the credit score interval for customers in different clusters and the corresponding sample frequency. The result is illustrated in Table 8.

Table 8. The credit rating for sample of small wholesale and retail enterprises

(1) No.	(2) Cluster center of credit score	(3) Credit rating	(4) Credit score interval	(5) Number of cases
1	85.497	AAA	[80.447, 100]	32
2	74.423	AA	[71.347, 80.447)	60
3	68.251	A	[65.264, 71.347)	54
4	62.147	BBB	[59.232, 65.264)	68
5	56.153	BB	[53.468, 59.232)	120
6	50.746	B	[47.179, 53.468)	73
7	43.464	CCC	[39.083, 47.179)	79
8	34.279	CC	[27.826, 39.083)	68
9	19.883	C	[0, 27.826)	124

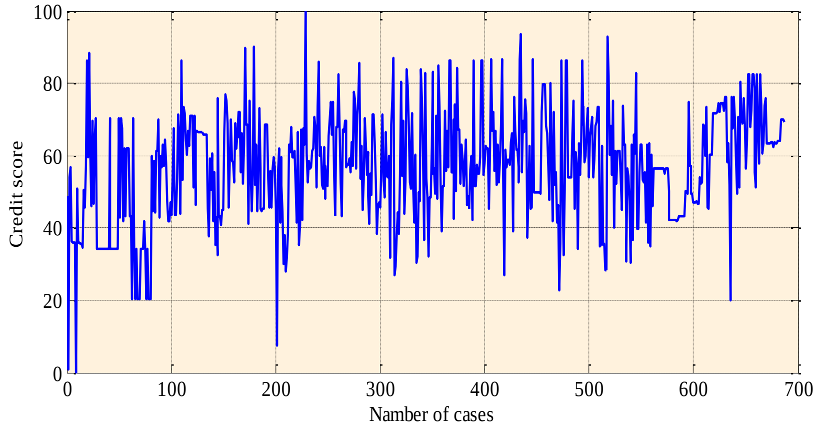

Figure 8. Distribution of credit score data for 687 small wholesale and retail enterprises

