Machine Intelligence Research
www.mi-research.net

21(2), April 2024, 257-271
DOI: 10.1007/s11633-023-1425-9

Corporate Credit Ratings Based on Hierarchical Heterogeneous Graph Neural Networks

Bo-Jing Feng* Xi Cheng* Hao-Nan Xu Wen-Fang Xue

Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

Abstract: In order to help investors understand the credit status of target corporations and reduce investment risks, the corporate credit rating model has become an important evaluation tool in the financial market. These models are based on statistical learning, machine learning and deep learning especially graph neural networks (GNNs). However, we found that only few models take the hierarchy, heterogeneity or unlabeled data into account in the actual corporate credit rating process. Therefore, we propose a novel framework named hierarchical heterogeneous graph neural networks (HHGNN), which can fully model the hierarchy of corporate features and the heterogeneity of relationships between corporations. In addition, we design an adversarial learning block to make full use of the rich unlabeled samples in the financial data. Extensive experiments conducted on the public-listed corporate rating dataset prove that HHGNN achieves SOTA compared to the baseline methods.

Keywords: Corporate credit rating, hierarchical relation, heterogeneous graph neural networks, adversarial learning.

Citation: B. J. Feng, X. Cheng, H. N. Xu, W. F. Xue. Corporate credit ratings based on hierarchical heterogeneous graph neural networks. *Machine Intelligence Research*, vol.21, no.2, pp.257–271, 2024. http://doi.org/10.1007/s11633-023-1425-9

1 Introduction

The financial risks of investing are relatively high if trading partners fail to deliver on their promises. Choosing an excellent corporation to invest in is the core issue of investment from the perspective of decision makers. How shall they evaluate the quality of corporations more accurately and efficiently before investment? Since the 21st century, many financial institutions have been using credit rating tools to estimate the likelihood[1] that potential trading partners will default on their debts[2]. Standard & Poor′s, Moody′s and Fitch are internationally renowned and authoritative credit rating agencies that monopolize the rating market in developed countries, while they lack ratings for Chinese companies. Chinese rating agencies, represented by Dagong global credit rating, CCXI and golden credit rating, started credit rating business relatively late. Their credit rating results are still based on the expert evaluation method, which leads to the problem of insufficient authority and waste of human resources.

Achieving high accuracy of corporate credit ratings is challenging[3] for the following reasons. First, the corporate financial reports data are usually used for credit analysis. But in fact, the financial reports are usually published several months after the current quarter. Therefore, the data mainly represents the historical information while lacking the reflection of the corporations′ current solvency. Second, a complex dynamic network is established between enterprises through various relationships. The asymmetric information transmission process in this network is difficult to dig. Specific examples will be given later. Third, the economic cycle changes over time according to Merrill Lynch Investment Clock[4]. It is difficult to model its substantial impact on corporate solvency. Moreover, the financial data of most small and medium-sized enterprises (SMEs) are unevenly distributed, the amount is insufficient, and some attributes are unfilled.

Credit rating is one of the problems of multivariate modeling, which is a mature research area of artificial intelligence[1]. Credit rating methods rely on Bayes theorem[5], transfer learning[6], data mining, machine learning, deep learning[7] and relevant economic theories, which can be mainly divided into three categories: statistical learning methods[8], integrated learning methods[9–11] and deep learning methods[12]. However, these machine learning models are unable to effectively capture the internal patterns of the extremely random and nonlinear financial data[8]. Deep learning, which requires massive structured data as training samples, has achieved great success in

Research Article
Special Issue on Commonsense Knowledge and Reasoning: Representation, Acquisition and Applications
Manuscript received on November 1, 2022; accepted on February 7, 2023; published online on January 12, 2024
Recommended by Associate Editor Zhi-Yuan Liu
Colored figures are available in the online version at https://link.springer.com/journal/11633
*These authors contribute equally to this work

computer vision (CV) and natural language processing (NLP)[13]. Unfortunately, it is difficult to obtain extensive structured data on credit ratings.

