

Research Article

Multiview Graph Learning for Small- and Medium-Sized Enterprises' Credit Risk Assessment in Supply Chain Finance

Cong Wang,¹ Fangyue Yu,¹ Zaixu Zhang,¹ and Jian Zhang ²

¹School of Economics and Management, China University of Petroleum, Qingdao, China

²School of Government, Central University of Finance and Economics, Beijing, China

Correspondence should be addressed to Jian Zhang; zjpolicy@163.com

Received 25 December 2020; Revised 28 January 2021; Accepted 1 February 2021; Published 15 February 2021

Academic Editor: Baogui Xin

Copyright © 2021 Cong Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, supply chain finance (SCF) is exploited to solve the financing difficulties of small- and medium-sized enterprises (SMEs). SME credit risk assessment is a critical part in the SCF system. The diffusion of SME credit risk may cause serious consequences, leading the whole supply chain finance system unstable and insecure. Compared with traditional credit risk assessment models, the supply chain relationship, credit condition of SME, and core enterprises should all be considered to rate SME credit risk in SCF. Traditional methods mix all indicators from different index systems. They cannot give a quantitative result on how these index systems work. Furthermore, traditional credit risk assessment models are heavily dependent on the number of annotated SME data. However, it is implausible to accumulate enough credit risky SMEs in advance. In this paper, we propose an adaptive heterogenous multiview graph learning method to tackle the small sample size problem for SMEs' credit risk forecasting. Three graphs are constructed by using indicators from supply chain operation, SME financial indicator, and nonfinancial indicator individually. All the graphs are integrated in an adaptive manner, providing a quantitative explanation on how the three parts cooperate. The experimental analysis shows that the proposed method has good performance for determining whether SME is risky or nonrisky in SCF. From the perspective of SCF, SME financing ability is still the main factor to determine the credit risk of SME.

1. Introduction

As a result of the COVID-19 pandemic, global industrial chain and supply chain have suffered a severe setback. Industrial chains across the world are experiencing an unprecedented crisis, which has rarely been encountered nearly the century. The accompanying adjustment of supply chain and industrial chain has a great impact on small- and medium-sized enterprises (SMEs). The survival predicament faced by SMEs also arises in China. According to the National Bureau of Statistics, the SMEs in China played an essential role in the national economy, accounting for more than 99% of the total number of enterprises, more than 80% employed people in industrial enterprises, more than 70% of the national total technological innovation results, more than 60% of the GDP, and more than 50% of the national total tax revenue. However, the financing difficulties faced by

SMEs have always been the key issues for their development and survival.

Supply chain finance (SCF) is a series of financing modes designed to solve the SMEs financing problem [1], which integrates capital into supply chain management. With the core enterprise as the center and the real trade as the background, SCF transforms the uncontrollable risk of a single enterprise into controllable risk of the whole supply chain enterprise through effective control of capital flow, information flow, and logistics. It effectively builds a benign industrial ecology of banks, core enterprises, and SMEs, promoting the interactive development of capital and industry. In SCF, credit risk is regarded as one of the core issues, which must be considered seriously, especially that of SMEs in the upstream or downstream of core enterprises. Compared with banks and core enterprises, SMEs are more prone to encounter credit risk. In the SCF system, once the

credit risk occurs, the credit status of enterprises in the chain will be magnified and even spread to the whole supply chain due to the connectivity of the supply chain [2–4]. However, credit risk is inevitable. Therefore, it is necessary to establish a credit risk assessment model for SMEs to effectively control risks, forming a stable SCF system, both in the construction and operation process.

The SMEs' credit risk assessment system under the SCF environment has always been the focus of the academic world and the financial world [5]. The deep integration between core enterprises and SMEs in SCF can reflect the future capital capacity and cash flow of SMEs. This characteristic provides a solid data base for credit risk prediction in SCF. The academic world constructs the credit risk assessment index system from different perspectives. Constructing the credit risk assessment system based on different factors is one of the branches. Lekakos et al. considered core enterprise qualification as one of the main factors for SME credit evaluation [6]. Wuttke found that the buyer has an impact on supply chain finance [7]. These literatures investigated the correlation between SCF and SME credit risk theoretically. However, some indicators are neither difficult to obtain in practice nor to analyze quantitatively [8].

Another branch established the credit risk assessment system for SMEs in SCF from different perspectives. Altman and Sabato [5] designed an SME credit risk evaluation system, including five financial indicators such as liquidity, earnings, leverage, coverage ratio, and business activity ratio. However, only financial indicators are selected as the main features for SME credit risk assessment and other nonfinancial indicators are neglected. Yurdakul [9] combined both finance and nonfinance indicators to evaluate the SME credit risk in Turkey, including the revenue generation capacity, cost control, operational efficiency and profitability, short-term liquidity, capital structure, and other solvency indicators. In order to consider the factors synthetically on SME credit risk assessment in SCF, many other indicators were also involved in the system. From the perspective of supply chain, Rostamzadeh et al. [10] proposed to take operational and major policy risks into considerations. Mou et al. [11] constructed the evaluation system from four aspects: industrial status, operation status, asset status, and credit history. Among these systems, there may be overlapping or contradictory phenomena among the credit risk assessment indicators selected from different views. It should be emphasized that this issue is difficult to solve for most of the conventional prediction model. Furthermore, most of existing credit risk assessment systems focus on selecting single discriminative indicator (e.g., three-class indicator), without considering the overall discrimination of the whole perspectives (e.g., one or two-class indicator). They cannot give quantitative results about the relationships of the different views.

Recently, machine learning approaches have been widely applied as a substitute to traditional qualitative and statistical analysis methods on credit risk assessment field [12, 13]. Logistic regression (LR) is the basic method to assess credit risk for its simplicity and usability [14]. However, it is

insufficient to describe the complex situation of credit risk. The prediction accuracy needs further enhancement. Support vector machine [15] (SVM) is another approach to tackle credit risk problem. It has been proved that the SVM-based credit risk assessment model is effective and advantageous than the LR model based on principal component analysis (PCA) [16]. Danenas and Garsva identified risky enterprises in imbalanced datasets based on linear SVM and particle swarm optimization [17]. The authors of [18] applied a nonlinear SVM classifier with genetic optimization for credit risk assessment. The artificial neural network (ANN) has also attracted a wide range of attention in credit risk forecasting. Adnan made a comprehensive investigation of different supervised neural models and learning schemes for credit risk evaluation [19]. The deep belief network (DBN), as a special paradigm of ANN, has been found to yield the best performance on the CDS dataset [20]. In addition, tree-structured [21] and graph-structured [22] methods have also been shown to be effective on credit risk assessment.