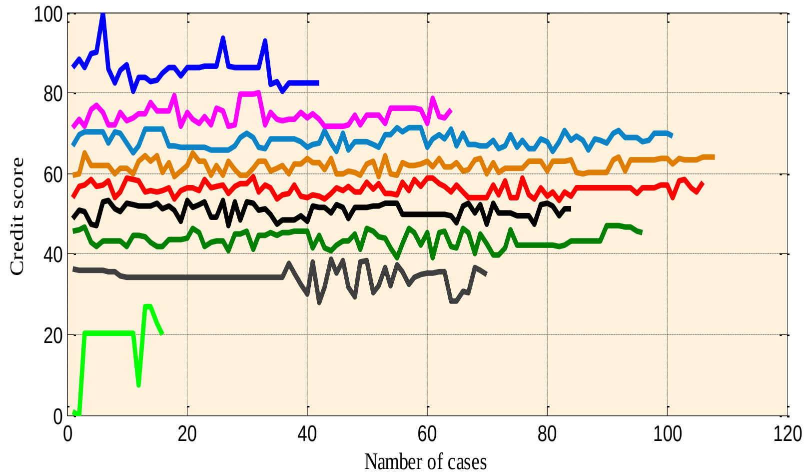

Figure 9. The classification of nine cluster centers

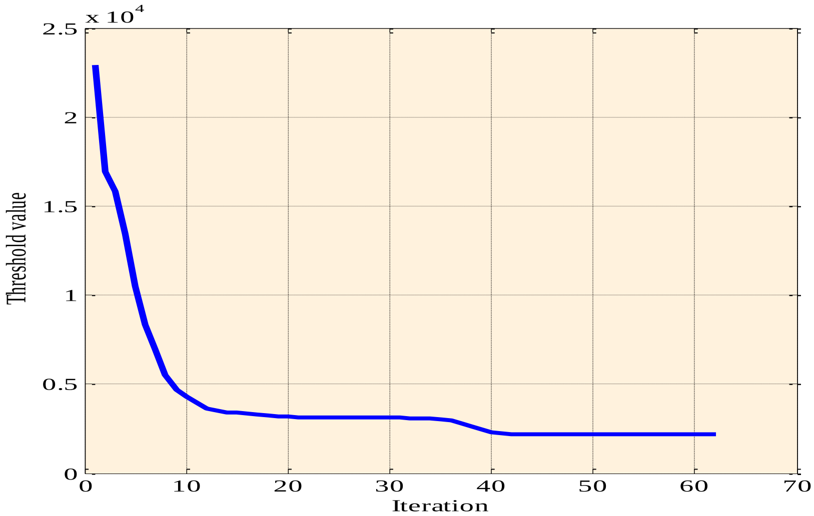

Figure 10. Credit rating division objective function change trend

4 Conclusion

Small enterprises have been central to China's economic development. However, because of imperfect financial information, urgent demand for loans but small amount of loan business, dispersed risks, and the absence of necessary guarantees, small enterprises have made it difficult for financial institutions such as commercial banks to depict their credit risks precisely, thus bringing about salient loan difficulties in terms of financing and high loan prices. This paper uses a sample of 687 small wholesale and retail enterprises to establish a credit rating system for such enterprises with a combination of metrological statistics and fuzzy decision. To begin with, we use partial correlation analysis to eliminate the indicators with repeated information and Probit regression to screen indicators that markedly influence the defaulting status of small wholesale and retail enterprises, establishing a credit risk evaluation indicator system composed of 17 indicators such as "X_{18} family monthly income" and "X_{20} the value of car and real estate of legal representatives" for these enterprises. Second, we calculate the credit scores of loan enterprises through the entropy-weighting TOPSIS method. Finally, we employ a fuzzy C-means (FCM) algorithm to grade the credit ratings of small wholesale and retail enterprises. The established indicator system, through defaulting state discrimination and testing the ROC curve, displays predictive accuracy of 85.40% and 90.09% respectively, demonstrating that the system has a relatively high default predictive ability, which can be used in applications at commercial banks and other financial institutions.

This paper is innovative in the following three respects. First, this paper establishes a credit rating system in accordance with the loan characteristics of small wholesale and retail enterprises. It is a useful supplement to the existing credit rating literature and can act as a decision-making reference for commercial banks and small wholesale and retail enterprises in their credit rating. Second, this paper introduces triangular fuzzy numbers into the dimensionless scoring process of non-financial statistics, avoiding the subjective arbitrariness in the present quantitative evaluation of qualitative indicators. Third, our empirical research shows that, for small wholesale and retail enterprises, non-financial factors are more capable of predicting default risks than financial factors. According to Figure 5, among the 17 influential rating indicators, the sum of the weights of non-financial factors and external micro indicators is 0.752, which is much higher than 0.248, the weight of internal financial indicators. Thus it can be seen that the non-financial factors and external micro conditions play a more important role in influencing the credit rating of small wholesale and retail enterprises; more attention should be paid to non-financial factors in terms of the prediction of small enterprises' default.

Although this paper makes progress in devising a credit rating system for small wholesale and retail enterprises, some limitations remain. Because it is difficult to obtain the real default loss data on loan enterprises, this paper only uses default status y_i as a dependent variable. This rating method has difficulty in explaining the objective reality that two different customers who default at the same time cause different losses to the same bank. With the accumulation of default data and the advance of data analysis technology, further breakthroughs and research on these problems can be produced.

Acknowledgements

The study was supported by the National Natural Science Foundation of China (Nos: 71873103, 72173096, 71503199 and 71731003), the Social Science Foundation of Shaanxi Province, China (No. 2018D51), the Tang Scholar Program of Northwest A&F University, China (No. 2021-04).

References

Abedin, M. Z., Chi, G. T., Colombage, S., & Moula, F. E., 2018. Credit default prediction using a support vector machine and a probabilistic neural network. *Journal of Credit Risk*, 14(2), 1-27.

Abedin, M. Z., Chi, G. T., Moula, F.E., Zhang, T., and Hassan. M. K., 2019. An optimized support vector machine intelligent technique using optimized feature selection methods: evidence from Chinese credit approval data. *Journal of Risk Model Validation*, 13(2): 1-46.

Abedin, M.Z., Chi, G.T., Hajek, P., Tong, Z., 2022. Combining Weighted SMOTE with Ensemble Learning for the Class-Imbalanced Prediction of Small Business Credit Risk. *Complex & Intelligent Systems*, Article in press, DOI: 10.1007/s40747-021-00614-4.

Akkoç, S. 2012. An Empirical Comparison of Conventional Techniques, Neural Networks and the Three-Stage Hybrid Adaptive Neuro Fuzzy Inference System (ANFIS) Model for Credit Scoring Analysis: The Case of Turkish Credit Card Data. *European Journal of Operational Research*, *222*(1): 168–78.

