Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 1783445, 14 pages
https://doi.org/10.1155/2022/1783445

Research Article

A Meta-Path-Based Evaluation Method for Enterprise Credit Risk

Marui Du,[1] **Yue Ma**,[2] **and Zuoquan Zhang**[1]

[1]*School of Science, Beijing Jiaotong University, Beijing, China*
[2]*Guanghua School of Management, Peking University, Beijing, China*

Correspondence should be addressed to Marui Du; 17118446@bjtu.edu.cn

Received 14 May 2022; Accepted 29 September 2022; Published 13 October 2022

Academic Editor: Abdul Rehman Javed

Nowadays, small and medium-sized enterprises (SMEs) have become an essential part of the national economy. With the increasing number of such enterprises, how to evaluate their credit risk becomes a hot issue. Unlike big enterprises with massive data to analyse, it is hard to find enough primary information of SMEs to assess their financial status, which makes the credit risk evaluation result less accurate. Limited by the lack of primary data, how to infer SMEs' credit risk from secondary data, such as information about their upstream, downstream, parent, and subsidiary enterprises, attracts big attention from industry and academy. Targeting on accurately evaluating the credit risk of the SME, in this study, we exploit the representative power of the information network on various kinds of SME entities and SME relationships to solve the problem. With that, a heterogeneous information network of SMEs is built to mine enterprise's secondary information. Furthermore, a novel feature named meta-path feature is proposed to measure the credit risk, which makes us able to evaluate the financial status of SMEs from various perspectives. Experiments show that our proposed meta-path feature is effective to identify SMEs with credit risks.

1. Introduction

Small and medium-sized enterprise (SME) is one of the backbones in the national economy, whose development directly affects it. However, due to the incomplete management system and the lack of appropriate financial indicators, the credit risk assessment process is usually time-consuming, and the evaluation result is often of low accuracy. Therefore, in this study, we are going to propose an appropriate method of credit risk assessment to target this problem.

Industry and academy always have a critical focus on how to measure enterprise credit risk. Conventional approaches of assessment mainly extract enterprise-related features, such as financial indicators, to predict enterprise solvency. However, with the expansion of global market size in recent years, conventional approaches have lost their power of discrimination in the situations, where relations and interactions between SMEs are numerous and complicated. An SME's financial status can be easily affected by some actions from its other related SMEs. For example, the contagion risk is caused by associated credit entities, which besets many SMEs with the risk of default even in good financial conditions. Therefore, rather than single financial indicators, relations and interactions between SMEs should be paid more attention in studying SME credit risk.

To model the relations and interactions, various entities and their relationships can be considered in the information networks [1]. In the previous one, most of the researchers studied the abovementioned problem with a homogeneous information network [2] consisting only one single relation type and one entity type. However, in SME setting, the structure of the homogeneous information network may be a bit simple to explain the relationships between SMEs. To not lose important information, a heterogeneous information network [3] with complicated graph structure is more suitable to study the interaction between SMEs. In the heterogeneous information network, meta-paths (MP) [3] are taken as a fundamental data structure to capture semantical relationships between entities. Through MP, complicated relationships between entities can be systematically and concisely defined. The path provides a clear view of how entities interact mutually in the information network. In this study, to assess the status of SMEs, we exploit

the power of meta-path to study how influences among financial entities spread in the information network of SMEs.

In our method, we first build a heterogeneous information network of SMEs to describe interactive relationships between different entities associated with SME. Figure 1 is a toy example of the Alibaba heterogeneous information network, which demonstrates some possible connections of Alibaba and its related entities. For example, path "Alibaba $\longrightarrow^{\text{subsi di ary}}$ Lazada" represents information that Lazada is a subsidiary of Alibaba; path "Alibaba $\longrightarrow^{C}$ EO Bob $\longrightarrow^{c}$ ontroller Taobao" represents information that Alibaba's CEO, Bob, is also Taobao's controller; and path "Alibaba $\longrightarrow^{c}$ ontrol YouKu $\longrightarrow^{r}$ eport news" represents information that Alibaba's control enterprise, YouKu, is criticized by the newspaper. It is easy to see that through information networks, the interrelated relations between entities can be easily obtained. By building the information network of SMEs, we can not only obtain the self-related information but also the interactive information associated with the target enterprise.

With the given information network of SMEs, we propose a novel feature, -meta-path feature, to measure the impact through meta-paths from one financial entity to another. Unlike conventional financial indicators, the meta-path feature can be defined and applied very flexibly. The flexibility makes us able to evaluate the credit status of SMEs from various perspectives more comprehensively. The proposed meta-path feature can also explicitly show how much one entity can be affected by a specific logical path, which can provide an intuitive view for banks, lenders, and relevant experts to understand the credit risk faced by SMEs. In this way, SME default can be effectively identified.

The main contributions of this study are as follows:

(i) Due to the low relationship capturing power of the conventional approaches, in our method, we build a heterogeneous information network of SMEs to describe interactive relationships between different entities associated with SME

(ii) Propose three meta-path features to measure the impact through meta-paths from one financial entity to another from different angles

(iii) Our proposed meta-path features improve the performance of the SME credit risk evaluation. We compared state-of-the-art SME credit risk evaluation features with our proposed three meta-path features. Our meta-path feature achieved better results compared to state-of-the-art features.

In the rest of this paper, Section 2 introduces the SME credit risk evaluation method and the application of information networks. Section 3 builds a model of SMEs' heterogeneous information network and proposes the meta-path feature. In Section 4, by considering the ability of risk identification, three features are proposed based on the meta-path. Section 5 presents the experiment on three real-world datasets, and Section 6 concludes the study.

2. Related Works

In this section, we will review the related studies from the following perspectives: SME credit risk evaluation methods and information network applications.

2.1. SME Credit Risk Evaluation Methods. The credit risk evaluation model of SMEs was first established by Edmister [4] in 1972, leading to the emergence of a large number of credit risk measurement index systems. Most of the early credit evaluation models for SME at home and abroad follow the index system of the credit evaluation model for large enterprises, that is, the extraction of some key financial indicators of enterprise financial statements. Among these key financial indicators, profitability indicators [5, 6], such as the operating profit ratio and ratio of profits to cost, and solvency indicators [4, 5, 7], such as the current and quick ratio, are used the most. Besides, operational capacity indicators [8], development capacity indicators [9], and liquidity indicators [9] are added in many studies. Since financial indicators alone cannot lineate the complete picture of an enterprise, nonfinancial indicators such as managers background [10, 11], working experience [6], and enterprise internal structure [12, 13] are added for evaluation. However, financial and nonfinancial indicators cannot capture the contagion credit risk among financial entities since they are independent and do not consider the casual chain.

With the development of big data technology, a large amount of unstructured data related to enterprises have been accumulated, such as enterprise news information, enterprise transaction data, and enterprise relational data. The information used for SME credit risk evaluation has been extended. With the help of natural language processing techniques, we are able to explore meaningful information from text information. For example, Mosteller and Wallace [14] proposed an approach to analyse the Federalist papers; Spafford and Weeber [15] proposed an approach to analyse software forensics; Akram et al. [16] proposed a short text clustering technique using the deep learning model. Abbasi [17] proposed a framework to extract author-related information from unstructured textual information. In the field of SME credit risk evaluation, Tsai and Wang [18] proposed a method to extract enterprise-related news information and used it to support credit risk evaluation; Yin et al. [19] utilized legal judgments to support the evaluate of SME credit risk. Other than textual information, different kinds of relational information were also used for SME credit risk evaluation. Letizia and Lillo [20] used payment relation between enterprises; Tobback et al. [21] used enterprise's common shareholder and common director relation to extract interenterprise information; Kou et al. [22] focused on three different kinds of enterprise information, namely, basic enterprise information, manager/shareholders information, and payment and transactional information, to extract useful information. A summary of different types of features used in SMEs credit risk evaluation is listed in Table 1.