Various approaches have shown their advantageous on credit risk evaluation. However, to assess the credit risks of the SMEs in SCF environment, we need to construct the model from a perspective of the supply chain, rather than only assessing repayment ability [23]. The funds and credit condition of SME, core enterprises, and supply chain relationship are to be involved in the machine learning algorithms. These information from different views need to get organized properly. Most of approaches concatenate the indicators together to evaluate SME credit risk in supply chain finance. By concatenating the indicators, Zhang et al. [23] and Xu and He [24] applied SVM and restricted Boltzmann machine (RBM), respectively, for assessment. From the perspective of feature fusion, these early fusion approaches treat all features in the same way, which weakens the internal connections among different views. Other literatures ensemble multiple machine learning models to make a comprehensive decision. Zhang et al. [25] considered the leading enterprise's credit status and the relationships developed in the supply chain by assembling SVM and BP. Zhu et al. [26] proposed a two-stage hybrid model by integrating the results of LR and ANN to boost the accuracy of the single model. Nevertheless, all methods mentioned above are heavily dependent on the annotated SME data tightly. Once the number of annotated samples decreases, the algorithm performance will decline rapidly. However, it seems implausible to accumulate enough credit risky SMEs in reality. Compared with the method above, graph-based learning can work with small number of annotated data by modeling the relationships of samples, which has been widely used in many other fields [27–29]. It is difficult to construct a graph to cover local manifolds of different view features. Therefore, how to integrate multiview heterogeneous features into graph learning is worth pondering.

In this paper, we propose a heterogeneous multiview graph learning method for SMEs' credit risk forecasting. The credit risky or nonrisky SME is recognized by diffusion process on the fused multiview manifolds. Considering different views having different impacts on credit risk recognition, we learn the view weights and credit risk scores

simultaneously. The main contributions of our work are summarized as follows:

- (1) Tackling the problem of lacking enough annotated data, we propose a multiview graph-based semi-supervised learning approach by modeling the local manifold between samples for SME credit risk assessment in SCF.
- (2) The proposed multiview graph-based learning approach integrates multiview features in a comprehensive way. The SME credit risk scores are achieved by learning the complementariness of different views automatically, which improves interpretability.
- (3) Comprehensive experiments are conducted to empirically analyze the proposed SME credit risk in the SCF method. The experimental results on the collected dataset demonstrate the effectiveness of the proposed method.

The remainder of this paper is organized as follows. In Section 2 and Section 3, we discuss the single graph learning and multiview graph learning. Section 4 describes the data and variables. Section 5 demonstrates the experimental results. The last section is the conclusion.

2. Single Graph-Based Learning for Classification

Graph-based learning [30, 31] has attracted great interests in classification tasks due to its effectiveness and flexibility to various areas. A graph describes the pairwise relationships based on the given data, where vertices are labeled and unlabeled samples and edges indicate the relationships of vertices. Generally, single graph-based learning can be formulated as follows.

Assume that there are N data points $X = \{x_1, x_2, \dots, x_N\}$ in the dataset. Each sample $x_i \in \mathfrak{R}^d$ can be represented by a d -dimensional feature vector. Without loss of generality, we assume the first N_l ($N_l \ll N$) samples are labeled from dataset $X_L = \{(x_1, y_1), (x_2, y_2), \dots, (x_{N_l}, y_{N_l})\}$ by experts in advance, where $y_i = \{-1, +1\}$ is the class label. Our goal is to assign an optimal label y_i for each of the rest $N - N_l$ unlabeled data points through graph learning.

A graph is defined as $G = (V, E, W)$. V is the set of vertices, where each sample x_i is a vertex v_i in G , including both labeled and unlabeled samples. $E = e_{ij}$, ($i = 1, 2, \dots, N; j = 1, 2, \dots, N$) is the set of edges, where e_{ij} links the adjacent vertices v_i and v_j . $W = w_{ij}$, ($i = 1, 2, \dots, N; j = 1, 2, \dots, N$) is termed as affinity matrix, where w_{ij} is the weight of e_{ij} , indicating the strength of vertices v_i and v_j . The value of w_{ij} can be either discrete (e.g., $\{0, 1\}$) or continuous (e.g., 0.54).

Generally, graph-based classification task is formulated in a regularization framework:

$$\arg \min_f \left\{ \Omega(f) + \lambda R_{\text{emp}}(f) \right\}, \quad (1)$$

where f is the to-be-learned relevance score vector defined on domain $(-1, 1)$, $R_{\text{emp}}(f)$ is an empirical loss function,

$\Omega(f)$ is a regularizer term on f , and λ is a nonnegative parameter.

$\Omega(f)$ implements the smoothness assumption on the graph by considering data points on the same manifold are likely to share the same label.

The regularizer on the graph is defined as follows:

$$\Omega(f) = \frac{1}{2} \sum_{i,j=1}^N W_{ij} \left(\frac{f_i}{\sqrt{D_{ii}}} - \frac{f_j}{\sqrt{D_{jj}}} \right)^2, \quad (2)$$

where D is a diagonal matrix given by $D_{ii} = \sum_j W_{ij}$, that is, D_{ii} is the sum of the i -th row of W .

Let $\Theta = D^{-1/2} W D^{-1/2}$, and the normalized graph Laplacian is denoted as follows:

$$L = I - \Theta, \quad (3)$$

here I is a unit diagonal matrix. Finally, the regularizer can be rewritten as follows:

$$\Omega(f) = f^T L f. \quad (4)$$

The empirical loss function $R_{\text{emp}}(f)$ conducts the consistency of labeled vertices by forcing the assigned labels close to the initial labels.

$$R_{\text{emp}}(f) = \sum_{i=1}^N (f_i - y_i)^2, \quad (5)$$

where y_i is the initial label vector. Three possible values $\{1, 0, -1\}$ can be assigned to y_i : 1 if vertex i is the positive sample, -1 if it is the negative sample, and 0 if it is unlabeled. The closed-form solution for the minimization is found to be

$$f = \left(I + \frac{1}{\lambda} L \right)^{-1} y. \quad (6)$$

After obtaining f , data point x_i can be classified according to its sign, i.e., positive if $f_i > 0$ and negative otherwise. In addition, the relative value of data points in X can also be ranked according to the learned f .

3. Multiview Graph-Based Learning for Classification

Single graph learning has improved to be an effective way for classification task, especially when labeled samples are limited. However, to make a comprehensive evaluation on SME credit risk, many different aspect information should be taken into consideration, such as the funds and credit condition of SME, core enterprises, and supply chain operation. These complementary knowledges contained in multiple views to comprehensively represent the credit status of SME. In this situation, more than one view features can be used to measure the affinity between vertices.

