Credit Allocation Considering Loaner's Credit Risk and Willingness for Acceptance: A Hybrid XGBoost-Topsis Enabled Optimization Approach

Genglin Zhu[1], Zixin Peng[1], Mu Li[1], Jiantao Fan[2], and Xinjun Lai[3](✉)

[1] School of Electro-Mechanical Engineering, Guangdong University of Technology, Guangzhou, China

[2] School of Economics and Commerce, Guangdong University of Technology, Guangzhou, China

[3] State Key Laboratory of Precision Electronic Manufacturing Technology and Equipment, Guangdong CIM Provincial Key Lab, School of Electro-Mechanical Engineering, Guangdong University of Technology, Guangzhou, China

xinjun.lai@gdut.edu

Abstract. Banks provide financial support for enterprises but may bear more risks lending to SMEs. This paper aims to reduce the credit risk of banks while maximizing their revenue. To provide banks with optimal credit strategy, this paper considers the accuracy of enterprise default risk. In this paper, banks screen out lending objects and enterprises choose whether to accept loans. The difficulty lies in that some enterprises lack credit records. Based upon the existing credit records, this paper uses XGBoost model to predict credit rating of these enterprises and evaluates the default risk through Topsis model. The solution results are in high accuracy and easy interpretation. To get the bank's optimal credit strategy, we obtained the data of non-credit record enterprises from the official network of Contemporary Undergraduate Mathematical Contest in Modeling. Starting from whether to consider the willingness, bank provides loans to 131 non-credit records enterprises. When considering the willingness, we find that bank can get an income of CNY 7.30189 million by providing a credit of CNY 96.44 million. Without considering it, the credit is CNY 49.27 million, and the income is CNY 5.10599 million. Therefore, the bank credit decision-making model considering the willingness to increases the income of CNY 2.1959 million.

Keywords: Credit Decision · Goal Programming · Topsis · Xgboost Classified Forecast

1 Introduction

Small and medium-sized enterprises (SMEs) play a crucial role in social services and national economic flowering [17]. They, however, face many difficulties in obtaining bank credit in light of their low anti-risk ability [12]. Therefore, banks often refuse to

N. Radojević et al. (Eds.): ICAID 2022, AHIS 7, pp. 985–995, 2023.
https://doi.org/10.2991/978-94-6463-010-7_99

lend. Constructing a reasonable risk assessment model provides an appraisal basis for bank's optimal credit decision, enables bank to obtain the highest return and promotes enterprises' growth.

Scholars have done a lot of research on credit classification, using traditional models like logical regression algorithms [15]. The advancement of technology paves the way for applying artificial intelligence algorithms like reinforcement learning and deep learning technology. SMEs' insufficient data, however, deprives them of deep learning technology. Machine learning algorithms like XGBoost, decision tree (DT), random forest (RF) and gradient boosting decision tree (GBDT) have lower requirements for the amount of data but have high accuracy in classification and regression. The XGBoost algorithm proposed by Chen and Guestrin [6] is an efficient boosting ensemble learning model based upon DT model. By studying the application of XGBoost method in credit evaluation, Huali et al. found that the XGBoost model has obvious advantages in feature selection and classification performance of personal loan scenarios [13]. Therefore, the XGBoost model can better predict and classify the credit rating of enterprises as loans of the small and medium-sized enterprises similar to personal loans.

Risk assessment plays an important role in risk control. Bolton employed the traditional logistics model [1]. Bryant developed an expert system [7], Yurdakull and Ic used the decision theory [8], Pavlenko and Chernyak applied the neural network model [14], Iazzolino employed the Data Envelopment Analysis method (DEA) [4], Zhang tested the Support Vector Machine (SVM) in credit score [18]. The Topsis model [5] proposed by Hwang and Yong has been widely used in evaluation research in light of its reliable results, easy understanding and interpretation.

Before that, many scholars used Topsis model to evaluate the size of enterprise credit. Yusuf [16] and others proposed a method based upon enterprise financial ratios and fuzzy Topsis model to determine the credibility of manufacturing enterprises for banks. Shen [3] proposed the extended intuitionistic fuzzy Topsis method with new distance measure. Roy [10] and others developed the credit score model of SMEs with Analytic hierarchy process and Topsis model. Shaw [11] and others used the fuzzy BWM and fuzzy Topsis to develop a multi-standard sustainable credit scoring system, which proved the model's practicability in credit scoring. This paper develops a credit evaluation model, which takes advantage of the high prediction accuracy of XGBoost model and the high interpretability of Topsis model.