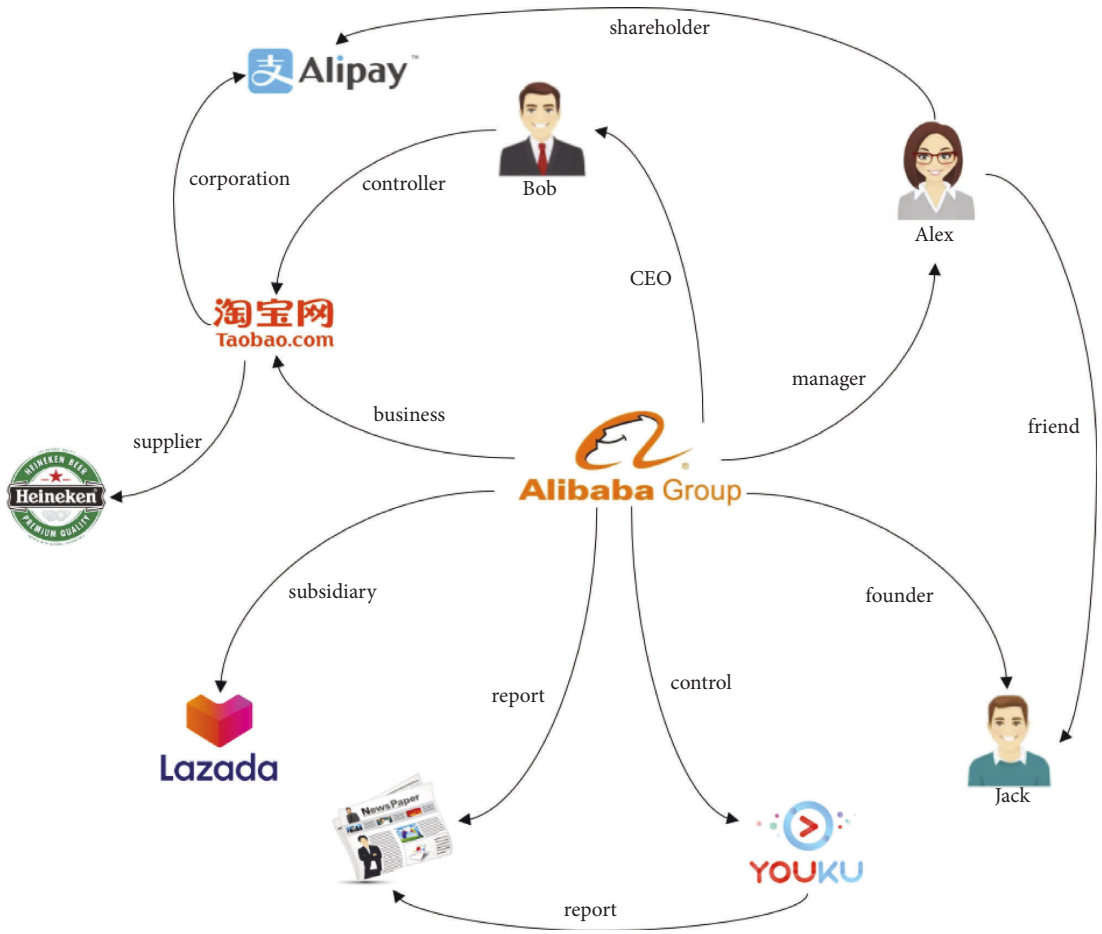

Figure 1: Alibaba heterogeneous information network example. There exist multiple types of nodes in the network, such as *enterprise* (Alibaba, Lazada, YouKu, Heineken), *person* (Bob, Alex, Jack), *commodity* (Taobao, Alipay), and *news* (newspaper). Links between nodes represent relation connect entities, for example, Jack is the founder of Alibaba, Heineken is the supplier of Taobao, and the newspapers report a piece of news of Alibaba's control company YouKu.

However, all of their works are built homogeneously, most of which do not consider heterogeneous information.

2.2. Information Network Applications. Recently, with the rapid improvement of computing capacity and the development of data mining technology, the information network has gained much attention from researchers and makes excellent work in the field of clustering [27–29], classification [30, 31], relation prediction [32, 33], and recommendation [34, 35]. Researchers often use two kinds of information networks, namely, the homogeneous information network and the heterogeneous information network. The homogeneous information network builds with same type of objects and link relations. For example, Jamali and Ester [36] built a social network for user recommendation based on user ratings; Ma et al. [37] built a friend relationship prediction network based on personal relations. These homogeneous information networks ignore the relationship between different objects and relations, which causes the loss of important information. The concept of heterogeneous information network was first proposed by Shi et al. [3] in 2009. It combines more information and contains logical semantics of different object types and link types. For example, Wang et al. [38] proposed a signed heterogeneous information network embedding to capture the sentiment links of online social information by considering users with sentiment and social relations; Hosseini et al. [39] used the heterogeneous information network with high dimensional data and rich relationships for medical diagnosis. The heterogeneous information network is usually used to capture complicated semantic and logical relationships among different entities.

2.3. Heterogeneous Information Network for SMEs. In the above-discussed related work, the state-of-the-art SME credit risk evaluation information is built on homogeneous information networks. It can only capture one single type of entity and one single type of relation, which is hard to capture the complicated relations of SMEs. Since massive data have been cumulated and many data analysis methods have been proposed, we are able to build a complicated network to capture more information of SMEs. The heterogeneous information network is able to capture more complicated graph structure, which is more suitable for

Table 1: Summary of feature types used for SMEs credit risk evaluation.

Work	Year	Features	Feature type
Edmister [4]	1972	19 accounting ratios	Financial information
Altman and Sabato [5]	2007	17 accounting ratios	Financial information
Chen et al. [23]	2010	Current liability, equity, asset, and closing stock price	Financial information
Psillaki et al. [10]	2010	Preinterest, pretax operating surplus/total assets, tangible assets/total assets, intangible assets/total assets, net growth, firm size, and managers background	Financial and nonfinancial information
Altman et al. [7]	2013	31 accounting ratios and 10 credit related variables	Financial and nonfinancial information
Hajek and Michalak [6]	2013	Enterprise size, enterprise reputation, profitability ratios, asset structure, business situation, market value ratios, and working experience	Financial and nonfinancial information
Moro and Fink [13]	2013	Economic and social environment, enterprise characteristics, and characteristics of the relationships between the loan manager and the SME manager	Financial and nonfinancial information
Angilella and Mazzù [24]	2015	Intangible assets/fixed assets, R&D/sales, ROA, short-term debt/equity, cash/total asset, development risk, production risk, technological risk, and market risk	Quantitative, financial, and nonfinancial information
Cultrera and Brédart [25]	2016	Current ratio, return on operating assets before depreciation, global degree of financial independence, proportion of gross value assed allocated to tax expenses, cash flow/total debt, business sector, enterprise size, and enterprise age	Financial and nonfinancial information
Gupta and Gregoriou [26]	2017	EBITDATA, taxes/total assets, total liability/total assets, short-term debt/equity book value, market-to-book ratio, excess return, standard deviation of past three months daily return, and price per share	Financial and nonfinancial information
Tsai and Wang [18]	2017	News information	Textual information
Letizia and Lillo [20]	2017	Enterprise payment relation	Relational information
Tobback et al. [21]	2017	Enterprise common shareholders and directors' relation	Relational information
Yin et al. [19]	2020	Current ratio, quick ratio, debt asset ratio, receivables turnover ratio, total asset turnover, operating profit ratio, missing ratio, enterprise age, registered capital, enterprise location, number of shareholders, number of insured, number of patents, and enterprise legal judgment	Financial, nonfinancial and textual information
Kou et al. [22]	2020	Basic enterprise information, managers/shareholders information, payment, and transactional information	Financial, nonfinancial and relational information

SMEs. Therefore, in this study, we build a heterogeneous information network for SMEs to more effectively evaluate SME credit risks, which considers both the heterogeneous information of SMEs and the semantic information carried by different SME entities. In this way, we are able to capture more information to accurately evaluate the credit risk of SMEs.

3. Model of SME Credit Risk

To evaluate SME credit risk, conventional methods adopted by experts usually make their judgments only based on the features directly affecting SME default, such as the asset-liability ratio, current ratio, and turnover rate, but not on logical relationships between SMEs, such as parent and subsidiary situations, upstream and downstream situations, enterprise director, and high-level manager related situations. For example, when a parent company defaults, the solvency of its subsidiaries will also be affected. If the influences exerted by the parent company are neglected, its subsidiary company's default conditions will be overestimated. Therefore, apart from the features directly affecting default, the logical relationships between SMEs should also be considered in evaluating SMEs' status. Paying attention to different connections between SMEs can improve both the reliability and the interpretability of the evaluation. This section will give a model of SME credit risk with logical relationships adopted.

3.1. SME Heterogeneous Information Network. A heterogeneous information network [3] is a classical data structure used to model objects and relations in a directed graph. This graph structure has shown its superiority in representing and storing knowledge about the natural world for many applications [40–42]. Given different objects in information networks, logical connections can be effectively constructed, and semantic relationships can be easily captured. Hence, we also build our model in an information network which is defined as follows:

Definition 1. With a schema $S = (\mathcal{A}, \mathcal{R})$, an information network is defined as a directed graph $G = (\mathcal{V}, \mathcal{E})$ with object type function $\tau: \mathcal{V} \longrightarrow \mathcal{A}$ and relation type function $\varphi: \mathcal{E} \longrightarrow \mathcal{R}$, where object $v \in \mathcal{V}$ belongs to object type $\tau(v) \in \mathcal{A}$ and link $e \in \mathcal{E}$ belongs to relation type $\varphi(e) \in \mathcal{R}$.

In this study, our model is built as a heterogeneous information network of SMEs. The SME schema is shown in Figure 2.

In our model, *enterprise* ($\mathcal{A}_e$), *commodity* ($\mathcal{A}_c$), *person* ($\mathcal{A}_p$), and *news* ($\mathcal{A}_n$) are four fundamental object types in studying SME credit risk. The studied relation types are summarized from public enterprise information and objective facts, such as the *shareholder* relation between enterprise and person, the *produce* relation between enterprise and commodity, and the *report* relation between enterprise and news. The types mentioned in this study are listed in Table 2.