Nevertheless, the transmission of information through the relationships between corporations is not adequately modeled in the above models. For example, the net profit of oil companies will be low when the price of crude oil falls. Chemical companies that produce products from crude oil are downstream in the oil supply chain. Besides, the institutions that invest in these oil companies will also be under financial pressure. The heterogeneous relationships mentioned in the above examples include industry relationship, supply chain relationship and investment relationship. The solvency of the relevant corporations will decrease, and credit risk will increase. Graph neural network (GNN) with topological properties has the natural advantage of mining relationships among entities. Obviously, information is transferred through different relationships between entities in different ways. However, current researches applying GNN to credit ratings do not consider various heterogeneous relationships but treat them equally. In recent years, graph models for modeling heterogeneous relationships have been divided into two main parts: manually defining meta-paths models (such as MAGNN[14] and MEGAE[15]) and automatically discovering meta-paths models (such as GTN[16], HGT[17], and HetSANN[18]). However, these methods do not consider the hierarchy and heterogeneity of corporate features. This work is the first work to apply heterogeneous graph neural network to the field of credit rating. At present, the mainstream rating models first construct indicator system as input and then analyze the correlation between indicators and rating results. Due to the different main business sectors, the importance of the same indicator varies, e.g., fixed assets play a more important role in heavy industry than in retail. Different from the previous works that subjectively assign certain weights, we use the attention mechanism to assign different weights to different indicators. The information used for credit rating is hierarchically divided into high and low levels. The financial data of one corporation belongs to the low level information. The relationship data between companies, on the other hand, belongs to the high level information.

Since the targets of rating agencies are large companies, there are no rating labels for SMEs although they have financial data. To make better use of both labeled and unlabeled data, we introduce adversarial semi-supervised learning. First, we train a plain rating model (PRM) on the labeled data, and then we use PRM to generate pseudo-labels for the unlabeled data to construct adversarial examples. However, since the pseudo-labels are not completely accurate, we develop an adversarial discriminator to train hierarchical heterogeneous graph neural network (HHGNN) by minimizing the loss between the labeled and pseudo-labeled data. As the loss decreases, the accuracy of the pseudo-labels embedding also increases. In this way, HHGNN can leverage the large amount of unlabeled data.

Aiming at solving the problems encountered in the process of corporate credit rating research (incomplete modeling level, single corporate relationship, and insufficient use of the large amount of unlabeled data), we propose a new novel model HHGNN. The network mainly consists of the data processing module, the feature graph module, the heterogeneous corporation graph module, and the downstream task module. To utilize the massive unlabeled data, the data processing module first predicts pseudo-labels for unlabeled data, and the downstream task module uses the pseudo-label classifier to learn. The feature graph module models the internal interaction among features. And the corporate heterogeneous graph module models the information transformation through heterogeneous relationships (industry relationship, main business relationship, conceptual relationship and fund position relationship). An adaptive graph relation module is designed to bridge the gap between heterogeneous relationships we use and the real corporate relationships. The combination of the feature graph module and the corporate heterogeneous graph module reflects the hierarchical characteristics of HHGNN. The attention mechanism is used to merge the interaction of features and relationships respectively. Then, the adversarial discriminator is developed to reduce the information gap between the labeled data and the pseudo-labeled data. Therefore, the collected data is maximally utilized.

Our contributions can be summarized as follows:

1) **Algorithm.** HHGNN model is proposed with four outstanding innovations. First, hierarchical modeling can integrate different levels (feature level and enterprise level) together. Second, heterogeneous modeling could encode different asymmetric relationships and use attention mechanism to propagate information. Third, adversarial network makes full use of the massive unlabeled data for auxiliary tasks. Fourth, we design an adaptive graph block to supplement the information of heterogeneous relationships.

2) **Experiments.** Comprehensive experiments on the real public-listed corporate rating dataset demonstrate that HHGNN achieves state-of-the-art performance for corporate credit rating, increasing the accuracy, recall and F1-Score to 0.970 31, 0.975 57 and 0.978 82, respectively.