Summarizing the existing literature of credit problems of SMEs, we found that the previous literature mainly focused on the prediction of anti-risk ability of enterprises, and seldom provided reasonable loans decision reference. This paper establishes an assessment model to evaluate the risk of SMEs and develops a planning model to help banks make credit decision. Through data verification, the model is more accurate in predicting anti-risk ability of enterprises and can provide a certain loans decision reference for banks.

2 Model Construction

This paper establishes an XGBoost classification prediction model based upon enterprises with existing credit records to predict the credit rating of non-credit record enterprises and to obtain the default risk level. After determining the default risk level, we establish the bank's optimal credit strategy. The specific solution is shown in Fig. 1.

2.1 Reputation Classification Prediction Model

The reputation of enterprises is crucial to their growth. In reality, banks will assess if enterprises can repay given their financial situation. After lending, whether be repaid will be retained in bank's records.

Using the existing credit records, we can obtain the original data set: corporate reputation $\mathrm{y_i}$, and the related feature variable $\mathrm{x_i} = \left(\mathrm{x_i^{(1)}}, \mathrm{x_i^{(2)}}, \ldots, \mathrm{x_i^{(v)}}\right)$, where v represents the number of features, i = 1, 2,…, n, and n is the number of samples, the prediction model takes the prediction of K-th iteration as the result. For the predicted reputation level of i-th enterprise, the predicted value is $\widehat{\mathrm{y_i}}$, that is:

$$\widehat{y_i} = \varphi(x_i) = \sum_{k=1}^{K} f_k(x_i) \tag{1}$$

The loss function in the training process of reputation grade prediction model is as follows:

$$\mathrm{Obj}^{(t)} = \sum_i L(y_i, \hat{y}_i) + \sum_k \Omega(f_k) \tag{2}$$

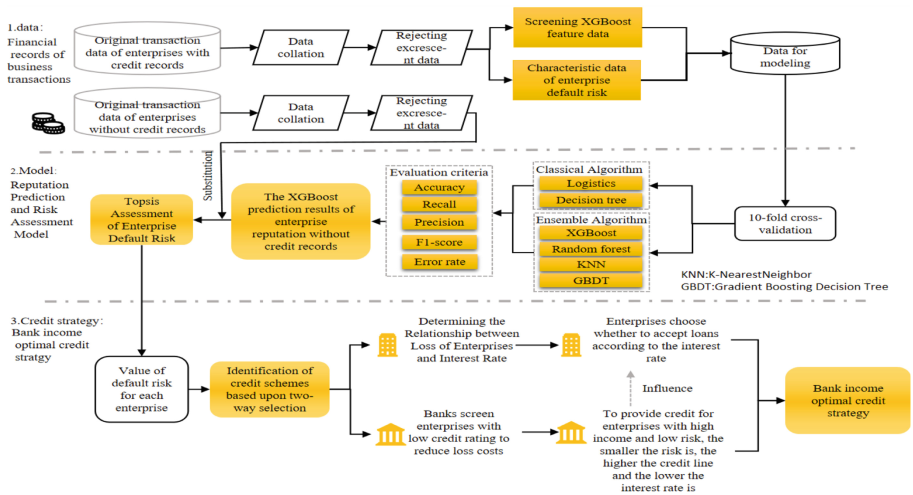

Fig. 1. Modeling process

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \|\theta_j\|^2 \tag{3}$$

In the equation, $\sum_i L(y_i, \hat{y}_i)$ denotes the loss function, and $\sum_k \Omega(f_k)$ denotes the regularization term. Where y_i represents the true value of enterprise reputation grade, $\widehat{y_i}$ represents the predicted value of enterprise reputation grade, T represents the number of leaf nodes, θ_j represents the leaf weight value, γ represents the leaf number penalty regular term, and λ represents the sub-weight penalty regular term [9].