With the SME schema defined, an example of SME heterogeneous information network is shown in Figure 3. We can see that v_1, v_2, and v_7 are the *enterprises*, that we have $\tau(v_1) = \mathcal{A}_e$, the same as $\tau(v_2)$ and $\tau(v_7)$ are. The v_6 and v_9 are the *commodities*, that we have $\tau(v_6) = \mathcal{A}_c$, the same as $\tau(v_9)$. The v_5, v_{10}, v_{11}, v_{12}, and v_{13} are *news*, that we have $\tau(v_5) = \mathcal{A}_n$, the same as $\tau(v_{10})$, $\tau(v_{11})$, $\tau(v_{12})$, and $\tau(v_{13})$ are. The v_3, v_4, and v_8 are *persons*, that we have $\tau(v_3) = \mathcal{A}_p$, the same as $\tau(v_4)$ and $\tau(v_8)$ are. The e_5 and e_8 are the relation of produces, that we have $\varphi(e_5) = \mathcal{R}_{\text{produce}}$, the same as $\varphi(e_8)$. The e_4, e_9, e_{10}, e_{11}, and e_{12} are the relation of *reports*, that we have $\varphi(e_4) = \mathcal{R}_{\text{report}}$, the same as $\varphi(e_9)$, $\varphi(e_{10})$, $\varphi(e_{11})$, and $\varphi(e_{12})$ are. The e_6 is the relation of *supply*, e_1 is the relation of *parent*, that we have $\varphi(e_6) = \mathcal{R}_{\text{supply}}$ and $\varphi(e_1) = \mathcal{R}_{\text{parent}}$. The e_7 and e_2 are the relations of *controller* and e_3 is the relation of *employee*, that we have $\varphi(e_7) = \mathcal{R}_{\text{control}}$, the same as $\varphi(e_2)$, and $\varphi(e_3) = \mathcal{R}_{\text{employee}}$. The e_{13} is the relation of *relate*, that we have $\varphi(e_{13}) = \mathcal{R}_{\text{relate}}$.

3.2. SME Meta-Path. In the SME network graph, we built in Section 3.1, a graph edge is used to present the relationship between two objects. Limited by the definition of edge, the represented relationships can only be some simple ones, which are insufficient to describe the relationships used in the problem of SME credit risk. In order to model complicated relationships, in this section, we introduce another data structure, meta-path (MP), to represent complicated and implicit relations in the SME network.

Definition 2. With a schema $S = (\mathcal{A}, \mathcal{R})$, a meta-path P is a path in the form $\mathcal{A}_1 \longrightarrow^{\mathcal{R}} 1\ \mathcal{A}_2 \longrightarrow^{\mathcal{R}} 2\ \ldots\ \longrightarrow^{\mathcal{R}} n\ \mathcal{A}_{n+1}$ which defines a composite relation $\mathcal{R} = \mathcal{R}_1^{\circ}\mathcal{R}_2^{\circ}\ldots\mathcal{R}_n^{\circ}$ between $\mathcal{A}_1$ and $\mathcal{A}_{n+1}$, where $\circ$ denotes the composition operator on relations.

For simplicity, we use the names of object types and relation types denoting the MP: $P = \mathcal{A}_1 \cdot \mathcal{R}_1 \cdot \mathcal{A}_2 \ldots \mathcal{R}_n \cdot \mathcal{A}_{n+1}$. With the definition of meta-path, a path $p = v_1 \cdot e_1 \cdot v_2 \ldots e_n \cdot v_{n+1}$ in graph G follows a meta-path P, if for any vertex $v_i \in \mathcal{V}$ and any edge $e_i \in \mathcal{E}$, the edge e_i is between v_i and v_{i+1}, $\tau(v_i) = \mathcal{A}_i$, and $\varphi(e_i) = \mathcal{R}_i$.

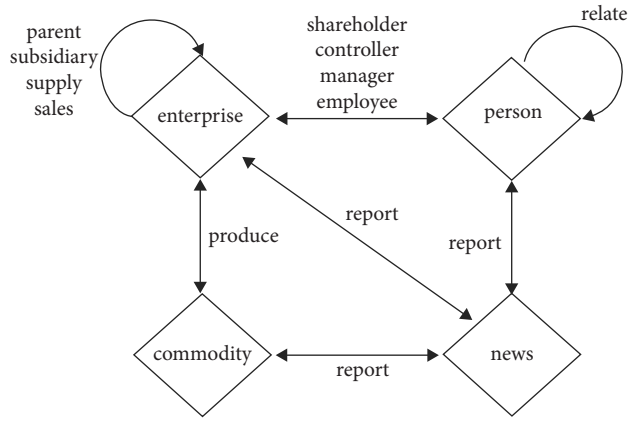

Figure 2: The SME network schema.

We also call p as a *path instance* of P with the denotation $p \in P$.

According to the definition, some examples of meta-paths can be seen in Figure 2. $P = \mathcal{A}_e \cdot \mathcal{R}_{\text{parent}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{report}} \cdot \mathcal{A}_n$ is a MP, which represents the information that the SME's parent enterprise has reported a news. According to Figure 3, there is a path instance $p = v_1 \cdot e_1 \cdot v_2 \cdot e_9 \cdot v_{10}$ of MP P. Because $\tau(v_1) = \mathcal{A}_e$, $\tau(v_2) = \mathcal{A}_e$, $\tau(v_{10}) = \mathcal{A}_n$, $\varphi(e_1) = \mathcal{R}_{\text{parent}}$, and $\varphi(e_9) = \mathcal{R}_{\text{report}}$.

The given MP definition structures logical connections between objects, making our model more expressive and interpretable. It not only can show explicit reasons for factors affecting SMEs on credit risk but also can explain implicit logics of correlation between objects having no direct links in the SME information network.

Compared to the information carried by objects, the information carried by meta-path is more critical in evaluating the credit risk of SMEs. The reason is that the expression ability of meta-path is stronger. Through different meta-paths, the same financial object may affect another financial object significantly differently. For instance, in Figure 3, we can see that there exist two paths from person v_4 to enterprise v_1. The first one is $p = v_1 \cdot e_3 \cdot v_4$ following meta-path $P = \mathcal{A}_e \cdot \mathcal{R}_{\text{employee}} \cdot \mathcal{A}_p$ and the second one is $p = v_1 \cdot e_2 \cdot v_3 \cdot e_{13} \cdot v_4$ following meta-path $P = \mathcal{A}_e \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{relate}} \cdot \mathcal{A}_p$. From the first path, the bribery scandal of an outsourcing employee v_4 may do limited harm to the enterprise v_1 since v_1 may have many other outsourcing employees to replace the role of v_4. However, from the second path, the bribery scandal of the outsourcing employee v_4 may do significant harm to enterprise v_1 since v_4 has a domestic relation with v_3 who directs enterprise v_1. Therefore, instead of inspecting each object's direct impact, our model regards a whole logical path consisting of objects and relations as a factor, in evaluating the credit risk of SMEs.

4. Meta-Path Impact on SME

In the above section, we have given the definition of MP, a well-patterned structure to represent various semantics relating to SME credit risk. It has been shown that even with

Table 2: Object-type and relation-type notations.

Notation	Descriptions
$\mathcal{A}_e$	The object type of *enterprise*
$\mathcal{A}_c$	The object type of *commodity*
$\mathcal{A}_p$	The object type of *person*
$\mathcal{A}_n$	The object type of *news*
$\mathcal{R}_{\text{parent}}$	The relation type of *parent* between enterprises
$\mathcal{R}_{\text{subsidiary}}$	The relation type of *subsidiary* between enterprises
$\mathcal{R}_{\text{supplier}}$	The relation type of *supply* between enterprises
$\mathcal{R}_{\text{saler}}$	The relation type of *sales* between enterprises
$\mathcal{R}_{\text{control}}$	The relation type of *controller* between enterprise and person
$\mathcal{R}_{\text{shareholder}}$	The relation type of *shareholder* between enterprise and person
$\mathcal{R}_{\text{manager}}$	The relation type of *manager* between enterprise and person
$\mathcal{R}_{\text{employee}}$	The relation type of *employee* between enterprise and person
$\mathcal{R}_{\text{produce}}$	The relation type of *produce* between enterprise and commodity
$\mathcal{R}_{\text{report}}$	The relation type of *report* between enterprise and news
$\mathcal{R}_{\text{relate}}$	The relation type of *relate* between person

no direct link given, the negative information of some SME may affect others heavily through meta-paths. For example, a piece of negative news about an enterprise director may lead to a bad reputation for his enterprise; a low-quality product of a parent enterprise may cause a loss of competitiveness to its subsidiary enterprises. Usually, potential risks brought from paths is nontrivial to be neglected when an SME is evaluated, but how to formulate such potential risk remains a question. In order to solve this question, in this section, we will propose several novel features, named meta-path feature, to represent the risk.

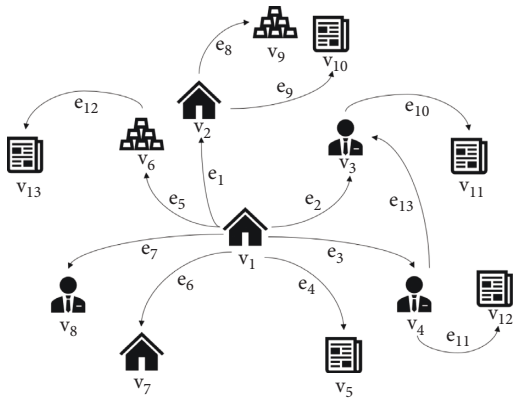

Figure 3: The SME heterogeneous information network.