3) **Interpretability.** We conduct analysis experiments from four aspects: the choice of pseudo-label rating module, unlabeled data volume, the knowledge learned by feature graph module, and corporate heterogeneous graph module, that increases the interpretability of HHGNN.

The remainder of this paper is organized as follows. Section 2 presents a brief literature review on the related works. The general structure and strategy of HHGNN is described in detail in Section 3. The experiment and interpretability analysis are presented in Section 4. Finally, the conclusions of this paper are summarized in Section 5.

2 Related works

Traditionally, credit rating experts spend a few months researching the pecuniary condition of companies, which requires a great deal of human, material, and financial resources. Since the mid-19th century, many financial institutions and scholars have designed various credit risk evaluation models[19] in order to simplify the complexity of credit rating process. These models can be broadly classified into three categories: statistical learning methods, integrated learning methods and deep learning methods.

2.1 Statistical learning methods

Statistical learning methods are based on the probabilistic credit rating models, which can be mainly divided into statistical inference method[8] and statistical classification method[20]. Regression analysis is a classical method of statistical inference[21], including linear regression, logistic regression, probit regression and so on. Classical statistical classification algorithms include discriminant analysis, classification tree method[22], nearest neighbor method and so on. Statistical learning is a relatively complete theoretical system and is easy to operate with a clear rating conclusion. However, the statistical theories impose too many preconditions and constraints, which leads to the limited scope of the rating in practice.

2.2 Integrated learning methods

The development of credit rating models based on integrated learning can be divided into two phases[23–25]: The integration of statistical learning methods and the integration of deep learning methods[26]. Specifically, it includes random forest method[27], shallow neural network combined with regression model[28], neural network combined with genetic algorithm, support vector machine and fuzzy calculation, etc. In addition, different models are usually used to model different datasets through integration learning, which can alleviate the imbalance rating category problem[29–31].

2.3 Deep learning methods

Compared with the method of constructing features by artificial rules, the deep learning method can better use large amounts of data to mine the implicit relationships between features and labels through appropriate representation of financial conditions. Deep learning methods applied to corporate credit rating[32] include: deep confidence network, deep convolution neural network[26, 33], deep recursive neural network, long and short term memory network[34] and so on.

GNN is a recently widely used deep learning technology that combines the advantages of deep learning and graph models. It extracts knowledge through topological structure and uses deep learning between the relationships among data, reducing the dependence on the quantity and completeness of credit rating samples[35–40]. Scientific SMEs credit data has the characteristics of short dimension, insufficient quantity and unfilled features, GNN demonstrates its significant advantages in the application of corporate credit evaluation of SMEs.

The emerging success of heterogeneous graph neural networks in recent years shows their advantages in exploring heterogeneous graphs whose nodes and edges belong to different types. The existing heterogeneous graph models[14, 15] usually define multiple meta-paths to capture composite relationships, and then transform heterogeneous graphs into homogeneous graphs through meta-paths. MAGNN[14] is composed of node content transformation, intra-metapath aggregation, and inter-metapath aggregation. It considers multiple metapaths and designs several encoders for distilling information from metapath instances. Metapath enhanced graph attention encoder (MEGAE)[15] uses graph attention encoder to learn graph structural information and provides interpretability. However, manually defining meta-paths requires a lot of expert knowledge. Graph transformer network (GTN)[16] automatically discovers efficient meta-paths and provides explanations for meta-path connectivity. Heterogeneous graph transformer (HGT)[17] automatically learns the most important meta-paths for downstream tasks. HGT presents the relative temporal encoding technique to capture graph dynamics. The heterogeneous mini-batch graph sampling algorithm extends the training of HGT to web-scale data. HetSANN[18] does not use meta-paths but encodes heterogeneous information directly. It maps the transformation relationships between different node types in a low-dimensional entity space and then uses GNN and an attention mechanism to aggregate information about multiple relationships between neighboring nodes. However, the heterogeneous graph models are rarely applied to credit rating tasks.