Suppose that the prediction result of i-th enterprise in **t**-round iteration is $\hat{y}_i^{(t)}$ and $f_t(x_i)$ is the newly added regression tree, which is deduced as follows:

$$\begin{aligned} \hat{y}_i^{(0)} &= 0 \\ \hat{y}_i^{(1)} &= f_1(x_i) = \hat{y}_i^{(0)} + f_1(x_i) \\ &\vdots \\ \hat{y}_i^{(t)} &= f_k(x_i) = \hat{y}_i^{(t-1)} + f_i(x_i) \end{aligned} \tag{4}$$

Replace the results in (4) into (2), there are

$$Obj^{(t)} = \sum_{k=1}^{t} L\left(y_i, \hat{y}_i^{(t-1)} + f_k(x_i)\right) + \Omega(f_k) + C \tag{5}$$

Where C is a constant, the objective function is expanded by second-order Taylor, and the regular term is introduced:

$$Obj^{(t)} = \sum_{j=1}^{T}\left[G_j\theta_j + \frac{1}{2}(H_i + \lambda)\theta_j^2\right] + \gamma T \tag{6}$$

In the formula (6), $g_i = \partial_{\overline{y_i}(t-1)} L\left(y_i, \hat{y}_i^{(t-1)}\right)$, $h_i = \dot{\partial}^2_{\widehat{y_i}(t-1)} L\left(y_i, \widehat{y_i}^{(t-1)}\right)$, $G_i = \sum_{i \in I_j} g_i$, $H_i = \sum_{i \in I_j} h_i$, T represents the number of leaf nodes. In the formula, the leaf node θ_j is an uncertain value, so the first derivative of objective function $Obj^{(t)}$ to θ_j can find the optimal value of leaf node j, and plug the $\theta^*{}_j$ value into the objective function to obtain the minimum value of $Obj^{(t)}$:

$$Obj^{(t)} = -\frac{1}{2}\sum_{j=1}^{T}\frac{G_i}{H_i + \lambda} + \gamma T \tag{7}$$

2.2 Enterprise Risk Assessment Model

Given the comprehensive risk measurement of enterprises, we examine the credit, development potential, and development scale of enterprise, in which the credit is measured by the credit rating CP_i and the purchase order efficiency OV_i. The development potential is measured by the number of customers CN_i and the annual average profit rate EA_i.

The development scale is measured by the total enterprise orders TO_i and the annual average profit EAA_i.

According to the comprehensive evaluation model within the Topsis model, the original data is used to reflect the gap between the above indicators like "the best scheme" and "the worst scheme". The comprehensive evaluation indexes of strength and reputation are obtained to measure the risk each enterprise. The credit risk assessment matrix is established by Topsis model, which are represented by x_{ij}, that is:

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix} \tag{8}$$

Standardize the evaluation indicators of an enterprise:

$$z_{ij} = x_{ij} / \sqrt{\sum_{i=1}^{n} x_{ij}^2} \tag{9}$$

By using the objective function, the maximum and minimum values of each index are compared:

$$Z^+ = \left(Z_1^+, Z_2^+, \dots, Z_m^+\right) \tag{10}$$

$$Z^- = \left(Z_1^-, Z_2^-, \dots, Z_m^-\right) \tag{11}$$

Among them:

$$Z_i^+ = max\{z_{1i}, z_{2i}, \dots, z_{ni}\} \tag{12}$$

$$Z_i^- = min\{z_{1i}, z_{2i}, \dots, z_{ni}\} \tag{13}$$

Calculate distance between the comprehensive index and the maximum and minimum of credit risk:

$$D_i^+ = \sqrt{\sum_{j=1}^{m} w_j \left(Z_j^+ - z_{ij}\right)^2} \tag{14}$$

$$D_i^- = \sqrt{\sum_{j=1}^{m} w_j \left(Z_j^- - z_{ij}\right)^2} \tag{15}$$

Make a comprehensive evaluation of the credit risk:

$$S_i = \frac{D_i^-}{D_i^+ + D_i^-} \tag{16}$$

Next, use the entropy method to modify the probability matrix P, and the objective function of each element p_{ij} in the probability matrix P is as follows:

$$p_{ij} = \frac{z_{ij}}{\sum_{i=1}^{m} z_{ij}} \tag{17}$$

Where $\sum_{i=1}^{m} p_{ij} = 1$, calculate the information entropy of $\boldsymbol{j}$ index:

$$e_j = -\frac{1}{\ln n}\sum_{i=1}^{n} p_{ij}\ln(p_{ij})(j = 1, 2, 3, \ldots, m) \tag{18}$$

Calculate the information redundancy:

$$d_j = 1 - e_j \tag{19}$$

The effective information is normalized, and the entropy weight of each index is solved by using the entropy weight solving function:

$$w_j = \frac{d_j}{\sum_{j=1}^{m} d_{ij}} \tag{20}$$

2.3 Credit Line Allocation Model

This paper establishes a linear goal programming model hinged upon the results of risk assessment. Bank uses the established evaluation model to make credit decision and enterprises often decide whether to accept loans according to the interest rate. To quantitatively describe the relationship between the interest rate and the loss of enterprises, this paper expresses them through the function *F(r)*.