4.1. Risk Inference from Object. Before introducing meta-path features, we first give a method to identify if there exists potential risk in financial objects themselves. According to the object types studied in Section 3.1, except the *news* object which is used to provide negative or positive information, a *commodity* object is regarded with potential risks if its quality is not reliable; a *person* object is regarded with potential risks if his capability is not qualified; an *enterprise* object is regarded with potential risks if it lacks credibility. In this study, in order to infer if potential risks exist, considering applicability and generality, we use the Naive Bayes model to infer if the mentioned objects are risky or not. Our probabilistic model is learnt from public historical data, such as financial statements, annual reports, and online public news. The definition of our Naive Bayes inference model is given as the following:

Definition 3. With the assumption that each attribute feature of an object is independent of each other, we define an inference function $\Gamma(x)$ to evaluate if object x is risky based on the probability $\mathbb{P}(y = 1|x)$ learnt from the Naive Bayes model.

$$\Gamma(x) = \begin{cases} 1, & \mathbb{P}(y = 1|x) > 0.5, \\ 0, & \text{otherwise}, \end{cases}$$

$$\mathbb{P}(y = 1|x) = \frac{\prod_i^n \mathbb{P}\left(x^{(i)}|y = 1\right)\mathbb{P}(y = 1)}{\prod_i^n \mathbb{P}\left(x^{(i)}|y = 1\right)\mathbb{P}(y = 1) + \prod_i^n \mathbb{P}\left(x^{(i)}|y = 0\right)\mathbb{P}(y = 0)}, \tag{1}$$

where $x^{(i)}$ is the ith attribute feature of object x, n is the number of all attributes, $y = 1$ indicates the risky object, and $y = 0$ indicates the nonrisky object.

With the inference function, we are able to identify the risk of a financial object by its own information. For instance, a *commodity* object with low sales volume, high repair rate, and high refund will be inferred as a risky one; a *person* object with irrelevant education background, irrelevant working experience, and short working years will be inferred as a risky one; an *enterprise* object with the low ROE ratio, low quick ratio, and high asset-liability ratio will be inferred as risky one. In the next section, we will study how to infer the potential risk from the MP level.

4.2. Risk Inference from Meta-Path. In an SME information network, an enterprise may have many paths linking to other financial objects, as shown in Figure 4. We can see enterprise J has 5 path instances for meta-path $P = \mathscr{A}_e \cdot \mathscr{R}_{\text{control}} \cdot \mathscr{A}_p \cdot \mathscr{R}_{\text{sharehol de r}} \cdot \mathscr{A}_e$ and enterprise K has 4 path instances for MP P.

With the inference function defined above, we are able to identify if objects in the above information network are risky or not. Thus, for a specific MP, with the objects linked by its path instances, it is natural to infer that an enterprise is most likely to be risky if potential risks exist in most of its linked objects. Based on this straight intuition, we next present several features to elaborate such risk from meta-path.

4.2.1. Meta-Path Feature. Given an enterprise x, the number of risky objects connected by a MP P are taken as an indicator to reflect the impact of meta-path P on target enterprise x. The larger the indicator is, the higher the potential risk exists. Formally, we call the indicator as naive MP feature and give its definition as the following:

Definition 4. Naive MP feature $N_P(x)$ is an indicator to reveal the impact of meta-path P on enterprise x:

$$N_P(x) = \frac{\left|\{x' \in D | \exists p_{x \rightsquigarrow x'} \in P, \Gamma(x') = 1\}\right|}{\left|\{x' \in D | \exists p_{x \rightsquigarrow x'} \in P\}\right|}, \quad (2)$$

where D is an SME object collection, $p_{x \rightsquigarrow x'}$ is a path instance from object x to object x', and $\Gamma(x)$ is the inference function defined in Section 4.1.

In Figure 4, if Q_2, Q_3, and Q_4 are the risky objects, then we have $N_P(J) = 3/5 = 0.6$, $N_P(K) = 3/4 = 0.75$.

4.2.2. Weighted Meta-Path Feature. Although the above-mentioned meta-path feature can effectively indicate the impact of MP, it may be argued that the impact of different objects on the same MP should not be the same. For all the objects in the network, irrelevant objects may affect small; relevant ones may matter big. Especially for an SME, the enterprise, which is its parent company, should influence it deeper than the enterprise, which only has one cooperation with it. Therefore, instead of treating all objects equally, it is more reasonable to treat them differently according to their relevance with the target SME. Next, considering relevance between objects, we will give a relevance-weighted version of meta-path feature accordingly.

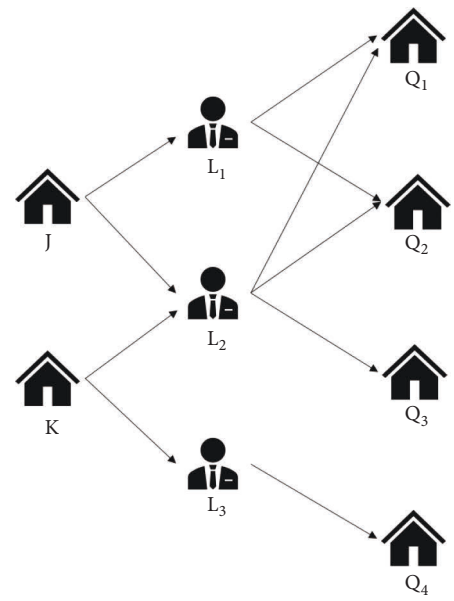

Figure 4: The path instances of MP $P = \mathscr{A}_e \cdot \mathscr{R}_{\text{control}} \cdot \mathscr{A}_p \cdot \mathscr{R}_{\text{sharehol de r}} \cdot \mathscr{A}_e$. J and K are the target SMEs, L_1, L_2, and L_3 are the controllers of J, and K. Q_1, Q_2, Q_3, and Q_4 are the associated enterprises of controllers L_1, L_2, and L_3.

Usually, relevance is used to measure how close two objects distance to each other. As there is no unified definition of relevance, different applications have unique and appropriate relevance measures. In SME application, there exists a usual fact that even though an enterprise is of well financial status, it may also default, which is caused by the propagated negative influence of its related upstream and downstream enterprises. Therefore, to measure the relevance between SME objects, a logical structure-based relevance measure is better than a textual context-based relevance measure.

A straightforward idea is that for any object pair, the two which have more paths should be more relevant. From this idea, we simply introduce a path count version of MP-weighted feature as follows:

Definition 5. CountSim MP weight feature $C_P(x)$ is an indicator to reveal the structure relevance impact of meta-path P on enterprise x. We call it CountSim MP feature.

$$C_P(x) = \frac{\left|\{x' \in D | \exists p_{x \rightsquigarrow x'} \in P\}\right|}{|\{x \in S\}| + \left|\{x' \in S'\}\right|}, \quad (3)$$

where S and S' are the SME object collections where all links from x and to x', respectively. D is another SME object collection which contains all objects.

The path count version is simple to apply but it makes little use of graph structure. In the SME heterogeneous information network, logical relationships between objects are captured by the structure of graph paths. Hence, compared to other measures, a path-based measure of relevance is more appropriate to be adopted in our model. At last, we apply HeteSim [43], an effective path-based similarity, to evaluate the relevance between objects.

Definition 6. HeteSim MP weight feature $H_P(x)$ takes HeteSim as the similarity measure to reveal the path relevance impact of meta-path P on enterprise x. We call it HeteSim MP feature.

$$H_P(x) = \frac{\sum_{x' \in \left\{x' \mid \exists p_{x \rightsquigarrow x'} \in P, \Gamma(x')=1\right\}} \text{HeteSim}(x, x')}{\sum_{x' \in \left\{x' \mid \exists p_{x \rightsquigarrow x'} \in P\right\}} \text{HeteSim}(x, x')}, \quad (4)$$

where $p_{x \rightsquigarrow x'}$ is a path instance from object x to object x', $\text{HeteSim}(x, x')$ is the relevance between object x and object x' under HeteSim, and $\Gamma(x)$ is the estimating function defined in Section 4.1.

5. Experiments

In this section, we are going to investigate the effectiveness of meta-path features. We conduct experiments on three real-world SME datasets. The result and explanation are detailed in this part.

5.1. Data and Settings. In our experiments, three datasets recording enterprises' statistics are used for comparison. GEM (The Growth Enterprise Market from Shenzhen Stock Exchange) and STAR (The Science and Technology Innovation Board from Shanghai Stock Exchange) datasets are about the SMEs of high technology, and SB (The Small and Medium-Sized Enterprise Board from Shenzhen Stock Exchange) dataset is about traditional enterprises. All the datasets can be downloaded from CSMAR (https://www.gtarsc.com). As this study only considers four types of financial entities (*person*, *commodity*, *enterprise*, and *news*), our experiments are only performed on the enterprises that at least relate to one person, one commodity, one other enterprise, and one piece of news.

The risk information about whether an enterprise lacks credibilities, a person lacks qualifications, and a commodity lacks reliabilities is obtained from CSMAR and CNINF (http://www.cninfo.com.cn), which provide an authoritative and professional assessment on the entities. The news information is collected from China Judgements Online (https://wenshu.court.gov.cn). The final details of datasets are shown in Table 3. As the gathered risk information may not be completed, for some important but unknown entities, we use the model in Section 4.1 to infer their risk. If an entity's inferred probability is larger than 0.75, it is deemed as risky.