Ala'Raj, M., and Abbod, M. F. 2016. A New Hybrid Ensemble Credit Scoring Model Based on Classifiers Consensus System Approach. *Expert Systems with Applications*, *64*: 36–55.

Altman, E. I. 1968. Financial Ratios, Discriminant Analysis and the Prediction of

Corporate Bankruptcy. *Journal of Finance*, *23*(4): 589–609.

Altman, E. I.; Haldeman, R.; and Narayanan, P. 1977. ZETA Analysis: A New Model to Identify Bankruptcy Risk of Corporations. *Journal of Banking and Finance*, *1*:29–54.

Bahnsen, A. C.; Aouada, D.; and Ottersten, B. 2015. Example-Dependent Cost-Sensitive Decision Trees. *Expert Systems with Applications*, *42*(19), 6609–6619.

Bai, C.G., Shi, B.F., Liu, F., Joseph, S., 2019. Banking credit worthiness: evaluating the complex relationships. *Omega*, 83, 26-38.

Bai, X.P. and Zhao, Z.C. 2022. An Optimal Credit Scoring Model Based on the Maximum Default Identification Ability for Chinese Small Business. *Discrete Dynamics in Nature and Society*, Vol. 2022, Article ID 1551937.

Bank of Dalian. 2014. *Dalian Bank Small Business Credit System*. Dalian Bank.

Chai, N.N., Wu, B., Yang, W.W., Shi, B.F., 2019. A multicriteria approach for modeling small enterprise credit rating: Evidence from China. *Emerging Markets Finance and Trade*, 55(11), 2523-2543.

Cheng, J.; Feng, Y.; Tan, J.; and Wei, W. 2008. Optimization of Injection Mold Based on Fuzzy Moldability Evaluation. *Journal of Materials Processing Technology*, *208*(1), 222–28.

Chern, C.C., Lei, W.U., Huang, K.L. et al. 2021. A decision tree classifier for credit assessment problems in big data environments. *Information Systems and e-Business Management*, 19, 363–386.

Chi, G.T., Abedin, M.Z., Moula, F.E, 2017. Modeling Credit Approval Data with Neural Networks: An Experimental Investigation and Optimization. *Journal of Business Economics and Management*, *18*(2): 224-240.

Chi, G. T., and Zhang, Z. 2017. Multi Criteria Credit Rating Model for Small Enterprise Using a Nonparametric Method. *Sustainability*, *9*(10), 1-23.

China Banking Regulatory Commission (CBRC). 2015. *Supervision by Law, Supervision of the People, Risk Supervision: The China Banking Regulatory Commission implements the reform of the regulatory framework*.

http://www.cbrc.gov.cn/chinese/home/docView/67163D0D8293499BA499D2A9705C61CD.html.

China Report Hall. 2017. *2017–2022 China Enterprise Management Project Industry Market In-depth Research and Investment Strategy Research Analysis Report*. China Report Hall.

Ciampi, F., and Gordini, N. 2013. Small Enterprise Default Prediction Modeling through Artificial Neural Networks: An Empirical Analysis of Italian Small Enterprises. *Journal of Small Business Management*, *51*(1), 23–45.

Dagong Global Credit Rating Co.,. 2010. *Credit Rating Methodology Framework of Dagong Global Credit Rating Co., Ltd*. Dagong Global Credit Rating Co.

Demircan, S., and Kahramanli, H. 2016. Application of Fuzzy C-Means Clustering Algorithm to Spectral Features for Emotion Classification from Speech. *Neural Computing & Applications*, *29*(8), 1–8.

Fitch Ratings. 2013. *Fitch Ratings Global Corporate Finance 2012 Transition and Default Study*. Credit Market Research, Fitch Ratings.

Florez-Lopez, R. 2007. Modelling of Insurers' Rating Determinants: An Application of Machine Learning Techniques and Statistical Models. *European Journal of Operational Research*, *183*(3): 1488–1512.

Florez-Lopez, R., and Ramon-Jeronimo, J. M. 2015. Enhancing Accuracy and Interpretability of Ensemble Strategies in Credit Risk Assessment: A Correlated-Adjusted Decision Forest Proposal. *Expert Systems with Applications*, *42*(13): 5737–5753.

Gu, W.; Meheli, B.; Zhang, C; and Li, R. W. 2017. A Unified Framework for Credit Evaluation for Internet Finance Companies: Multi-Criteria Analysis through AHP and DEA. *International Journal of Information Technology & Decision Making*, *3*(3):597–624.