We set the loan term as one year, so the return is the product of the loan line and the interest rate. The objective function is:

$$max\, L = \sum_{i=1}^{n} m_{ik} \times r_i \times N_i \; k = 1, 2, 3 \tag{21}$$

Where m_{ik} represents the loan line of i-th enterprise with credit rating n. For convenience, we use 1, 2, and 3 to represent rating A, B, and C respectively; r_i indicates the interest rate of the i-th enterprise loan. N_i indicates whether the i-th enterprise is loaned. If N_i is 1, it accepts loans, and if N_i is 0, it denies.

We set the total credit line as CNY 100 million per year. All the enterprises studied in this paper are SMEs, so we set the loan between CNY 0.1 million and CNY 1 million, and the interest rate between 4% and 15%. Either enterprises accept the loan is related to the interest rate, so we determine a probability function for N_i. Grounded upon the function expression obtained by previous fitting curve, we randomly determine if enterprises accept loans to imitate the reality of enterprise choice. Therefore, the restrictions are determined as follows:

$$\text{s.t.}\left\{\begin{array}{c}\sum_{i=1}^{258} m_{ik}N_i \leq 10000 \\ N_i = f(1 - F_k(r_i)) \\ N_i = 0, 1 \\ 10 \leq m_{ik} \leq 100 \\ 0.04 \leq r_{ij} \leq 0.15 \\ S_{imax} \times m_{ikmax} \geq S_{imin} \times m_{ikmin} \\ S_{imax} \times r_{imax} \leq S_{imin} \times r_{imin}\end{array}\right. \quad (22)$$

Among them, S_{imax} and S_{imin} refer to enterprises with relatively low and high risk in enterprises with credit rating j; m_{ikmax} and m_{ikmin} refer to the loan line of these enterprises; r_{imax} and r_{imin} refer to the interest rate of these enterprises.

3 Case Analysis

In this paper, the original transaction data of 302 enterprises without credit records and 123 enterprises with credit records are obtained from the official website of Contemporary Undergraduate Mathematical Contest in Modeling. The first five characteristic variables affecting the enterprise reputation grade are whether they have made profits for three consecutive years. The impact of total sales amount, total number of sales, 2017 sales amount, 2016 profit, and whether continuously profitable to the reputation grade is 0.244, 0.224, 0.209, 0.163, and 0.160 respectively.

We randomly split the 123 enterprises, including 98 as training sets and 25 as test sets. By using the XGBoost classification prediction model, the prediction accuracy of test set reached 76%. To test if the XGBoost classification prediction model has obvious advantages in solving the credit rating prediction, we set up Logistics regression algorithm, Decision tree, Random forest, KNN and GBDT as comparative models. The prediction effect of each model is shown in Table 1.

Among them, Accuracy, the percentage of total sample where the prediction is correct, is the most commonly used evaluation index of classification problem. Error

Table 1. Model comparison.

Model	recall	Precision	F1-score	Error rate	Accuracy
XGBoost	0.76	0.818	0.747	0.24	0.76
Logistics	0.44	0.495	0.431	0.56	0.44
Decision tree	0.52	0.672	0.431	0.48	0.52
Random forest	0.6	0.624	0.607	0.40	0.60
KNN	0.52	0.484	0.495	0.48	0.52
GBDT	0.56	0.624	0.569	0.44	0.56

Table 2. Solving results

method	Total loan amount*	Lending income L*	Number of loan enterprises
Simulated annealing algorithm	96.44	7.30189	131
Genetic algorith	52.60	3.85908	112
Interior-point method	96.12	6.86109	134