Since the brought impact from a meta-path decreases with its length increasing, we only consider the meta-paths with length less than 6. The meta-paths which do not start with SME type are not selected for our experiments. With the proposed MP features, we test their performance using a default prediction model which is used to learn the weights associated with those features. The logistic regression model is taken as the prediction model, which is optimized by MLE (maximum likelihood estimation).

Table 3: Dataset information.

	GEM	STAR	SB
Number of enterprise	528	297	722
Related enterprise information	58478	26554	80729
Related person information	360462	38663	515504
Related news information	13026	3748	24718
Related commodities information	17450	8987	36735

In this section, all experiments were performed using Python 2.7.17 in Win 8.1+ with CPU *i*5 − 9300+ processor and 8G+ RAM.

5.2. Selection of Meta-Path Features. Even limited by the length constraint, there may still exist numerous meta-paths. Among all possible meta-path features, which ones are the most valuable ones? In this section, we will run experiments to show the importance of meta-path features.

We first generate 40 meta-path features according to Definition 4 for simplicity. Then, each feature is tested under the Wald test, and the p value of the feature associated with its meta-path is used to evaluate the feature's importance. The test is performed on all three datasets. Tables 4–6 list the top 20 significant meta-path features for each dataset and Tables 7–9 the bottom 20 meta-path features. From Tables 4–6, we can see that for all three datasets, the controller's ability ($\mathscr{A}_e \cdot \mathscr{R}_{\text{control}} \cdot \mathscr{A}_p$), parent enterprise financial status ($\mathscr{A}_e \cdot \mathscr{R}_{\text{parent}} \cdot \mathscr{A}_e$), and news reported for enterprise ($\mathscr{A}_e \cdot \mathscr{R}_{\text{report}} \cdot \mathscr{A}_n$) play very significant roles in determining SME status. However, from Tables 7–9, there is a trend that the longer the relation chains, the worse the performance of MP features. This may be due to the fact that longer links contain less valuable information as the longer relation chains means a more distant relationships with the enterprise. The longer the chain, the more distracting and inaccurate information it contains. Look into details, we find that for GEM and STAR datasets (high-technology SMEs), the MP features containing personnel relations are most significant, while those containing enterprise relations are the least. For SB dataset (conventional SMEs), the opposite is true. It is reasonable that the conventional SME, due to their own resource constraints, will pay more attention to the relationship with stakeholders in order to ensure stable development. The high-technology SME mainly focuses on technology research and development, so the ability of personnel has a significant impact on the enterprise.

5.3. Overall Comparisons of MP Feature. In this section, we compare our three kinds of MP features with four kinds of other state-of-the-art features proposed for evaluating SME credit risk. First kind of the compared features is conventional features [44], such as current liquidity, quick ratio,

Table 4: Top 20 significant meta-path features for the GEM dataset.

	Meta-path feature	P value	Significance level 1
1	$\mathcal{A}_e \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p$	$\mathbf{3.7876e-46}$	****
2	$\mathcal{A}_e \cdot \mathcal{R}_{\text{parent}} \cdot \mathcal{A}_e$	$\mathbf{5.3500e-37}$	****
3	$\mathcal{A}_e \cdot \mathcal{R}_{\text{report}} \cdot \mathcal{A}_n$	$\mathbf{1.7758e-32}$	****
4	$\mathcal{A}_e \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_e$	$1.0156e-32$	****
5	$\mathcal{A}_e \cdot \mathcal{R}_{\text{pro du ce}} \cdot \mathcal{A}_c \cdot \mathcal{R}_{\text{report}} \cdot \mathcal{A}_n$	$3.9645e-29$	****
6	$\mathcal{A}_e \cdot \mathcal{R}_{\text{manager}} \cdot \mathcal{A}_p$	$8.3629e-26$	****
7	$\mathcal{A}_e \cdot \mathcal{R}_{\text{pro du ce}} \cdot \mathcal{A}_c$	$2.2358e-26$	****
8	$\mathcal{A}_e \cdot \mathcal{R}_{\text{boar dm ember}} \cdot \mathcal{A}_p$	$6.1598e-23$	****
9	$\mathcal{A}_e \cdot \mathcal{R}_{\text{sharehol de r}} \cdot \mathcal{A}_p$	$2.4664e-15$	****
10	$\mathcal{A}_e \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{report}} \cdot \mathcal{A}_n$	$1.6067e-9$	****
11	$\mathcal{A}_e \cdot \mathcal{R}_{\text{parent}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{manager}} \cdot \mathcal{A}_p$	$3.7876e-6$	****
12	$\mathcal{A}_e \cdot \mathcal{R}_{\text{subsi di ary}} \cdot \mathcal{A}_e$	$5.3500e-5$	****
13	$\mathcal{A}_e \cdot \mathcal{R}_{\text{manager}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{report}} \cdot \mathcal{A}_n$	0.00121	***
14	$\mathcal{A}_e \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{relate}} \cdot \mathcal{A}_p$	0.00160	***
15	$\mathcal{A}_e \cdot \mathcal{R}_{\text{subsi di ary}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{report}} \cdot \mathcal{A}_n$	0.00236	***
16	$\mathcal{A}_e \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{manager}} \cdot \mathcal{A}_e$	0.00246	***
17	$\mathcal{A}_e \cdot \mathcal{R}_{\text{subsi di ary}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p$	0.00396	***
18	$\mathcal{A}_e \cdot \mathcal{R}_{\text{parent}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{report}} \cdot \mathcal{A}_n$	0.00615	***
19	$\mathcal{A}_e \cdot \mathcal{R}_{\text{parent}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p$	0.00758	***
20	$\mathcal{A}_e \cdot \mathcal{R}_{\text{supply}} \cdot \mathcal{A}_e$	0.00823	***

* $P < 0.1$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$.

Enterprise controller, enterprise parent company, and enterprise news are the top three most significant features in the GEM dataset.

Table 5: Top 20 significant meta-path features for the STAR dataset.

	Meta-path feature	P value	Significance level 2
1	$\mathcal{A}_e \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p$	$\mathbf{7.4107e-44}$	****
2	$\mathcal{A}_e \cdot \mathcal{R}_{\text{parent}} \cdot \mathcal{A}_e$	$\mathbf{3.3610e-37}$	****
3	$\mathcal{A}_e \cdot \mathcal{R}_{\text{sharehol de r}} \cdot \mathcal{A}_p$	$\mathbf{1.8247e-29}$	****
4	$\mathcal{A}_e \cdot \mathcal{R}_{\text{report}} \cdot \mathcal{A}_n$	$1.8709e-22$	****
5	$\mathcal{A}_e \cdot \mathcal{R}_{\text{subsi di ary}} \cdot \mathcal{A}_e$	$1.925e-17$	****
6	$\mathcal{A}_e \cdot \mathcal{R}_{\text{manager}} \cdot \mathcal{A}_p$	$2.7723e-11$	****
7	$\mathcal{A}_e \cdot \mathcal{R}_{\text{boar dm ember}} \cdot \mathcal{A}_p$	$9.2910e-8$	****
8	$\mathcal{A}_e \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{report}} \cdot \mathcal{A}_n$	$2.8380e-4$	****
9	$\mathcal{A}_e \cdot \mathcal{R}_{\text{subsi di ary}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{report}} \cdot \mathcal{A}_n$	0.000929	****
10	$\mathcal{A}_e \cdot \mathcal{R}_{\text{pro du ce}} \cdot \mathcal{A}_c$	0.00175	***
11	$\mathcal{A}_e \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_e$	0.00277	***
12	$\mathcal{A}_e \cdot \mathcal{R}_{\text{pro du ce}} \cdot \mathcal{A}_c \cdot \mathcal{R}_{\text{report}} \cdot \mathcal{A}_n$	0.00283	***
13	$\mathcal{A}_e \cdot \mathcal{R}_{\text{boar dm ember}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{report}} \cdot \mathcal{A}_n$	0.00341	***
14	$\mathcal{A}_e \cdot \mathcal{R}_{\text{supply}} \cdot \mathcal{A}_p$	0.0044	***
15	$\mathcal{A}_e \cdot \mathcal{R}_{\text{parent}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p$	0.00455	***
16	$\mathcal{A}_e \cdot \mathcal{R}_{\text{sales}} \cdot \mathcal{A}_e$	0.00476	***
17	$\mathcal{A}_e \cdot \mathcal{R}_{\text{parent}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{manager}} \cdot \mathcal{A}_p$	0.00496	***
18	$\mathcal{A}_e \cdot \mathcal{R}_{\text{manager}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{report}} \cdot \mathcal{A}_n$	0.00510	***
19	$\mathcal{A}_e \cdot \mathcal{R}_{\text{supply}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{report}} \cdot \mathcal{A}_n$	0.00528	***
20	$\mathcal{A}_e \cdot \mathcal{R}_{\text{parent}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{report}} \cdot \mathcal{A}_n$	0.00741	***

* $P < 0.1$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$.

assets turnover, a total of 16 financial indicators, and age of the enterprise, employment, a total of 5 nonfinancial indicators. In our experiments, we call it *SME CV*. The second kind of the compared features is textual feature [19], which is modeled from unstructured textual information. It not only contains enterprise basic financial and nonfinancial information but also the enterprise legal information. In our experiments, we call it *SME TF*. The third kind of the compared features is homogeneous path feature [21], which is modeled from homogeneous information networks. It contains only one object type and only one relation type, for example, two SMEs are related if they share a high-level manager. In our experiments, we call it *SME HPF*. The last kind of the compared features is multiple homogeneous path feature [22], which is modeled from more than one homogeneous information networks. It not only contains basic enterprise information but also three kinds of homogeneous path features, namely, manager network-based features, shareholder network-based features, and payment network-based features. In our experiments, we call it *SME MHPF*. For our MP features, we, respectively, select the Naive MP features, CountSim MP features, and HeteSim MP features

Table 6: Top 20 significant meta-path features for the SB dataset.