Suppose we have M sets of features. Accordingly, M graphs can be constructed, denoted as G_1, G_2, \dots, G_M . The regularization framework in equation (2) can be extended to handle multiview features by combining each graph with a set of weighting coefficient. Then, the framework is written as follows:

$$\Omega_{\text{fix}}(f) = \frac{1}{2} \sum_{g=1}^M \sum_{i,j=1}^N \alpha_g W_{g,ij} \left(\frac{f_i}{\sqrt{D_{g,ii}}} - \frac{f_j}{\sqrt{D_{g,jj}}} \right)^2. \quad (7)$$

The coefficient is correlative with the importance of the g -th view, constrained by $0 < \alpha_g < 1$ and $\sum_{g=1}^M \alpha_g = 1$. W_g and D_g are the affinity matrix and the diagonal matrix for graph G_g , respectively. The value of α_g can be set as prior knowledge in advance by domain experts according to the importance of different views. Then, the solution of equation (1) can be derived as follows:

$$f = \left(I + \frac{\sum_{g=1}^M \alpha_g L_g}{\lambda} \right)^{-1} y, \quad (8)$$

where L_g is the normalized graph Laplacian of G_g . Equation (8) amounts to learning on a fused graph, where several normalized graph Laplacians are combined through the

coefficient α_g . However, there has a lot of confusion about which view is the more important one for classification task. Subsequently, in most cases, it is not an easy way to set value for α_g in advance. It is desired to incorporate the influence of α_g into the above learning framework. To avoid trivial solutions, a relaxation on weight coefficient by changing α_g to α_g^r is performed on equation (7).

$$\Omega_{\text{adp}}(f) = \frac{1}{2} \sum_{g=1}^M \sum_{i,j=1}^N \alpha_g^r W_{g,ij} \left(\frac{f_i}{\sqrt{D_{g,ii}}} - \frac{f_j}{\sqrt{D_{g,jj}}} \right)^2, \quad (9)$$

where r is a hyperparameter larger than 1 and r is used to avoid only the smoothest graph play the key role ($\alpha_g = 1$), while graphs from the other views are invalid ($\alpha_g = 0$). Finally, the cost function of multiview graph-based learning is reformulated as follows:

$$H(f, \alpha) = \left\{ \underbrace{\frac{1}{2} \sum_{g=1}^M \sum_{i,j=1}^N \alpha_g^r W_{g,ij} \left(\frac{f_i}{\sqrt{D_{g,ii}}} - \frac{f_j}{\sqrt{D_{g,jj}}} \right)^2}_{\Omega_{\text{adp}}(f)} + \lambda \underbrace{\sum_{i=1}^N (f_i - y_i)^2}_{R_{\text{emp}}(f)} \right\} \quad (10)$$

$$= \sum_{g=1}^M \alpha_g^r f^T L_g f + \lambda \|f - y\|^2.$$

Equation (10) holds that data are in local manifold spaces for different views, respectively. Combination of multiple features can be done through a weighted union of graphs generated by different views. Each graph resembles a weak classifier generated from a single cue, and together they form a stronger classifier.

The relevance value α_g and weight coefficient α_g are achieved by minimizing equation (10).

$$\begin{aligned} & \operatorname{argmin}_{f, \alpha_1, \alpha_2, \dots, \alpha_M} H(f, \alpha), \\ & \text{s.t. } \alpha_g > 0, \quad \sum_{g=1}^M \alpha_g = 1. \end{aligned} \quad (11)$$

The optimization strategy is to learn f and α_g simultaneously by fixing one and optimizing another for each iteration. We first fix α and optimize f .

The partial derivative of the objective function with respect to f is

$$\frac{\partial H(f, \alpha)}{\partial f} = \sum_{g=1}^M \alpha_g^r L_g f + \lambda (f - y). \quad (12)$$

By setting $((\partial H(f, \alpha))/\partial f) = 0$, we have

$$f = \left(I + \frac{\sum_{g=1}^M \alpha_g^r L_g}{\lambda} \right)^{-1} y. \quad (13)$$

Then, we fix f to optimize α_g . The partial derivative of the objective function with respect to α is

$$\frac{\partial H(f, \alpha)}{\partial \alpha_g} = r \alpha_g^{r-1} f^T L_g f. \quad (14)$$

Similarly, we can obtain α_g with Lagrange multipliers because $\sum_{g=1}^M \alpha_g = 1$.

$$\alpha_g = \frac{(1/(f^T L_g f))^{1/(r-1)}}{\sum_{g=1}^M (1/(f^T L_g f))^{1/(r-1)}}. \quad (15)$$

4. Data and Variable

4.1. Data. To validate the effectiveness of the multiview graph-based learning method in forecasting SMEs' credit risk in SCF, we construct a proper database of Chinese SME. SCF is a new and developing branch of financing. Because of imperfection of the theoretical study, development of SCF in China is less than satisfactory. Until recent years, SCF business is gradually on the right track, making the supporting businesses and data disclosure relatively perfect. However, it is still difficult to gather a complete dataset of SCF, especially unlisted companies. Therefore, the listed companies are selected as the main data source.

Due to the huge difference in factors and characteristics of different industries, it may lead to the decline of the prediction model by using the whole industry data. Manufacturing industry is the key field of SCF, which attracted widely attention of researchers. However, traditional manufactured products (e.g., auto industry) were so complex, which has long chains from suppliers to distributors, that thorough collection of SCF data was impossible. Compared with traditional manufacturing, the mobile manufacturing supply chain data scale is moderate (e.g., a mobile phone has more than 80 parts, while a car has tens of thousands of parts). The division of labor among upstream suppliers, midstream manufacturers, and downstream distributors in mobile phone manufacturing industry is relatively clear. Furthermore, mobile phones are time sensitive, which requires enterprises on the chain to cooperate closely. Taking all these into account, we select 104 quoted SMEs data in Wind dataset. After deleting invalid data, our final dataset consists of 924 samples over the period (2011–2019).

Some literatures [26, 32] identified credit risk by considering whether the enterprise is special treatment (ST). They held that the ST and *ST companies are in financial crisis [33]. These companies may have higher credit risks than the general ones. ST company has suffered losses for two consecutive years. However, it cannot reflect the credit risk when a company suffered losses in the first year. The Z-score model [34] is adopted to define the credit risk, which is widely used in enterprise financial health measurement. Finally, the samples, whose Z-score < 1.8, are considered as risk enterprises in the experiments, while the remaining samples are nonrisk type. In total, the dataset consists of 265 risky enterprises and 659 nonrisky enterprises. In the experimental settings, \$40\$ samples are treated as known risky enterprises denoted by -1 and 40 are known as nonrisky enterprises denoted by $+1$. The rest unknown samples (denoted by 0) need to be validated by the multiview graph-based learning. The sample distribution is given in Table 1.

4.2. Independent Variables. In compliance with the widely accepted 5C principle, the credit risk assessment system is constructed by three views, such as supply chain operation, SME financial indicator, and nonfinancial indicator. To facilitate the observation, the descriptions of the independent variables for supply chain operation, SME financial indicator, and nonfinancial indicator are shown in Tables 2–4, respectively.

4.3. Evaluations. To evaluate the effectiveness of proposed methods, average accuracy, Type I error, and Type II error are adopted in the experiments. These evaluation criteria are defined as follows:

TABLE 1: Sample distribution of dataset.