* Unit: CNY million

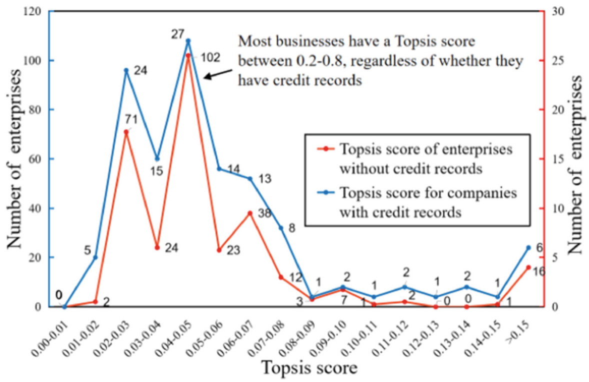

Fig. 2. Topsis results

rate means the percentage of total sample where the prediction is wrong. In the two-classification problem, we often divide the classification target into Positive and Negative. Precision is the probability of Positive predicted as Positive among all samples. Recall is the probability of Positive predicted as Positive among the original samples. F1 is the harmonic average of Precision and Recall. The higher the F1 is, the better the performance of the model is. From Table1, we see that the accuracy of XGBoost classification prediction model is 76%, significantly higher than other models. And the Precision, Recall and F1 of the XGBoost classification prediction model are also significantly better than other models. The results show that XGBoost has obvious advantages in reputation classification and prediction of SMEs.

Derived from the prediction of enterprise credit rating by XGBoost model, this paper uses Topsis model to obtain the possibility of enterprise default. We obtain the eigenvalues of each enterprise from packet sorting and plug them into the Topsis model to solve the results as shown in Fig. 2.

From the results, we see that the overall distribution of Topsis scores of enterprises is normal distribution, in line with the general law of reality. So the results have high credibility. We know that enterprises with credit rating D would default, so banks will not provide loans for them. The proportion of loans selected by enterprises at all levels varies with the interest rate, as shown in Fig. 3. The relationship tends to be a quadratic

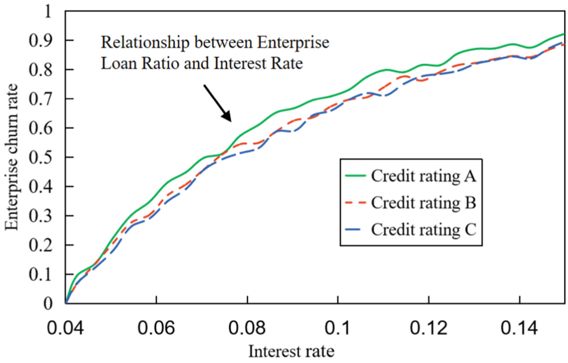

Fig. 3. Interest rate and enterprise churn rate

Table 3. Information on video and audio files that can accompany a manuscript submission.

ID	r_i	m_{ik}^*
124	0.100	56.72
125	0.118	57.04
126	0.076	76.69
...	...	...

* Unit: CNY ten thousand

function or a power function in Fig. 3. To quantitatively express the relationship, we use Curve Fitting Tool in Matlab and power function to fit the changes of loan ratios with the interest rate. The turnover function of an enterprise with a credit rating A is $F_A(r) = -0.1716r^{-0.6825} + 1.548$; the turnover function of an enterprise with a credit rating B is $F_B(r) = -0.2966r^{-0.546} + 1.721$; the turnover function of an enterprise with a credit rating C is $F_C(r) = -0.5104r^{-0.4510} + 2.045$. The residual squares and SSE of three fitting curves are 0.0045, 0.0028 and 0.0038 respectively, indicating that the fitting effect is excellent.

We know that enterprise with credit rating D has high default risk, so we exclude 44 enterprises with credit rating D, leaving 258 enterprises. Given the enterprise's turnover rate, we simulate the loan situation of enterprises under certain interest rate. We solved the objective programming model by simulated annealing algorithm, genetic algorithm and interior-point method. From the results in Table 2, we see that the simulated annealing algorithm has the highest expected income. Under the same constraints, the bank has the highest income and lowest cost. Then we can get the results under a credit of CNY 100 million, as shown in Table 3.

From the results in Table 4, we find that 131 enterprises accept loans. Without considering the willingness, the credit is CNY 49.27 million, with an income of CNY 5.10599 million. When considering it, the bank can increase income of CNY 2.1959 million. Therefore, the bank credit decision-making model established in this paper can significantly improve the income of banks.