	Meta-path feature	P value	Significance level 3
1	$\mathcal{A}_e \cdot \mathcal{R}_{\text{report}} \cdot \mathcal{A}_n$	$\mathbf{1.2831e-48}$	****
2	$\mathcal{A}_e \cdot \mathcal{R}_{\text{parent}} \cdot \mathcal{A}_e$	$\mathbf{3.0306e-45}$	****
3	$\mathcal{A}_e \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p$	$\mathbf{1.5510e-36}$	****
4	$\mathcal{A}_e \cdot \mathcal{R}_{\text{subsi di ary}} \cdot \mathcal{A}_e$	$6.5260e-35$	****
5	$\mathcal{A}_e \cdot \mathcal{R}_{\text{subsi di ary}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{report}} \cdot \mathcal{A}_n$	$3.7263e-35$	****
6	$\mathcal{A}_e \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_e$	$4.4973e-33$	****
7	$\mathcal{A}_e \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{manager}} \cdot \mathcal{A}_e$	$2.3524e-33$	****
8	$\mathcal{A}_e \cdot \mathcal{R}_{\text{supply}} \cdot \mathcal{A}_e$	$1.1475e-28$	****
9	$\mathcal{A}_e \cdot \mathcal{R}_{\text{boar dm ember}} \cdot \mathcal{A}_p$	$6.8367e-27$	****
10	$\mathcal{A}_e \cdot \mathcal{R}_{\text{parent}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{manager}} \cdot \mathcal{A}_p$	$5.2674e-13$	****
11	$\mathcal{A}_e \cdot \mathcal{R}_{\text{pro du ce}} \cdot \mathcal{A}_c$	$1.2831e-11$	****
12	$\mathcal{A}_e \cdot \mathcal{R}_{\text{boar dm ember}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{report}} \cdot \mathcal{A}_n$	$3.0306e-9$	****
13	$\mathcal{A}_e \cdot \mathcal{R}_{\text{sharehol de r}} \cdot \mathcal{A}_p$	$1.5510e-8$	****
14	$\mathcal{A}_e \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{sharehol de r}} \cdot \mathcal{A}_e$	$6.5260e-6$	****
15	$\mathcal{A}_e \cdot \mathcal{R}_{\text{sales}} \cdot \mathcal{A}_e$	$3.7263e-5$	****
16	$\mathcal{A}_e \cdot \mathcal{R}_{\text{parent}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{pro du ce}} \cdot \mathcal{A}_p$	$4.4973e-4$	****
17	$\mathcal{A}_e \cdot \mathcal{R}_{\text{manager}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_e$	$2.3524e-4$	****
18	$\mathcal{A}_e \cdot \mathcal{R}_{\text{sales}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{report}} \cdot \mathcal{A}_n$	0.00114	***
19	$\mathcal{A}_e \cdot \mathcal{R}_{\text{parent}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{report}} \cdot \mathcal{A}_n$	0.00526	***
20	$\mathcal{A}_e \cdot \mathcal{R}_{\text{supply}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{report}} \cdot \mathcal{A}_n$	0.00683	***

* $P < 0.1$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$.
Enterprise news, enterprise parent company, and enterprise controller are the top three most significant features in the SB dataset.

Table 7: Bottom 20 significant meta-path features for the GEM dataset.

	Meta-path feature	P value	Significance level 4
1	$\mathcal{A}_e \cdot \mathcal{R}_{\text{sales}} \cdot \mathcal{A}_e$	0.0783	*
2	$\mathcal{A}_e \cdot \mathcal{R}_{\text{supply}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{report}} \cdot \mathcal{A}_n$	0.0778	*
3	$\mathcal{A}_e \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{sharehol de r}} \cdot \mathcal{A}_e$	0.0788	*
4	$\mathcal{A}_e \cdot \mathcal{R}_{\text{sharehol de r}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_e$	0.0832	*
5	$\mathcal{A}_e \cdot \mathcal{R}_{\text{sharehol de r}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{sharehol de r}} \cdot \mathcal{A}_e$	0.0854	*
6	$\mathcal{A}_e \cdot \mathcal{R}_{\text{sharehol de r}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{report}} \cdot \mathcal{A}_n$	0.0861	*
7	$\mathcal{A}_e \cdot \mathcal{R}_{\text{sales}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{report}} \cdot \mathcal{A}_n$	0.0874	*
8	$\mathcal{A}_e \cdot \mathcal{R}_{\text{sharehol de r}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{manager}} \cdot \mathcal{A}_e$	0.0889	*
9	$\mathcal{A}_e \cdot \mathcal{R}_{\text{supply}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{pro du ce}} \cdot \mathcal{A}_c$	0.0893	*
10	$\mathcal{A}_e \cdot \mathcal{R}_{\text{manager}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_e$	0.0896	*
11	$\mathcal{A}_e \cdot \mathcal{R}_{\text{sales}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{pro du ce}} \cdot \mathcal{A}_c$	0.0899	*
12	$\mathcal{A}_e \cdot \mathcal{R}_{\text{parent}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{pro du ce}} \cdot \mathcal{A}_c$	0.0932	*
13	$\mathcal{A}_e \cdot \mathcal{R}_{\text{supply}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{manager}} \cdot \mathcal{A}_p$	0.1775	—
14	$\mathcal{A}_e \cdot \mathcal{R}_{\text{supply}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p$	2.4662	—
15	$\mathcal{A}_e \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{employee}} \cdot \mathcal{A}_p$	3.9645	—
16	$\mathcal{A}_e \cdot \mathcal{R}_{\text{sales}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{sharehol de r}} \cdot \mathcal{A}_p$	6.1598	—
17	$\mathcal{A}_e \cdot \mathcal{R}_{\text{manager}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{sharehol de r}} \cdot \mathcal{A}_p$	7.4662	—
18	$\mathcal{A}_e \cdot \mathcal{R}_{\text{sales}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p$	10.6710	—
19	$\mathcal{A}_e \cdot \mathcal{R}_{\text{sharehol de r}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{employee}} \cdot \mathcal{A}_e$	12.4639	—
20	$\mathcal{A}_e \cdot \mathcal{R}_{\text{employee}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{manager}} \cdot \mathcal{A}_p$	16.0762	—

* $P < 0.1$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$.

according to the ranking result in Section 5.2 as the candidate features for comparison. All the comparisons are still conducted on the mentioned three datasets. To compare the mentioned methods, we first select the top 10 performed features of each method. Then, we use their average AUC score as the overall score of each mentioned method. The comparison results are summarized in Table 10.

We can see that the heterogeneous MP features outperform all the comparison features in all three datasets. For the proposed MP features, it turns out that (1) all the MP features show better classification performance than the SME conventional features, textual features, and homogeneous path features; (2) the classification performance of the CountSim MP features and the HeteSim MP features beats the Naive MP features; (3) the classification performance of the CountSim MP features and the HeteSim MP features are similar. The above results demonstrate the effectiveness of our proposed features in classifying default SMEs.

5.4. Discussion. In this section, we will discuss some interesting point which we found in our experiments. In general, prediction accuracy increases with data size

Table 8: Bottom 20 significant meta-path features for the STAR dataset.