	Dataset	Labeled (training)	Unlabeled (testing)
Risky	265	40 (-1)	225 (0)
Nonrisky	659	40 ($+1$)	619 (0)

TABLE 2: List of independent variables for supply chain operations (V_1).

Indexes	Variables	Mean	Std. Deviation
C_1	Customer concentration	27.78245	28.70553
C_2	Supplier concentration	19.9285	23.77576
C_3	CE current ratio	0.71919	0.73108
C_4	CE quick ratio	0.51692	0.52582
C_5	CE return on net assets	1.90677	2.52005
C_6	CE net profit rate of sales	2.05564	2.28251
C_7	CE asset liability ratio	26.04776	17.08954
C_8	CE inventory turnover days	45.55526	37.55354
C_9	CE accounts receivable turnover	3.68589	3.02582

*CE represents the core enterprise in supply chain

TABLE 3: List of independent variables for supply chain operations (V_2).

Indexes	Variables	Mean	Std. Deviation
C_{10}	EBIT	44186.84275	195438.89078
C_{11}	Gross profit	116784.07683	489589.42188
C_{12}	Net interest rate of total assets	6.14171	11.5233
C_{13}	Retained earnings	109346.28328	476524.70955
C_{14}	Cash turnover	953.97242	2781.79251
C_{15}	Inventory turnover days	85.66379	140.64208
C_{16}	Accounts receivable turnover	13.29238	98.98973
C_{17}	Turnover rate of current assets	1.26009	0.88023
C_{18}	Turnover of fixed assets	28.04207	146.34457
C_{19}	Asset liability ratio	38.03026	25.63465
C_{20}	Quick ratio	1.836	2.5333
C_{21}	Interest cover ratio	170.74379	2559.79827
C_{22}	Current ratio	2.22639	2.70785

TABLE 4: List of independent variables for supply chain operations (V_3).

Indexes	Variables	Mean	Std. Deviation
C_{23}	Total R&D expenditure	22777.54545	82601.12925
C_{24}	Market value of the company	88.86682	242.42438
C_{25}	Enterprise value	9.64595	58.71563
C_{26}	Staff quality	15.19087	21.52928
C_{27}	GDP	10.505	2.51509
C_{28}	CPI	2.589	1.12239
C_{29}	Consumer spending	25.60751	7.66701

$$\begin{aligned}
\text{Average Accuracy} &= \frac{N}{M}, \\
\text{type I error} &= \frac{E_p}{M_p}, \\
\text{type II error} &= \frac{E_n}{M_n},
\end{aligned} \tag{16}$$

here M is the total number of samples in the testing set, M_p and M_n are the number of positive and negative samples, respectively ($M = M_p + M_n$), N is the number of all correctly predicted samples, and E_p and E_n are the incorrectly predicted positive and negative samples.

Average accuracy is used to evaluate the overall performance of the model. The higher, the better. Sometimes, the imbalance between positive and negative samples may lead an inaccurate evaluation. Type I and Type II errors are also adopted as a supplement to evaluate the performance of the model. The lower, the better. Type I error describes the accuracy of positive samples. In SME credit risk assessment, Type II error with a high ratio will result in a loss of potential customers, which incorrectly classify nonrisky enterprises into risky. Type II error presents the accuracy of negative samples. Incorrectly classifying risky enterprises into nonrisky will lead a high ratio of Type II error, which may expose great risk on bank, core enterprises, and SMEs in SCF. In order to build a stable SCF system, Type II error is more vital than the other two criteria.

4.4. Experimental Settings. The experiments are performed on a computer which has Intel Xeon (R) 2.13 GHz 8 processors 8 GB RAM and the 64 bit Windows10 system. The experiments are conducted based on Python, NumPy, and Scikit-learn. The data are firstly normalized by the following:

$$X_{\text{norm}} = \frac{X - \mu}{\sigma}, \tag{17}$$

where X_{norm} is the normalized data, X is the source data, and μ and σ are the mean value and standard deviation of X . Then, principal component analysis (PCA) is performed on X_{norm} for dimension reduction. By preserving 90% energy, 18-dimensional features are prepared for LR, SVM, and GL. Particularly for M-GL, PCA is performed on each view separately. For GL and M-GL, a vertex x_i is linked with its nearest 5 neighbors and the Euclidean distance is used for measurement. The parameter λ in equation (6) and equation (8) is set to 0.01, and r in equation (8) is set to 1.9. The initial weights α_g for the three views are set equally.

5. Results and Discussion

5.1. Experimental Settings. In this part, we compare the proposed multiview graph learning for SME credit risk assessment in SCF with several related works.

- (1) *LR*. Logistic regression [35] is performed on the processed features from all the three views.

- (2) *SVM*. Support vector machine is performed on the processed features from all the three views.
- (3) *GL*. Single graph learning is performed by using the same features as LR and SVM. Only one graph is constructed by using the processed features.
- (4) *M-GL*. The proposed multiview graph-based learning approach integrates multiview features individually, which can adaptively assign different weights for the three views. Three graphs are constructed by using the processed features from the three views correspondingly.

Figures 1–3 describe the obtained results for the four models when the number of negative and positive training samples is 40, respectively. Upon inspecting the results in Figures 1–3, we notice that the proposed M-GL performs best on all the three evaluation criteria for SME credit risk assessment in SCF. It demonstrates the superiority of the proposed M-GL. From the perspective of feature fusion, all methods listed above integrate the features from the three views. However, M-GL treats these features as heterogeneous elements, which assigns different weights of each view for credit risk assessment. The other works treat the three views as a whole. They cannot distinguish which view is more valuable for SME credit risk assessment in SCF. Though some works (e.g., LR) can obtain the most important independent variables, the models depend too much on these features, leading a high data sensibility. On the contrary, graph-based methods (GL and M-GL) exploit the smoothness between samples. The loss or inaccuracy of one feature has little influence on the assessment model. During data collection, not all indicators can be easily obtained for every SME.

The noise resistance is an important characteristic in practical financial applications. In the next subsection, we will investigate the impact of different financial views for SME credit risk assessment in SCF.