The above credit rating models do not take the hierarchy, heterogeneity and unlabeled data into account. First of all, at present some graph models usually regard corporations as nodes and the relationships among corporations as edges in the graph, lacking feature level modeling. In the research of feature graph, features are used as nodes, and the relationships among features are used to build the graph edges[41], without considering the relationships among corporations, e.g., in the heterogeneous graph models mentioned above. These models are widely used in the knowledge graph domain, but have never been applied to the field of credit rating. In addition, the heterogeneous graph models only propagate information based on the heterogeneous relationship between nodes, but do not represent the information interaction inside nodes. However, the information interaction inside the nodes represents the mutual restriction and influence of the financial data inside the corporation, which cannot be ignored for the problem of credit rating. Secondly, some current methods model only one kind of relationship separately in the process of building corporate graph, such as supply chain relationship between corporations. However, the complicated relationships among corporations in the real world lead to the difficulty of modeling the corporation-to-corporation networks by only one relationship. Finally, many credit data are unlabeled. The reliance on big data poses significant challenges to deep learning methods. HHGNN provides a new way to address these shortcomings.

3 The proposed method: HHGNN

In this section, we introduce the proposed framework: HHGNN.

3.1 Problems statement

In this paper, the entire input financial dataset is represented by the bold capital letter C. C_L is composed of two parts C_L and C_U, where $C_L=\{c_1,c_2,\cdots,c_L\}$ denotes the set of corporations with rating labels and $C_U=\{c_1,c_2,\cdots,c_U\}$ denotes the set of corporations without rating labels. Each sample $c\in\mathbf{R}^d$ in the dataset C_L represents the data of a company, where d represents the feature dimension. Generally speaking, the number of SMEs without rating labels in actual application scenarios is much larger than the number of large corporations with labels, i.e., $U\gg L$.

In dataset C_L, every corporation sample c has a corresponding rating label, e.g., A, B, C, etc. We use $Y=\{y_1,y_2,\cdots,y_m\}$ to represent all rating outcomes, and m represents the number of rating levels. The objective of the whole credit rating model is to predict the credit rating level of the corporations based on dataset C. The prediction result will be represented by a probability distribution $\hat{y}$, where the label with the highest probability is the final predicted credit rating result.

3.2 Design overview

Our goal is to alleviate the above problem and make predictions about the corporate credit rating by hierarchical heterogeneous graph with adversarial learning. We have developed HHGNN for such specific scenarios. The framework of HHGNN is shown in Fig. 1. From left to right, the model is mainly composed of four modules: data processing module, feature graph module, corporate heterogeneous graph module and final downstream task module. Adversarial semi-supervised learning is realized by the data processing module and the downstream task module. The combination of the feature graph network and the corporate heterogeneous graph network shows the hierarchy of financial data representation. The corporate heterogeneous graph network mainly models the heterogeneity of relationships between companies through adaptive graph learning.

1) **Data processing module**

To create a dataset suitable for training, the data processing module performs preprocessing on the original dataset, including data cleaning, data feature selection, and data feature normalization. In order to make better use of unlabeled data, we first train a pre-classifier with fast fitting speed and good generalization performance such as GBDT, XGBoost, etc. This simple pre-classifier makes predictions on huge amounts of unlabeled data while generating pseudo-labels. Although a distinction is made between pseudo-labels and real labels in the data processing module, this has no effect on the subsequent