Table 4. Information on video and audio files that can accompany a manuscript submission.

	Total loan amount*	Lending income*
Our approach	96.44	7.30189
Conventional	49.27	5.10599

* Unit: CNY million

4 Conclusion

To improve bank's income and reduce costs of loans to SMEs without credit records, this paper establishes a risk assessment model. Given the existing credit records of enterprises, we can predict the credit rating of enterprises without credit records, and the accuracy of XGBoost prediction model reaches 76%. According to the predicted reputation level, the Topsis model is used to evaluate default risk. The results show that the risk distribution of each enterprise is generally normal distribution, aligned with the general reality. To provide optimal credit strategy for banks, we establish a goal programming model rooted in the evaluation results. Then we discuss two situations. In the case of without considering the willingness, the bank gets an income of CNY 5.10599 million by providing a credit of CNY 49.27 million to 131 enterprises. When considering it, the bank can get an income of CNY 7.30189 million by providing CNY 96.44 million of credit. That is, the bank credit decision model considering the willingness can increase the income of CNY 2.1959 million.

Acknowledgment. The work described in this paper was jointly supported by the National Natural Science Foundation of Guangdong Province (No. 2021A1515012015), National Natural Science Foundation of China (No. 71601052), Hundred Young Talent Project of the Guangdong University of Technology (No. 220413637), and High-Level University Development for the Guangdong University of Technology (No. 262511006).

References

1. Bolton C et al (2010) Logistic regression and its application in credit scoring, Ph.D. dissertation, University of Pretoria
2. ShuPing C Contemporary undergraduate mathematical contest in modeling. http://www.mcm.edu.cn/
3. Shen F, Ma X, Li Z, Xu Z, Cai D (2018) An extended intuitionistic fuzzy topsis method based on a new distance measure with an application to credit risk evaluation. Inf Sci 428:105–119
4. Iazzolino G, Bruni ME, Beraldi P (2013) Using DEA and financial ratings for credit risk evaluation: an empirical analysis. Appl Econ Lett 20(14):1310–1317
5. Tzeng G-H, Huang J-J (2011) Multiple attribute decision making: methods and applications. CRC Press
6. Li H, Cao Y, Li S, Zhao J, Sun Y (2020) Xgboost model and its application to personal credit evaluation. IEEE Intell Syst 35(3):52–61

7. Bryant K (2001) ALEES: an agricultural loan evaluation expert system. Expert Syst Appl 21(2):75–85
8. Yurdakul M, İç YT (2004) AHP approach in the credit evaluation of the manufacturing firms in Turkey. Int J Prod Econ 88(3):269–289
9. Meng X, Zhang Z, Yang Y, Li G, Tao C, Niu Y (2021) Improved xgboost stray current prediction and interpretable model. Laser Optoelectron 1–13
10. Roy PK, Shaw K (2021) A credit scoring model for SMEs using AHP and Topsis. Int J Finance Econ
11. Roy PK, Shaw K (2021) Developing a multi-criteria sustainable credit score system using fuzzy BWM and fuzzy topsis. Environ Dev Sustain 1–32
12. Angilella S, Mazzù S (2015) The financing of innovative SMEs: a multicriteria credit rating model. Eur J Oper Res 244(2):540–554
13. Chen T, Guestrin C (2016) Xgboost: a scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, ser. KDD 2016, New York, NY, USA. Association for Computing Machinery, pp 785–794
14. Pavlenko T, Chernyak O (2010) Credit risk modeling using Bayesian networks. Int J Intell Syst 25(4):326–344
15. Zhang W, He H, Zhang S (2019) A novel multi-stage hybrid model with enhanced multi-population niche genetic algorithm: an application in credit scoring. Expert Syst Appl 121:221–232
16. İÇ YT, Yurdakul M (2010) Development of a quick credibility scoring decision support system using fuzzy topsis. Expert Syst Appl 37(1):567–574
17. Liu Z, Cruz JM (2012) Supply chain networks with corporate financial risks and trade credits under economic uncertainty. Int J Prod Econ 137(1):55–67
18. Zhang Z, Gao G, Shi Y (2014) Credit risk evaluation using multi-criteria optimization classifier with kernel, fuzzification and penalty factors. Eur J Oper Res 237(1):335–348