5.2. The Impact of Different Financial Views. For traditional SME credit risk assessment methods, it is difficult to make a quantitative investigation on which one is the most informative view on risk assessment. They either concentrate the features as a whole or treat each indicator individually. Few works [32] give discriminative analysis on each indicator. However, the effectiveness of single indicator cannot provide evidences to guarantee the overall effectiveness of the corresponding view. The proposed M-GL constructs three individual subgraphs by using the features from supply chain operation, SME financial indicator, and nonfinancial indicator. By means of novel graph construction and optimization method, M-GL can adaptively distinguish the importance of different views. The upside of this kind of SME credit risk assessment is suitable for building the secure and stable SCF system. For example, the SME can improve their credit artificially by tampering a certain financial data to deceive the bank or core enterprises. It is of great danger for the whole SCF system. Tampering with a small amount of data will not affect the risk assessment results obviously for

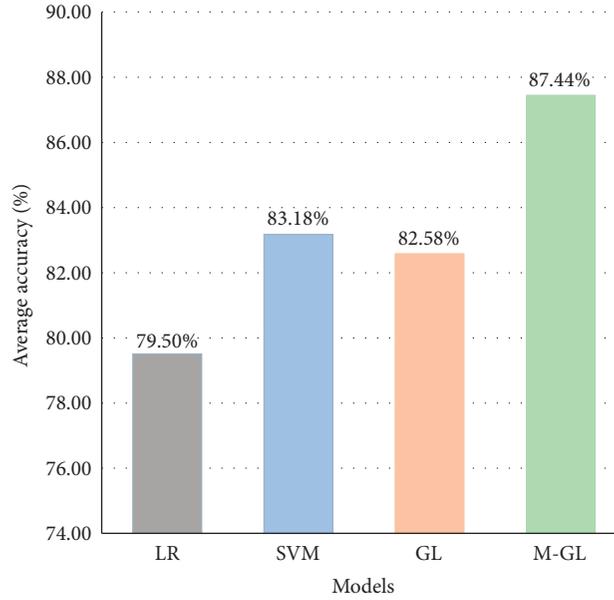


FIGURE 1: Average accuracy for different methods.

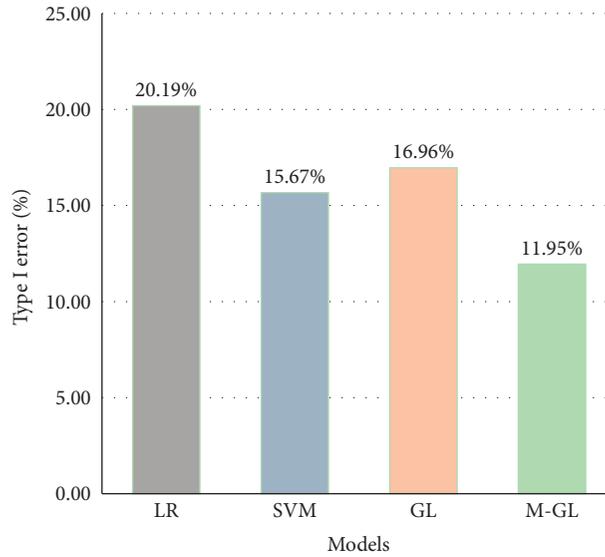


FIGURE 2: Type I error for different methods.

the proposed M-GL, which is one of the strengths of our approach. Table 5 gives the learned weight coefficient α_g for each view. α_1 , α_2 , and α_3 are the learned weight coefficients for supply chain operation, SME financial indicator, and nonfinancial indicator, respectively. From Table 5, we can see that SME financial indicator is the main factor for SME credit risk assessment in SCF. However, this does not mean that supply chain relationship is useless. The view of supply chain relationship is also an effective supplement, which is also informative to assess the SME credit risk assessment in SCF. In contrast, nonfinancial indicator has the lowest value

among the three views. Partly because we investigate the mobile manufacturing supply chain data, the difference of influencing factors between different industries have been weakened during data collection phase.

5.3. The Impact on the Number of Labeled Samples. The proposed M-GL is a graph-based transductive classification. An advantage of transductive methods is that it may be able to make better predictions with fewer labeled samples. In this part, we investigate the performance for different

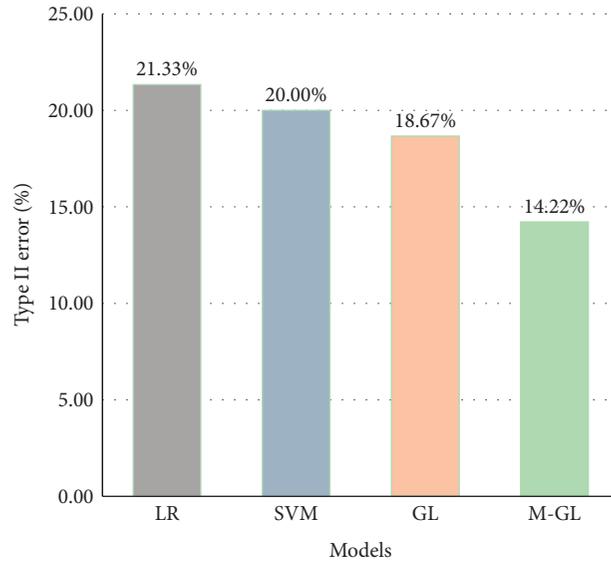


FIGURE 3: Type II error for different methods.

TABLE 5: The impact of different financial views.

α_1	α_2	α_3
0.284	0.598	0.118

amounts of initial labeled samples. Figure 4 reports the average accuracy, and Figures 5 and 6 report Type I and Type II errors of SVM and M-GL with varying amounts of positive and negative samples. M-GL can be regarded as label diffusion processes through three subgraphs. A good graph method needs to propagate labels to unlabeled samples. It is observed that M-GL outperforms SVM throughout the experiments. Especially, when the number of labeled samples is 10, the advantage of M-GL is more apparent than SVM on the three evaluation criteria. In the practical financial forecasting, labeling samples are always inconvenient, especially the risky SME. It has a great influence on traditional methods when the labeled samples are insufficient. Nevertheless, we can obtain lots of unlabeled SME information. M-GL models the relationship of these few labeled samples and large amounts of unlabeled samples through label diffusion processes on the smooth manifold for SME credit risk forecasting. Exploiting the large amounts of unlabeled samples to improve the accuracy of credit risk

forecasting is also an advantage of M-GL. In the next subsection, we will investigate the influence of smoothness on the classifier's performances.

5.4. The Impact on the Number of Nearest Neighbors. In our experiment, the number of nearest neighbors k is set to 5 in graph construction. During label diffusion, we need a smooth graph. A proper value of k will lead a stable and excellent performance. A large or a small k will degrade the performance. When k is small, it fails to capture the relationship of samples. On the contrary, when k is large, mismatched vertices may be brought into the graph. This explains the trends of curves shown in Figures 7–9. After k reaches an appropriate value, the performances begin to decline.

Because of the smoothness of the graph, we can obtain acceptable results through label diffusion. When we need to select partners from some SMEs, M-GL can give an appropriate credit ranking among the candidate enterprises.