Hai, L.; Shi, B. F.; and Peng, G. 2013. A Credit Risk Evaluation Index System Establishment of Petty Loans for Farmers Based on Correlation Analysis and Significant Discriminant. *Journal of Software*, *8*(9), 2344–2351.

Harris, T. 2015. Credit Scoring Using the Clustered Support Vector Machine. *Expert*

Systems with Applications, *42*(2):741–50.

Hens, A. B., and Tiwari, M. K. 2012. Computational Time Reduction for Credit Scoring: An Integrated Approach Based on Support Vector Machine and Stratified Sampling Method. *Expert Systems with Applications*, *39*(8): 6774–6781.

İç, Y. T., and Yurdakul, M. 2010. Development of a Quick Credibility Scoring Decision Support System Using Fuzzy TOPSIS. *Expert Systems with Applications*, *37*(1):567–74.

Industrial and Commercial Bank of China. 2005. *Notice on Printing and Distributing the Measures for the Evaluation of Credit Ratings of Small Business Enterprises of Small Industrial Enterprises of Industrial and Commercial Bank of China*. Industrial and Commercial Bank of China, no. 78.

iResearch. 2021. *Report on Chinese Financing Development of Micro, Medium and Small Enterprises in 2021*. iResearch Inc., 2021.11.

Khalili-Damghani, K.; Sadi-Nezhad, S.; Lotfi, F. H.; and Tavana, M. 2013. A Hybrid Fuzzy Rule-Based Multi-Criteria Framework for Sustainable Project Portfolio Selection. *Information Sciences*, *220*(1), 442–62.

Liang, X. C.; Chen, S. F.; and Liu-Yan. 2007. The Study of Small Enterprises Credit Evaluation Based on Incremental AntClust. In *IEEE International Conference on Grey Systems and Intelligent Services* (pp. 294–98). IEEE Xplore.

Marcano-Cedeño, A.; Marin-De-La-Barcena, A.; Jimenez-Trillo, J.; Piñuela, J. A.; and Andina, D. 2011. Artificial Metaplasticity Neural Network Applied to Credit Scoring. *International Journal of Neural Systems*, *21*(4), 311–17.

Mijid, N., and Bernasek, A. 2013. Gender and the Credit Rationing of Small Businesses. *Social Science Journal*, *50*(1), 55–65.

Nami, S., and Shajari, M. 2018. Cost-Sensitive Payment Card Fraud Detection Based on Dynamic Random Forest and K-Nearest Neighbors. *Expert Systems with Applications*. 110: 381–92.

Promentilla, M. A.; Furuichi, T.; Ishii, K.; & Tanikawa, N. 2008. A Fuzzy Analytic

Network Process for Multi-Criteria Evaluation of Contaminated Site Remedial Countermeasures. *Journal of Environmental Management*, *88*(3), 479–95.

Reichert, A.; Chien-ChingCho; and Wagner, G. 1983. An Examination of the Conceptual Issues Involved in Developing Credit-Scoring Models. *Journal of Business & Economic Statistics*, *1*(2), 101–14.

Robillard, M. P.; Maalej, W.; Walker, R. J.; and Zimmermann, T. 2014. An Approach on Fault Detection in Diesel Engine by Using Symmetrical Polar Coordinates and Image Recognition. *Advances in Mechanical Engineering*, *2014*(4), 1–9.

Ruan, J. H.; Wang, Y.; Chan, F. T. S.; Hu, X.; Zhao, M.; Zhu, F.; Shi, B. F.; Shi, Y.; and Lin, F. 2018. A Life-Cycle Framework of Green IoT Based Agriculture and Its Finance, Operation and Management Issues. *IEEE Communications Magazine*, 2018, doi:10.1109/MCOM.2018.1800332

Rui, L., and Mendes, R. V. 2017. Detecting and Quantifying Ambiguity: A Neural Network Approach. *Soft Computing 22*(8), 1–9.

Shi, B. F., and Chi, G. T. 2014. A Model for Recognizing Key Factors and Applications Thereof to Engineering. *Mathematical Problems in Engineering*, *2014*(1), 368–81.

Shi, B. F.; Chen, N.; and Wang, J. 2016. A Credit Rating Model of Microfinance Based on Fuzzy Cluster Analysis and Fuzzy Pattern Recognition: Empirical Evidence from Chinese 2,157 Small Private Businesses. *Journal of Intelligent & Fuzzy Systems*, *31*(6), 3095–3102.