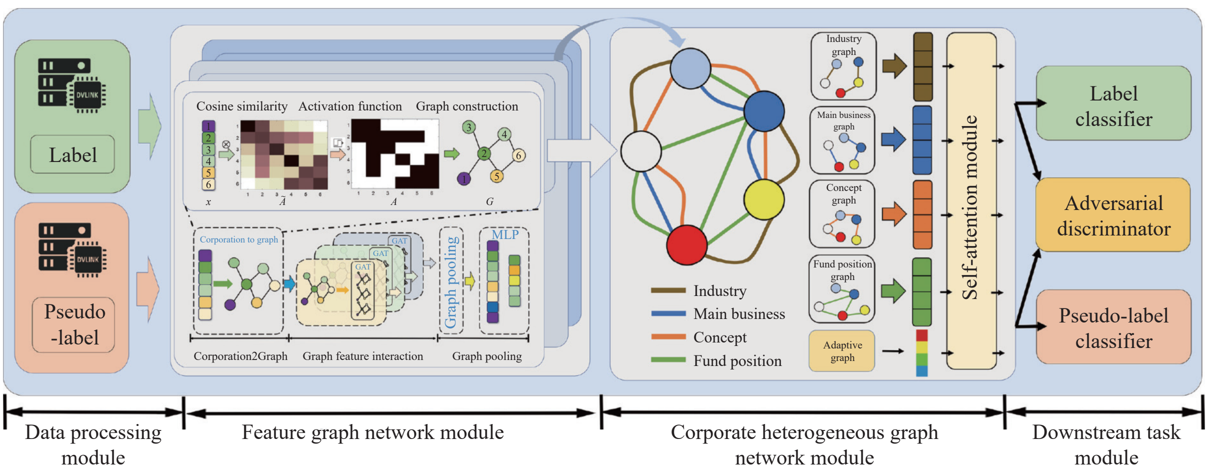

Fig. 1 The framework of HHGNN

modules (feature graph module and corporate heterogeneous graph module). The subsequent modules do not know whether the current input is a real or a pseudo-labeled sample.

2) **Feature graph module**

After passing through the data processing module, the clean and tidy data goes into the feature graph module. This module consists of three sublayers: graph transformation layer, graph feature interaction layer, and graph pooling layer. Feature data of each corporation is first converted into an intra-corporate feature graph in the graph transformation layer. One intra-corporate feature graph represents the internal relationships inside the corporation. In the graph feature interaction layer, the intra-corporate feature graph network uses the powerful aggregation ability of GNN to simulate the feature interaction. The features are combined according to the topological relationship of the graph, and then we obtain the interactive feature node representations. Finally, in the graph pooling layer, we use the pooling method to obtain the final embedding of the entire intra-corporate feature graph.

3) **Corporate heterogeneous graph module**

Each intra-corporate feature graph is represented as one embedding in the previous module, which is taken as the initial representation of each node in the corporate heterogeneous graph module. In this module, nodes represent corporations and edges represent various heterogeneous relationships (industry relationship, main business relationship, concept relationship, fund position relationship, and graph adaptive relationship). GNN models the information interactions among corporations through various heterogeneous relationships and outputs the representation aggregated by the corresponding topological graph network.

4) **Downstream task module**

The previous modules (feature graph module and corporate heterogeneous graph module) have already modeled feature interaction within one corporation (intra-corporate feature graph) and feature interaction between corporations (inter-corporate feature graph). The final downstream task module needs to make full use of the extracted feature representation. There are three components in this module, the label classifier, the pseudo-label data classifier, and the adversarial discriminator. The label classifier and the pseudo-label classifier are used to predict the rating results of the labeled and pseudo-labeled data respectively. The labels fed to the label classifier are exactly correct. However, for the pseudo-label classifier, the pseudo-labels given by the data processing module are partially wrong. Therefore, we design the adversarial discriminator, which is used to identify whether the current sample is from a real labeled sample or a pseudo-labeled sample.

3.3 Data processing module

The data processing module can be roughly divided into two stages: the data preprocessing stage and the pseudo-label generation stage.

3.3.1 Data preprocessing stage

In the data preprocessing stage, the collected raw data is cleaned. The samples with more than 90% missing features and the features with more than 95% missing values will be discarded. We use the industry mean of the feature to fill in the remaining missing values. Then, the data is standardized and regularized. We use the Z-score standardization method to avoid the influence of outliers. The sample data c subtracts the mean of the dataset $\mu = \frac{1}{|C|}\sum\limits_{c_i \in C} c_i$ and is divided by the standard deviation $\sigma = \sqrt{\frac{1}{|C|}\sum\limits_{c_i \in C} \|c_i - \mu\|^2}$. It is expressed as $c = \frac{c-\mu}{\sigma}$.