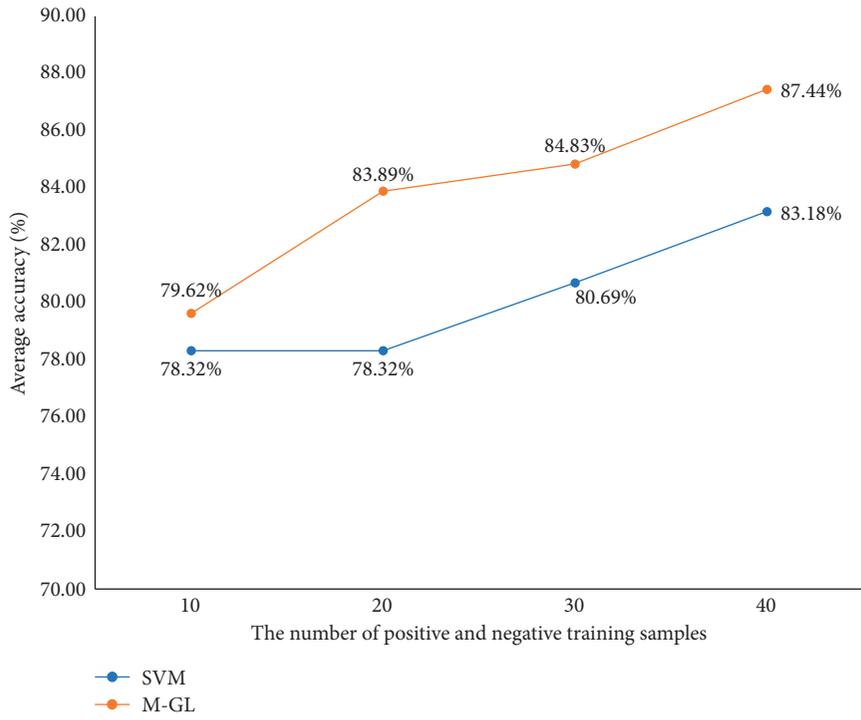


FIGURE 4: Variations of average accuracy with the increased number of labeled samples for M-GL.

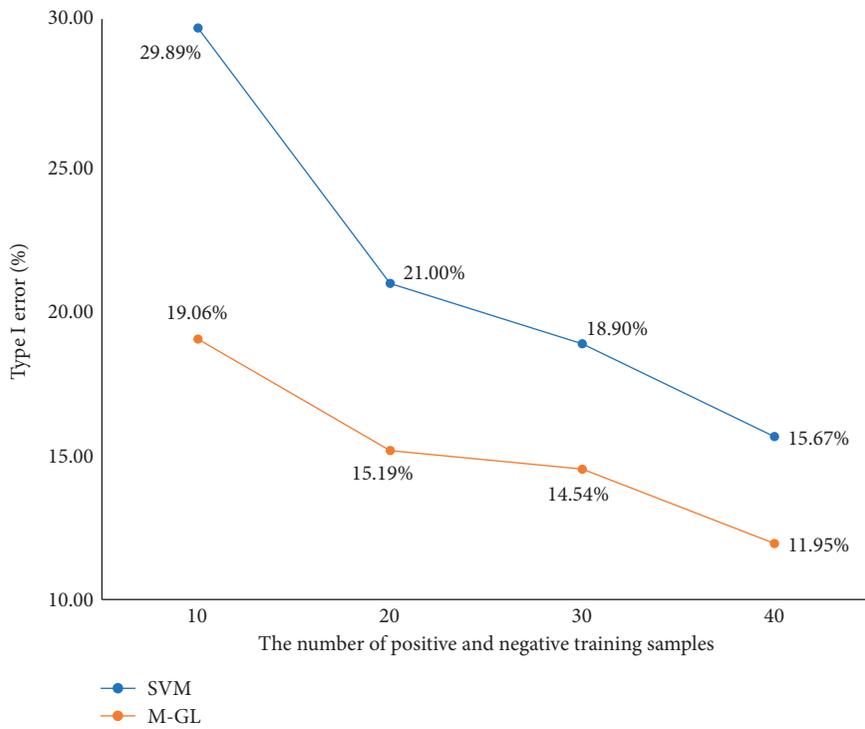


FIGURE 5: Variations of Type I error with the increased number of labeled samples for M-GL.

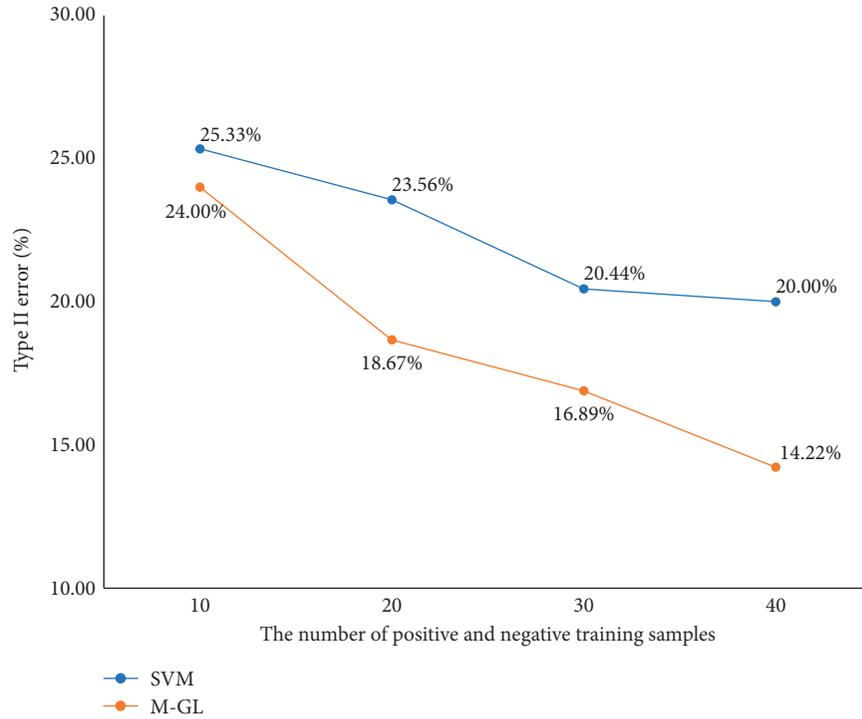


FIGURE 6: Variations of Type II error with the increased number of labeled samples for M-GL.