	Meta-path feature	P value	Significance level 5
1	$\mathcal{A}_e \cdot \mathcal{R}_{\text{sharehol de r}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{relate}} \cdot \mathcal{A}_p$	0.0538	*
2	$\mathcal{A}_e \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{manager}} \cdot \mathcal{A}_e$	0.0598	*
3	$\mathcal{A}_e \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{sharehol de r}} \cdot \mathcal{A}_e$	0.0641	*
4	$\mathcal{A}_e \cdot \mathcal{R}_{\text{sharehol de r}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{report}} \cdot \mathcal{A}_n$	0.0870	*
5	$\mathcal{A}_e \cdot \mathcal{R}_{\text{subsi di ary}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p$	0.0873	*
6	$\mathcal{A}_e \cdot \mathcal{R}_{\textit{sharehol de r}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\textit{sharehol de r}} \cdot \mathcal{A}_e$	0.0881	*
7	$\mathcal{A}_e \cdot \mathcal{R}_{\text{manager}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p$	0.0886	*
8	$\mathcal{A}_e \cdot \mathcal{R}_{\text{parent}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{pro du ce}} \cdot \mathcal{A}_c$	0.0928	*
9	$\mathcal{A}_e \cdot \mathcal{R}_{\text{subsi di ary}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{pro du ce}} \cdot \mathcal{A}_c$	0.0941	*
10	$\mathcal{A}_e \cdot \mathcal{R}_{\text{sharehol de r}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{manager}} \cdot \mathcal{A}_e$	0.0951	*
11	$\mathcal{A}_e \cdot \mathcal{R}_{\text{subsi di ary}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{manager}} \cdot \mathcal{A}_p$	0.0974	*
12	$\mathcal{A}_e \cdot \mathcal{R}_{\text{supply}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p$	0.0976	*
13	$\mathcal{A}_e \cdot \mathcal{R}_{\text{sales}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{report}} \cdot \mathcal{A}_n$	0.0982	*
14	$\mathcal{A}_e \cdot \mathcal{R}_{\text{sales}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{pro du ce}} \cdot \mathcal{A}_c$	0.0987	*
15	$\mathcal{A}_e \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{employee}} \cdot \mathcal{A}_e$	4.6731	—
16	$\mathcal{A}_e \cdot \mathcal{R}_{\text{sales}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p$	7.7232	—
17	$\mathcal{A}_e \cdot \mathcal{R}_{\text{supply}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{pro du ce}} \cdot \mathcal{A}_c$	9.2910	—
18	$\mathcal{A}_e \cdot \mathcal{R}_{\text{sharehol de r}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_e$	12.8380	—
19	$\mathcal{A}_e \cdot \mathcal{R}_{\text{supply}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{manager}} \cdot \mathcal{A}_p$	14.4176	—
20	$\mathcal{A}_e \cdot \mathcal{R}_{\text{sharehole de r}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{employee}} \cdot \mathcal{A}_e$	17.5919	—

* $P < 0.1$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$.

Table 9: Bottom 20 significant meta-path features for the SB dataset.

	Meta-path feature	P value	Significance level 6
1	$\mathcal{A}_e \cdot \mathcal{R}_{\text{pro du ce}} \cdot \mathcal{A}_c \cdot \mathcal{R}_{\text{report}} \cdot \mathcal{A}_n$	0.0714	*
2	$\mathcal{A}_e \cdot \mathcal{R}_{\text{manager}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{report}} \cdot \mathcal{A}_n$	0.0730	*
3	$\mathcal{A}_e \cdot \mathcal{R}_{\text{sharehol de r}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{relate}} \cdot \mathcal{A}_p$	0.07551	*
4	$\mathcal{A}_e \cdot \mathcal{R}_{\text{parent}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p$	0.07652	*
5	$\mathcal{A}_e \cdot \mathcal{R}_{\text{parent}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{sharehol de r}} \cdot \mathcal{A}_p$	0.08352	*
6	$\mathcal{A}_e \cdot \mathcal{R}_{\text{subsi di ary}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p$	0.08497	*
7	$\mathcal{A}_e \cdot \mathcal{R}_{\text{subsi di ary}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{pro du ce}} \cdot \mathcal{A}_c$	0.08632	*
8	$\mathcal{A}_e \cdot \mathcal{R}_{\text{sales}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{pro du ce}} \cdot \mathcal{A}_c$	0.08756	*
9	$\mathcal{A}_e \cdot \mathcal{R}_{\text{sales}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p$	0.09367	*
10	$\mathcal{A}_e \cdot \mathcal{R}_{\text{sharehol de r}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{sharehol de r}} \cdot \mathcal{A}_e$	0.09526	*
11	$\mathcal{A}_e \cdot \mathcal{R}_{\text{subsi di ary}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{manager}} \cdot \mathcal{A}_p$	0.09831	*
12	$\mathcal{A}_e \cdot \mathcal{R}_{\text{sharehol de r}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p$	0.09836	*
13	$\mathcal{A}_e \cdot \mathcal{R}_{\text{manager}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{manager}} \cdot \mathcal{A}_e$	5.5101	—
14	$\mathcal{A}_e \cdot \mathcal{R}_{\text{employee}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{manager}} \cdot \mathcal{A}_e$	6.5260	—
15	$\mathcal{A}_e \cdot \mathcal{R}_{\text{supply}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p$	9.7263	—
16	$\mathcal{A}_e \cdot \mathcal{R}_{\text{sharehol de r}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{employee}} \cdot \mathcal{A}_p$	14.4973	—
17	$\mathcal{A}_e \cdot \mathcal{R}_{\text{sales}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{sharehol de r}} \cdot \mathcal{A}_p$	23.5246	—
18	$\mathcal{A}_e \cdot \mathcal{R}_{\text{supply}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{sharehol de r}} \cdot \mathcal{A}_p$	27.7731	—
19	$\mathcal{A}_e \cdot \mathcal{R}_{\text{control}} \cdot \mathcal{A}_p \cdot \mathcal{R}_{\text{employee}} \cdot \mathcal{A}_p$	28.3672	—
20	$\mathcal{A}_e \cdot \mathcal{R}_{\text{supply}} \cdot \mathcal{A}_e \cdot \mathcal{R}_{\text{manager}} \cdot \mathcal{A}_p$	31.5267	—

* $P < 0.1$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$.

Table 10: Average AUC score comparison for three datasets.

	SME CV	SME TF	SME HPF	SME MHPF	Naive MP	CountSim MP	HeteSim MP
GEM	0.716	0.732	0.728	0.744	0.747	0.771	0.774
STAR	0.654	0.698	0.707	0.728	0.759	0.767	0.791
SB	0.721	0.734	0.733	0.747	0.752	0.756	0.783

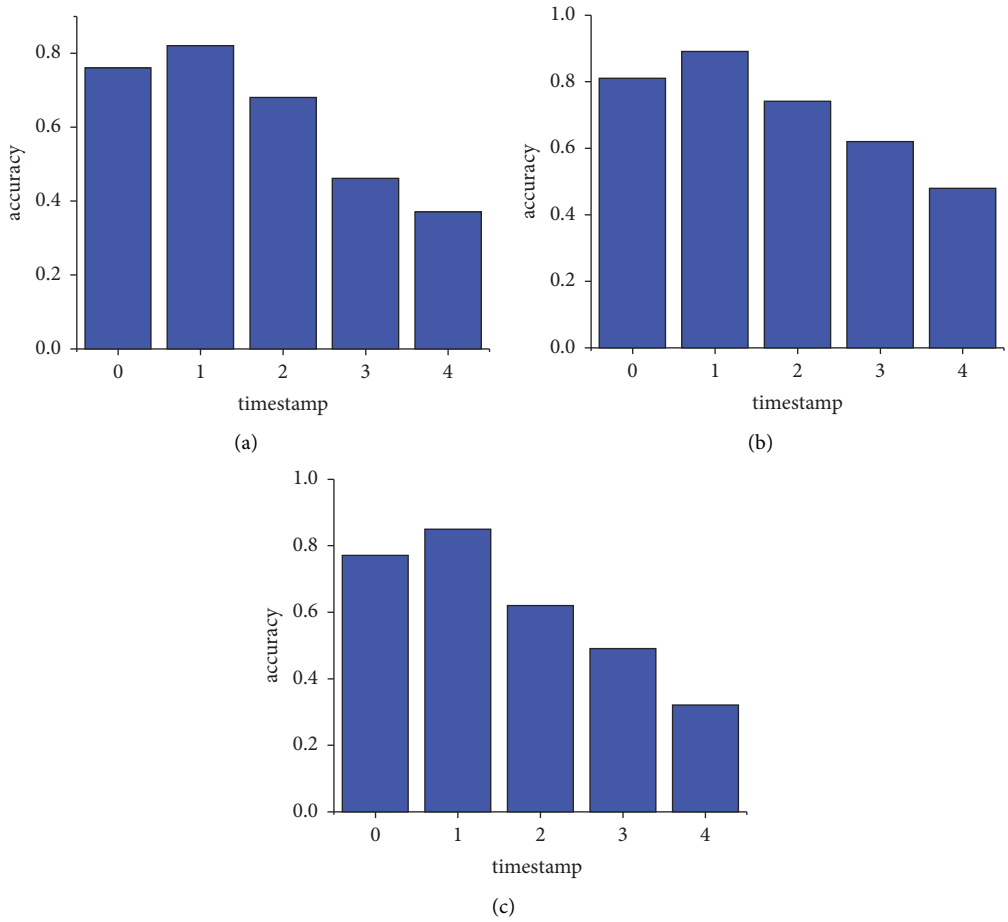

Figure 5: Classification accuracy of MP features under different timestamps. (a) The GEM dataset. (b) The STAR dataset. (c) The SB dataset.

increasing. However, we found that for SMEs, the impact of data size is affected by the timestamp of data. Next, we will detail and discuss how this affection comes. Figures 5(a)–5(c) show the classification accuracy of meta-path features under different timestamps.

It is interesting that when we extend SME data used in our model with the latest data in one year, the accuracy of the model increases for all three datasets. But if we extend that with data before last year, the accuracy of the model shows a declining trend. This phenomenon may be due to the fact that if the additional data are still in its valid duration, our model can be learnt more fully within the life circle of the enterprise. But if the additional data are out of its valid duration, our model may be learnt out of the life circle and lose its effectiveness. For example, employee turnover rate over two years cannot reflect the truth about the target enterprise now. The number of corporate enterprises over two years may be changed.