Shi, B. F.; Meng, B.; Yang, H. F., Wang, J., and Shi, W. L. 2018. A Novel Approach for Reducing Attributes and Its Application to Small Enterprise Financing Ability Evaluation. *Complexity*, (2018): 1–17. Article ID 1032643. doi:10.1155/2018/1032643

Shi, B. F.; Wang, J.; Qi, J.; and Cheng, Y. 2015. A Novel Imbalanced Data Classification Approach Based on Logistic Regression and Fisher Discriminant. *Mathematical Problems in Engineering*, *2015*(6), 1–12.

Sobehart, J., and Keenan, S. 2001. Measuring Default Accurately. *Risk, 14(3)*:

31–33.

Standard & Poor's Ratings Services. 2011. *S&P's Study of China's Top Corporates Highlights Their Significant Financial Risks*. New York: Standard & Poor's.

State Council of the People's Republic of China (SCPRC). 2016 *Notice of the State Council on Printing and Promoting Inclusive Financial Development Plan (2016–2020)*. http://www.gov.cn/zhengce/ Content/2016–01/15/content_10602.htm.

State Council of the People's Republic of China (SCPRC). 2017. *Notice of the General Office of the State Council on the Establishment of the "Made in China 2025" National Demonstration Zone*. http://www.gov.cn/zhengce/content/2017 11/23 /content_5241727.htm.

Sun, Y., Chai, N.N., Dong, Y.Z., Shi, B.F., 2022. Assessing and predicting small industrial enterprises' credit ratings: A fuzzy decision making approach. International Journal of Forecasting, Forthcoming.

Tao, Y.; Zhang, Y.; and Wang, Q. 2018. Fuzzy C-Mean Clustering-Based Decomposition with GA Optimizer for FSM Synthesis Targeting to Low Power. *Engineering Applications of Artificial Intelligence*, *68*:40–52.

Tomczak, J. M., and Zięba, M. 2015. Classification Restricted Boltzmann Machine for Comprehensible Credit Scoring Model. *Expert Systems with Applications*, *42*(4):1789–1796.

Wang, J., Ding, D., Liu, O. and Li, M. 2016. A Synthetic Method for Knowledge Management Performance Evaluation Based on Triangular Fuzzy Number and Group Support Systems. *Applied Soft Computing*, *39*: 11–20.

Wang, Y.T. and Leng, H. Y, 2021. Credit decision of SMEs based on Improved TOPSIS and decision tree. *2021 4th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE)*, 489-492, doi: 10.1109/AEMCSE51986.2021.00106.

Wu, Z., Ahmad, J., and Xu, J. 2016. A Group Decision Making Framework Based on Fuzzy VIKOR Approach for Machine Tool Selection with linguistic information. *Applied Soft Computing*, *42*:314–24.

Yu, F., Xu, H., Wang, L., and Zhou, X. 2010. *An Improved Automatic FCM Clustering Algorithm. International Workshop on Database Technology and Applications* (pp.1–4). IEEE. DOI: 10.1109/DBTA.2010.5659043

Yurdakul, M., and Iç, Y. T. 2015. Development of a Performance Measurement Model for Manufacturing Companies Using the AHP and TOPSIS Approaches. *International Journal of Production Research*, *43*(21): 4609–4641.

Zhang, J. H., Liu, H. Y., Zhu, R., and Liu, Y. 2017. Emergency Evacuation of Hazardous Chemical Accidents Based on Diffusion Simulation. *Complexity*, *2017*, 1–16.

Zhong, Y.; Zhang, L.; Xing, S.; Li, F.; and Wan, B. 2014. The Big Data Processing Algorithm for Water Environment Monitoring of the Three Gorges Reservoir Area. *Abstract and Applied Analysis*, *2014*(5), 1–7.

Zhu, P., and Hu, Q. 2013. Rule Extraction from Support Vector Machines Based on Consistent region covering reduction. *Knowledge-Based Systems*, *42*(2), 1–8.

Baofeng Shi is a Professor at College of Economics and Management, Northwest A&F University and the director of the Research Center on Credit and Big Data Analytics, Northwest A&F University, China. He received his doctoral degree in financial engineering in 2014 from Dalian University of Technology, China. He is interested in credit risk assessment, Fintech, Agriculture-related risk management, and rural finance. Dr. Shi has co-organised a number of international conferences on the themes of microfinance, rural finance, risk management and financial stability. He has published more than 50 research papers in peer reviewed journals, including *Nature Food*, *Omega-International Journal of Management Science*, *International Journal of*

Forecasting, *Annals of Operations and Research*, *Finance Research Letters*, and *Economic Modelling*, and a referee for more than 40 peer-reviewed journals.