In addition, the data should be regularized and normalized. For each sample c, we first calculate the Lp-norm of each sample c and then map it to 1 as described in (1) and (2).

$$L_p(c) = \left(|c^{(1)}|^p + |c^{(2)}|^p + \cdots + |c^{(d)}|^p\right)^{\frac{1}{p}} \tag{1}$$

$$c = \frac{c}{L_p(c)}. \tag{2}$$

3.3.2 Pseudo-label generation stage

In the pseudo-label stage, we first train a plain rating model (PRM) using the labeled data and then use the PRM to make predictions on the unlabeled data. In theory, any corporate credit rating model can be used as a PRM, such as logistic regression, multi-layer perceptron, GBDT, and XGBoost. However, in practical applications, we need to make a balance between the complexity and the performance of the model to select an appropriate PRM. The best approach is to choose a simple model that is easy to train but performs relatively well. We do an exhaustive experimental analysis for different pseudo-label models, as detailed in Section 4.3. After this module, all samples have their corresponding labels.

3.4 Feature graph module

The feature graph module simulates the information interaction of features inside the corporations, which can be divided into three layers: corporation to graph layer, graph feature interaction layer, and graph pooling layer.

3.4.1 Corporation to graph layer

$c \in \mathbf{R}^{d\times 1}$ represents the features of one corporate. The embedded vector of the corporation $x \in \mathbf{R}^{d\times d\_hid}$ is obtained through an embedding layer $x = Sigmoid(cW + b)$, where $W \in \mathbf{R}^{1\times d\_hid}, b \in \mathbf{R}^{d\times d\_hid}$ are learning model parameters, d_hid is the dimension of the embedded layer and $Sigmoid$ is a nonlinear activation function.

After the embedding layer, we use the cosine similarity to calculate the interaction matrix $\tilde{A} \in R^{d\times d}$ as shown in (3). $A \in [-1, 1]$, and 0 represents a weak similarity and ± 1 represents a strong similarity. This is the core design of the corporation to graph layer, that is, modeling the efficient interaction of internal features of the corporations. Recent works have shown that explicitly modeling feature interactions is more meaningful and concatenation is suboptimal[42–44].

$$\tilde{A} = cos\_similarity(x, x) = \frac{xx^{\mathrm{T}}}{\|x\|\|x\|}. \tag{3}$$

Not all information in the interaction matrix $\tilde{A}$ is necessary, so we use the activation function $g(\cdot)$ to extract the obvious interactions of the internal feature of the corporations and ignore some trivial information. Equation (4) shows the detail. $A \in \mathbf{R}^{d\times d}$ is the output matrix. A_{ij} represents the elements in row i and column j of the matrix A and r is an adaptive parameter.

$$A_{ij} = g\left(\widetilde{A}_{ij}\right) = \begin{cases} 0, & \text{if } \widetilde{A}_{ij} < r \\ 1, & \text{if } \widetilde{A}_{ij} \geq r. \end{cases} \tag{4}$$

Adjacency matrix A is used to build corporation feature graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. $\mathcal{V} = \{v_1, v_2, \cdots, v_d\}$ represents the nodes of the graph, and every corporation feature graph contains d feature nodes. If $A_{ij} = 1$, then we construct one edge between v_i and v_j. The entire corporate feature graph is constructed at this step.

3.4.2 Graph feature interaction layer

After the corporation to graph layer, we use GNN to simulate the interactive influence between different features within the corporation. Since the attention point is different for features, we introduce the graph attention networks (GAT) to automatically utilize the attention mechanism to model the interaction between features. In addition, since the interaction between financial features is not limited to the neighboring features, the graph feature interaction layer stacks more GAT layers to aggregate information from farther graph neighbors, which models higher-order feature interactions.

$$X_i^{(l)} = \alpha_{i,i}^{(l)}\Theta^{(l)}(X_i^{(l-1)} + X_i^{(l-2)}) + \sum_{j\in\mathcal{N}_{(i)}} \alpha_{i,j}^{(l)}\Theta^{(l)}(X_j^{(l-1)} + X_j^{(l-2)}). \tag{5}$$