6. Conclusion

This study proposes a meta-path-based SME credit risk evaluation method that models SME-related information as a heterogeneous information network. In detail, we first build an SME heterogeneous information network based on four entity types and ten relation types. The heterogeneous information network of SMEs can capture the relationship among related enterprises and provide more comprehensive and reliable information for the credit risk measurement of SMEs. Then, we extracted meta-path features associated with SME based on the information network schema, which represents the situation of the SME credit risk. Finally, we developed three features to evaluate the effect of meta-path on SME credit risks. The experimental result shows that our proposed SME credit risk measuring method has a higher significance than the state-of-the-art features.

Data Availability

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

Disclosure

The presentation of the manuscript is used as Arxiv in [45].

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Project of Science and Technology Research and Development of China State Railway Group Co., Ltd. (K2020Z002).

References

[1] M. Gupta, P. Kumar, and B. Bhasker, "Heteclass: a meta-path based framework for transductive classification of objects in heterogeneous information networks," *Expert Systems with Applications*, vol. 68, pp. 106–122, 2017.

[2] Y. Sun and J. Han, "Mining heterogeneous information networks," *Acm Sigkdd Explorations Newsletter*, vol. 14, no. 2, pp. 20–28, 2013.

[3] C. Shi, Y. Li, J. Zhang, Y. Sun, and P. S. Yu, "A survey of heterogeneous information network analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 1, pp. 17–37, 2017.

[4] R. O. Edmister, "An empirical test of financial ratio analysis for small business failure prediction," *Journal of Financial and Quantitative Analysis*, vol. 7, no. 2, pp. 1477–1493, 1972.

[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] P. Hajek and K. Michalak, "Feature selection in corporate credit rating prediction," *Knowledge-Based Systems*, vol. 51, pp. 72–84, 2013.

[7] E. I. Altman, A. Giannozzi, O. Roggi, and G. Sabato, "Building sme rating: is it necessary for lenders to monitor financial statements of the borrowers?" *Bancaria*, vol. 10, pp. 54–71, 2013.

[8] J. Bauer and V. Agarwal, "Are hazard models superior to traditional bankruptcy prediction approaches? a comprehensive test," *Journal of Banking & Finance*, vol. 40, pp. 432–442, 2014.

[9] G. Sermpinis, S. Tsoukas, and P. Zhang, "Modelling market implied ratings using lasso variable selection techniques," *Journal of Empirical Finance*, vol. 48, pp. 19–35, 2018.

[10] M. Psillaki, I. E. Tsolas, and D. Margaritis, "Evaluation of credit risk based on firm performance," *European Journal of Operational Research*, vol. 201, no. 3, pp. 873–881, 2010.

[11] M. Bu, "Performance evaluation of enterprise supply chain management based on the discrete hopfield neural network," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 3250700, 2021.

[12] L. Lugovskaya, "Predicting default of Russian smes on the basis of financial and non-financial variables," *Journal of Financial Services Marketing*, vol. 14, no. 4, pp. 301–313, 2010.

[13] A. Moro and M. Fink, "Loan managers' trust and credit access for SMEs," *Journal of Banking & Finance*, vol. 37, no. 3, pp. 927–936, 2013.

[14] F. Mosteller and D. L. Wallace, "Inference in an authorship problem," *Journal of the American Statistical Association*, vol. 58, no. 302, pp. 275–309, 1963.

[15] E. H. Spafford and S. A. Weeber, "Software forensics: can we track code to its authors?" *Computers & Security*, vol. 12, no. 6, pp. 585–595, 1993.

[16] M. W. Akram, M. Salman, M. F. Bashir, S. M. S. Salman, T. R. Gadekallu, and A. R. Javed, "A novel deep auto-encoder based linguistics clustering model for social text," *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*vol. 2022, Article ID 3527838, 2022.

[17] A. Abbasi, A. R. Javed, F. Iqbal, Z. Jalil, T. R. Gadekallu, and N. Kryvinska, "Authorship identification using ensemble learning," *Scientific Reports*, vol. 12, no. 1, p. 9537, 2022.

[18] M. F. Tsai and C. J. Wang, "On the risk prediction and analysis of soft information in finance reports," *European Journal of Operational Research*, vol. 257, no. 1, pp. 243–250, 2017.

[19] C. Yin, C. Jiang, H. K. Jain, and Z. Wang, "Evaluating the credit risk of smes using legal judgments," *Decision Support Systems*, vol. 136, no. 113, p. 113364, 2020.

[20] E. Letizia and F. Lillo, "Corporate Payments Networks and Credit Risk Rating," *EPJ Data Science*, vol. 8, 2017.

[21] E. Tobback, T. Bellotti, J. Moeyersoms, M. Stankova, and D. Martens, "Bankruptcy prediction for smes using relational data," *Decision Support Systems*, vol. 102, no. oct, pp. 69–81, 2017.

[22] G. Kou, Y. Xu, Y. Peng et al., "Bankruptcy prediction for smes using transactional data and two-stage multiobjective feature selection," *Decision Support Systems*, vol. 140, Article ID 113429, 2021.

[23] 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.

[24] S. Angilella and S. Mazzù, "The financing of innovative smes: a multicriteria credit rating model," *European Journal of Operational Research*, vol. 244, no. 2, pp. 540–554, 2015.

[25] L. Cultrera and X. Brédart, "Bankruptcy prediction: the case of belgian smes," *Review of Accounting and Finance*, vol. 21, 2016.

[26] J. Gupta and A. Gregoriou, "Impact of market-based finance on smes failure," *Economic Modelling*, vol. 1, 2017.

[27] Y. Sun, Y. Yu, and J. Han, "Ranking-based clustering of heterogeneous information networks with star network schema," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 797–806, Paris, France, July 2009.

[28] Y. Sun, C. C. Aggarwal, and J. Han, "Relation strength-aware clustering of heterogeneous information networks with incomplete attributes," 2012, https://arxiv.org/abs/1201.6563.

[29] Q. Zhang, "A big data-driven approach to analyze the influencing factors of enterprise's technological innovation," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–14, Article ID 3785685, 2022.

[30] M. Ji, Y. Sun, M. Danilevsky, J. Han, and J. Gao, "Graph regularized transductive classification on heterogeneous information networks," in *Proceedings of the Machine Learning and Knowledge Discovery in Databases. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 570–586, Springer, Barcelona, Spain, September 2010.

[31] C. Wang, Y. Song, H. Li, M. Zhang, and J. Han, "Text classification with heterogeneous information network kernels," in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, Beijing China, June 2016.

[32] Y. Sun, J. Han, C. C. Aggarwal, and N. V. Chawla, "When will it happen? relationship prediction in heterogeneous information networks," in *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, pp. 663–672, Seattle, WA, USA, February 2012.

[33] A. Popescul and L. H. Ungar, "Statistical relational learning for link prediction," *IJCAI workshop on learning statistical models from relational data*, vol. 2003, p. 125, 2003.

[34] C. Shi, Z. Zhang, P. Luo, P. S. Yu, Y. Yue, and B. Wu, "Semantic path based personalized recommendation on

weighted heterogeneous information networks," in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, p. 453, 462 Melbourne, Australia, October 2015.

[35] H. Ma, I. King, and M. R. Lyu, "Learning to recommend with social trust ensemble," in *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 203–210, MA, Boston, USA, July 2009.

[36] M. Jamali and M. Ester, "A matrix factorization technique with trust propagation for recommendation in social networks," in *Proceedings of the Fourth ACM Conference on Recommender Systems*, pp. 135–142, Barcelona, Spain, September 2010.

[37] H. Ma, H. Yang, M. R. Lyu, and I. King, "Sorec: social recommendation using probabilistic matrix factorization," in *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, pp. 931–940, California, Napa Valley, USA, October 2008.

[38] H. Wang, F. Zhang, M. Hou, X. Xie, M. Guo, and Q. Liu, "Shine: Signed heterogeneous information network embedding for sentiment link prediction," in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pp. 592–600, CA, Marina Del Rey, USA, February 2018.

[39] A. Hosseini, T. Chen, W. Wu, Y. Sun, and M. Sarrafzadeh, "Heteromed: heterogeneous information network for medical diagnosis," in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pp. 763–772, Torino, Italy, October 2018.

[40] E. Zhong, F. Wei, Z. Yin, and Y. Qiang, "Modeling the dynamics of composite social networks," in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Illinois, Chicago, USA, August 2013.

[41] Y. Xiao, R. Xiang, Y. Sun, B. Sturt, and J. Han, "Recommendation in heterogeneous information networks with implicit user feedback," in *Proceedings of the Acm Conference on Recommender Systems*, Hong Kong, China, October 2013.

[42] B. Hu, Z. Zhang, C. Shi, J. Zhou, X. Li, and Y. Qi, "Cash-out user detection based on attributed heterogeneous information network with a hierarchical attention mechanism," in *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 946–953, Hawaii, Honolulu, USA, January 2019.

[43] C. Shi, X. Kong, Y. Huang, P. Yu, and B. Wu, "Hetesim: a general framework for relevance measure in heterogeneous networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 10, pp. 2479–2492, 2014.

[44] A. Ptak-Chmielewska, "Predicting micro-enterprise failures using data mining techniques," *Journal of Risk and Financial Management*, vol. 12, no. 1, p. 30, 2019.

[45] M. Du, Y. Ma, and Z. Zhang, "A meta path based evaluation method for enterprise credit risk," 2021, https://arxiv.org/abs/2110.11594.