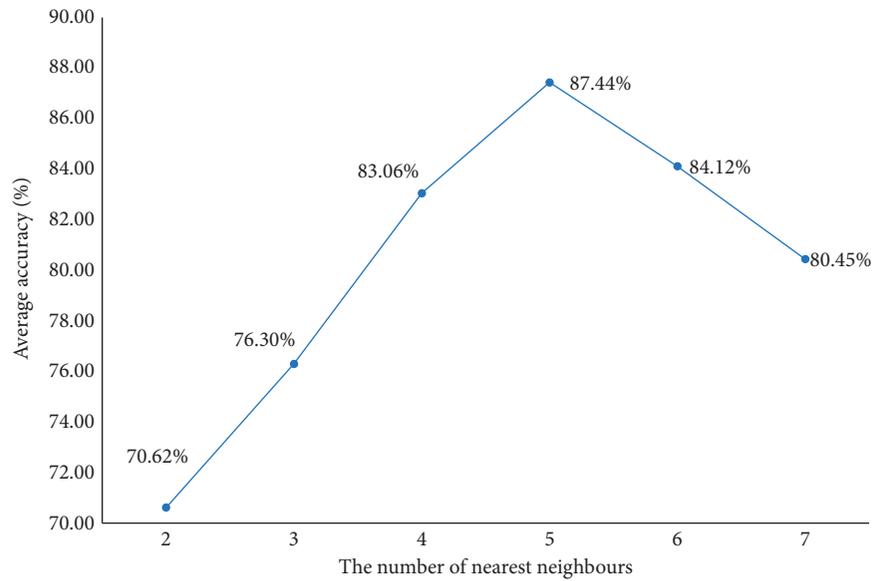


FIGURE 7: Variations of average accuracy with the increased number of nearest neighbors for M-GL.

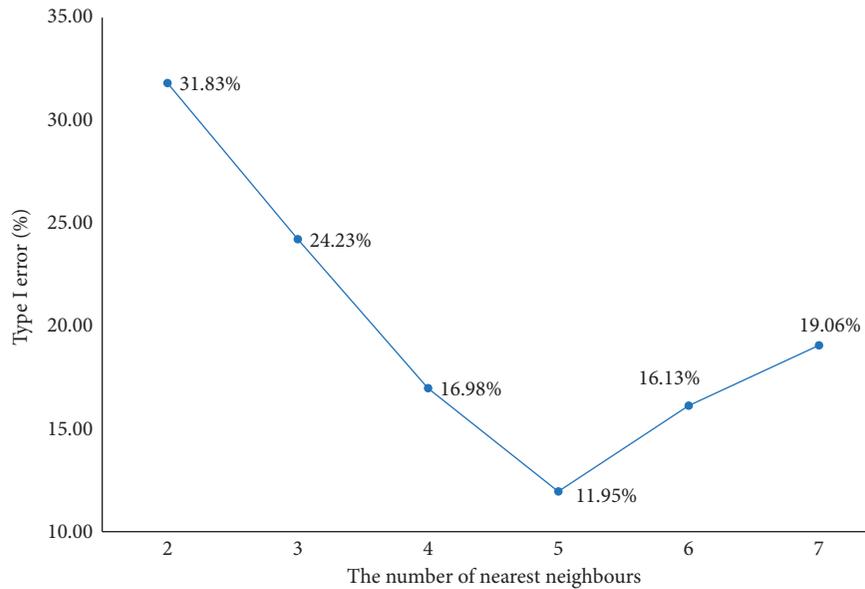


FIGURE 8: Variations of Type I error with the increased number of nearest neighbors for M-GL.

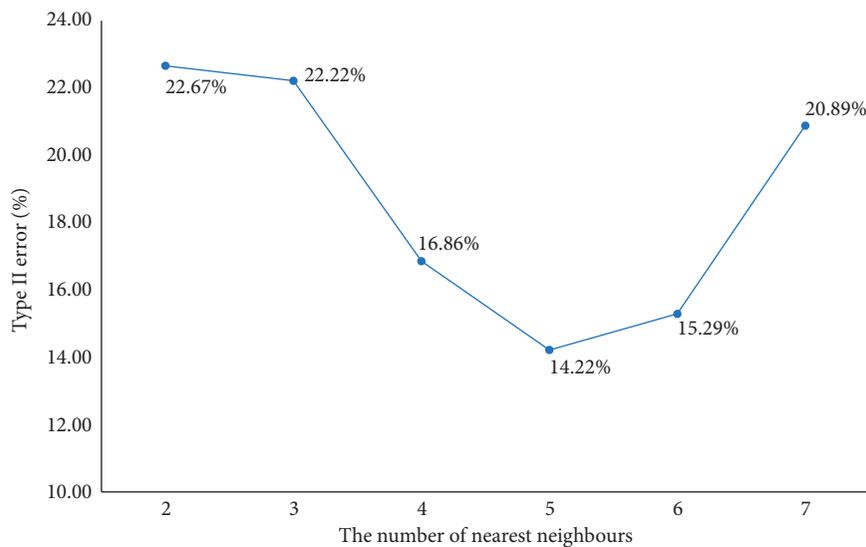


FIGURE 9: Variations of Type II error with the increased number of nearest neighbors for M-GL.

6. Conclusions

The primary purpose of the paper is to assess SME credit risks in SCF by means of multiview graph-based learning. The credit risk evaluation system is constructed from three aspects: supply chain operation, SME financial indicator, and nonfinancial indicator. Different from traditional methods, the proposed methods treat the three views individually, aiming to give a quantitative comparison for which view is more important for credit risk assessment. The paper also provides a detailed analysis on the impact of different financial views for SME credit risk assessment in SCF. To sum up, M-GL shows its effectiveness to assess SME credit risks in SCF. Especially, M-GL has good performance in dealing with

few labeled samples. In addition, SME financial indicator is the most critical view for SME credit risk assessment in SCF. Supply chain operation is a supplementary view, which is also a beneficial view to enhance the performance of assessment. Furthermore, M-GL not only uses the features from the three views but also exploits the relationships among samples. Therefore, the proposed M-GL for SME credit risk assessment is robust, which is applicable to prevent credit fraud by tampering with few data. When the bank or core enterprise needs to pick a partner, M-GL could give a relative sort order from the candidate enterprises.

Machine learning methods have shown their effectiveness in credit risk assessment. However, assessing SME credit risk in SCF is still a real challenge. Data

collection and annotation is a natural barrier to apply the new machine learning method for SME credit risk assessment in SCF, such as deep learning approaches. Moreover, mobile manufacturing supply chain data are utilized in this paper, where differences between industries are neglected. In the future, we will investigate SME credit risk assessment in SCF with a large dataset for comprehensive analysis.

Data Availability

The data used to support the findings of this study are available from the corresponding authors upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities (no. 18cx04007B), Qingdao Social Science Planning Project (no. QDSKL2001044), and Shandong Social Science Planning Research Project (no. 18CJ113).