Specifically, in GAT, the update of nodes v_i in graph $\mathcal{G}$ can be described as (5), where l represents the layer number, $\Theta^{(l)} \in \mathbf{R}^{d^{(l)}\times d^{(l-1)}}$ and $a \in \mathbf{R}^{2d^{(l)}}$ are the learnable parameters of GAT, $x^{(l-1)} + x^{(l-2)} \in \mathbf{R}^{d\times d^{(l-1)}}$ is the input of GAT. To alleviate the gradient vanishing problem and speed up the model training, we add the residual to the input. $x^{(l)} \in \mathbf{R}^{d\times d^{(l)}}$ is model output. The calculation process of the attention coefficient $\alpha_{i,j}^{(l)}$ is shown in (6), where $k \in \mathcal{N}(i) \cup \{i\}$ and LR means LeakReLU function.

$$\alpha_{i,j}^{l} = \exp\left(LR\left(a^{T^{(l)}}\left[\Theta^{(l)}X_i^{(l-1)} \parallel \Theta^{(l)}X_j^{(l-1)}\right]\right)\right) \div \sum_k \exp\left(LR\left(a^{T^{(l)}}\left[\Theta^{(l)}X_i^{(l-1)} \parallel \Theta^{(l)}X_k^{(l-1)}\right]\right)\right) \tag{6}$$

$$X_{fg} = Sig\left(concat(X^{(0)}, \cdots, X^{(l)})W^{fg} + b_{fg}\right). \tag{7}$$

It is worth noting that there are no residuals but the original feature $X^{(0)}$ in the input of the first layer. The feature embedding matrix of all nodes after the final interaction is shown in (7). $X_{fg} \in \mathbf{R}^{d\times d\_hid}$ is the output, Sig means Sigmoid function. $W_{fg} \in \mathbf{R}^{\sum_{l\in L} l\times d\_hid}$ and $b_{fg} \in \mathbf{R}^{d\times d\_hid}$ are the trainable model parameters. Higher-order financial feature interactions inside the corporation have been modeled in neural networks with stacked GAT layers.

3.4.3 Graph pooling layer

The representation of the whole graph is obtained in the graph pooling layer. Specifically, there are two types of pooling, max graph pooling and average graph pooling. The output of the graph pooling layer $x_{fg} \in \mathbf{R}^{d\_hid}$ is not only the graph representation of the internal features of the corporation but also the node's initial representation of the next module. This shows the hierarchical characteristics of HHGNN.

3.5 Corporate heterogeneous graph module

After the features of the same corporation interact sufficiently, the information is passed to a higher level of the network, i.e., the corporate heterogeneous graph level. In such a module, the information interactions through heterogeneous relationships among corporations are modeled.

There are a variety of complex relationship networks between corporations, such as industry relationship, main business relationship, concept relationship, fund position relationship, and so on. However, no matter how much prior knowledge and data are collected, it is impossible to model the inter-corporate relationships completely. Therefore, we add an adaptive graph module as a supplement. Finally, use the self-attention module to fuse multiple heterogeneous relations. In this section, set $\mathcal{R} = \{r_{id}, r_{bs}, r_{cc}, r_f, r_{ad}\}$ represent the heterogeneous relationships we used. We use id, bs, cc, fd and ad to represent industry relationship, main business relationship, concept relationship, fund position and adaptive relation for convenience.

3.5.1 Heterogeneous relationships

A variety of heterogeneous relationships among corporations constitute a complex graph. This paper collects some of the main heterogeneous relationships between corporations for research, such as industry relationship, main business relationship, concept relationship and fund position relationship.

1) **Industry relationship.** The most important relationship between corporations is the industry relationship, which refers to the national economic industry category that corporations belong to. This largely determines whether the relationship between corporations is cooperation or competition. For example, Tencent and Alibaba both belong to the information technology industry and compete with each other. According to the industry standards stipulated in the 2017 National Economic Industry Classification (GB/T 4754-2017), this research divided the collected companies into 20 categories.