References

- [1] W. A. Abbasi, Z. Wang, and A. Alsakarneh, "Overcoming smes financing and supply chain obstacles by introducing supply chain finance," *HOLISTICA-Journal of Business and Public Administration*, vol. 9, no. 1, pp. 7–22, 2018.
- [2] E. Letizia and F. Lillo, "Corporate payments networks and credit risk rating," *EPJ Data Science*, vol. 8, no. 1, p. 21, 2019.
- [3] R. C. Basole, M. A. Bellamy, H. Park, and J. Putrevu, "Computational analysis and visualization of global supply network risks," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 3, pp. 1206–1213, 2016.
- [4] Z. Zhao, D. Chen, L. Wang, and C. Han, "Credit risk diffusion in supply chain finance: a complex networks perspective," *Sustainability*, vol. 10, no. 12, p. 4608, 2018.
- [5] E. I. Altman and G. Sabato, "Modelling credit risk for SMEs: evidence from the U.S. Market," *Abacus*, vol. 43, no. 3, pp. 332–357, 2007.
- [6] S. D. Lekkakos, A. Serrano, and A. E. Ellinger, "Supply chain finance for small and medium sized enterprises: the case of reverse factoring," *International Journal of Physical Distribution & Logistics Management*, vol. 46, 2016.
- [7] D. A. Wuttke, C. Blome, H. Sebastian Heese, and M. Protopappa-Sieke, "Supply chain finance: optimal introduction and adoption decisions," *International Journal of Production Economics*, vol. 178, pp. 72–81, 2016.
- [8] Q. An and Y. Zhao, "Construction of credit risk evaluation system for small and medium-sized enterprises: based on principal component analysis and logistic model," in *Proceedings of the 5th International Conference on Economics, Management, Law and Education (EMLE 2019)*, pp. 244–250, Atlantis Press, Krasnodar, Russia, October 2020.
- [9] M. Yurdakul, "AHP as a strategic decision-making tool to justify machine tool selection," *Journal of Materials Processing Technology*, vol. 146, no. 3, pp. 365–376, 2004.
- [10] R. Rostamzadeh, M. K. Ghorabae, K. Govindan, A. Esmaili, and H. B. K. Nobar, "Evaluation of sustainable supply chain risk management using an integrated fuzzy TOPSIS- CRITIC approach," *Journal of Cleaner Production*, vol. 175, pp. 651–669, 2018.
- [11] W. Mou, W.-K. Wong, and M. McAleer, "Financial credit risk evaluation based on core enterprise supply chains," *Sustainability*, vol. 10, no. 10, p. 3699, 2018.
- [12] Y. Liu and L. Huang, "Supply chain finance credit risk assessment using support vector machine-based ensemble improved with noise elimination," *International Journal of Distributed Sensor Networks*, vol. 16, no. 1, Article ID 1550147720903631, 2020.
- [13] H. Jiang, W.-K. Ching, K. F. C. Yiu, and Y. Qiu, "Stationary mahalanobis kernel svm for credit risk evaluation," *Applied Soft Computing*, vol. 71, pp. 407–417, 2018.
- [14] D. Fantazzini and S. Figini, "Default forecasting for small-medium enterprises: does heterogeneity matter?" *International Journal of Risk Assessment and Management*, vol. 11, no. 1-2, pp. 138–163, 2009.
- [15] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [16] H.-q. Hu, L. Zhang, D.-h. Zhang, and L. CHEN, "Research on finance credit risk assessment of supply chain based on svm," *Soft Science*, vol. 5, pp. 1–7, 2011.
- [17] P. Danenas and G. Garsva, "Selection of support vector machines based classifiers for credit risk domain," *Expert Systems with Applications*, vol. 42, no. 6, pp. 3194–3204, 2015.
- [18] J. Zhou and T. Bai, "Credit risk assessment using rough set theory and ga-based SVM," in *Proceedings of the 2008 The 3rd International Conference on Grid and Pervasive Computing-Workshops*, pp. 320–325, IEEE, Kunming, China, May 2008.
- [19] A. Khashman, "Neural networks for credit risk evaluation: investigation of different neural models and learning schemes," *Expert Systems with Applications*, vol. 37, no. 9, pp. 6233–6239, 2010.
- [20] C. Luo, D. Wu, and D. Wu, "A deep learning approach for credit scoring using credit default swaps," *Engineering Applications of Artificial Intelligence*, vol. 65, pp. 465–470, 2017.
- [21] N. Ghatasheh, "Business analytics using random forest trees for credit risk prediction: a comparison study," *International Journal of Advanced Science and Technology*, vol. 72, pp. 19–30, 2014.
- [22] P. Giudici, B. Hadji-Misheva, and A. Spelta, "Network based credit risk models," *Quality Engineering*, vol. 32, no. 2, pp. 199–211, 2020.
- [23] L. Zhang, H. Hu, and D. Zhang, "A credit risk assessment model based on svm for small and medium enterprises in supply chain finance," *Financial Innovation*, vol. 1, no. 1, p. 14, 2015.
- [24] R.-Z. Xu and M.-K. He, "Application of deep learning neural network in online supply chain financial credit risk assessment," in *Proceedings of the 2020 International Conference on Computer Information and Big Data Applications (CIBDA)*, pp. 224–232, IEEE, Guiyang, China, April 2020.
- [25] H. Haiqing, Z. Lang, and Z. Daohong, "Research on smes credit risk assessment from the perspective of supply chain finance-a comparative study on the svm model and bp model," *Management Review*, vol. 11, 2016.
- [26] Y. Zhu, C. Xie, G.-J. Wang, and X.-G. Yan, "Predicting China's SME credit risk in supply chain finance based on machine learning methods," *Entropy*, vol. 18, no. 5, p. 195, 2016.
- [27] W. Liu, J. Wang, and S.-F. Chang, "Robust and scalable graph-based semisupervised learning," *Proceedings of the IEEE*, vol. 100, no. 9, pp. 2624–2638, 2012.

- [28] J. Liu, M. Li, Q. Liu, H. Lu, and S. Ma, "Image annotation via graph learning," *Pattern Recognition*, vol. 42, no. 2, pp. 218–228, 2009.
- [29] Z. Kang, H. Pan, S. C. Hoi, and Z. Xu, "Robust graph learning from noisy data," *IEEE Transactions on Cybernetics*, vol. 50, no. 5, pp. 1833–1843, 2019.
- [30] M. Culp and G. Michailidis, "Graph-based semisupervised learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 1, pp. 174–179, 2007.
- [31] L. Wang, W. Sun, Z. Zhao, and F. Su, "Modeling intra- and inter-pair correlation via heterogeneous high-order preserving for cross-modal retrieval," *Signal Processing*, vol. 131, pp. 249–260, 2017.
- [32] Y. Zhu, L. Zhou, C. Xie, G.-J. Wang, and T. V. Nguyen, "Forecasting SMEs' credit risk in supply chain finance with an enhanced hybrid ensemble machine learning approach," *International Journal of Production Economics*, vol. 211, pp. 22–33, 2019.
- [33] X. Chen, X. Wang, and D. D. Wu, "Credit risk measurement and early warning of smes: an empirical study of listed smes in China," *Decision Support Systems*, vol. 49, no. 3, pp. 301–310, 2010.
- [34] E. I. Altman, "Predicting financial distress of companies: revisiting the z-score and zeta models," in *Handbook of Research Methods and Applications in Empirical Finance* Edward Elgar Publishing, Cheltenham, EN, UK, 2013.
- [35] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*, Vol. 398, John Wiley & Sons, Hoboken, NJ, USA, 2013.