2) **Main business relationship.** The main business refers to the daily business activities of the company. The main business relationship is more refined compared with the industry relationship.

3) **Concept relationship.** Wind forms 135 concept indices. For example, the Internet finance index represents the overall development trend of companies engaged in the Internet finance business.

4) **Fund position relationship.** More than 80% of the fund is invested in stocks. First of all, we establish the bipartite graph of the fund and the stocks according to whether the stocks are held by the fund. Then we establish the fully connected graph network of the stocks belonging to the same fund, to form the fund position relationship graph.

If two companies are in the same industry, then corporate heterogeneous graph creates an edge between them to form the graph. The construction process of the main business relationship graph and concept relationship graph is the same as above. $\mathcal{G} = \{\mathcal{G}_{id}, \mathcal{G}_{bs}, \mathcal{G}_{cc}, \mathcal{G}_{fd}\}$ represents the corporate heterogeneous graph, the corresponding adjacency matrix is $\mathcal{A} = \{\mathcal{A}_{id}, \mathcal{A}_{bs}, \mathcal{A}_{cc}, \mathcal{A}_{fd}\}$, where $\mathcal{A} \in \mathbf{R}^{n_c \times n_c}$ is a symmetric bool matrix.

In this paper, we use GAT to model the interaction between corporations, and the residual block is added after the interaction. The specific calculation process is shown in (8),

$$\begin{cases} X_{id} = AF\left(GAT\left(AF\left(GAT(X_{fg}, \mathcal{A}_{id})\right) + X_{fg}, \mathcal{A}_{id}\right)\right) \\ X_{bs} = AF\left(GAT\left(AF\left(GAT(X_{fg}, \mathcal{A}_{bs})\right) + X_{fg}, \mathcal{A}_{bs}\right)\right) \\ X_{cc} = AF\left(GAT\left(AF\left(GAT(X_{fg}, \mathcal{A}_{cc})\right) + X_{fg}, \mathcal{A}_{cc}\right)\right) \\ X_{fd} = AF\left(GAT\left(AF\left(GAT(X_{fg}, \mathcal{A}_{fd})\right) + X_{fg}, \mathcal{A}_{fd}\right)\right) \end{cases} \tag{8}$$

where $X_{fg} \in \mathbf{R}^{n_c \times d\_hid}$ is the output of the previous module, AF is the active function implemented by $Sigmoid$ or $ReLU$. The outputs $\{X_{id}, X_{bs}, X_{cc}, X_{fd}\} \in \mathbf{R}^{n_c \times d\_hid}$ are the node embeddings under the heterogeneous relationships.

3.5.2 Adaptive graph block

The adaptive graph block complements the four kinds of relationships in the corporate heterogeneous graph, which is a fully connected graph network. It uses attention mechanism to automatically learn relation aggregations, which is calculated in a similar way to the corporate heterogeneous graph as shown in (9).

$$X_{ad} = AF\left(GAT\left(AF\left(GAT(X_{fg}, \mathcal{A}_{ad})\right) + X_{fg}, \mathcal{A}_{ad}\right)\right) \tag{9}$$

It is worth noting that we used GAT twice in the HHGNN model, once in the graph feature interaction layer of feature graph module, and the other time in the corporate heterogeneous graph module. In the first time, the GAT graph represents one corporation, its node represents one feature of that corporation, and the information interaction is only between the internal features of that corporation. In the second time, the graph of GAT represents all corporations, its nodes represent one corporation, and the information interaction is among all corporation nodes.

3.5.3 Self-attention block

This module uses the attention mechanism to fuse the embeddings of the aforementioned four heterogeneous graphs $(x_{id}, x_{bs}, x_{cc}, x_{fd}, x_{ad})$ and the embeddings of the adaptive graph $(x_{adaptive})$, which can be described as $x_{cg} = \sum_{r \in \mathcal{R}} \beta_r x_r$. The attention coefficient β is calculated as follows,

$$\beta_r = \frac{\exp\left(Sigmoid\left(x_r W_{att} + b_{att}\right)\right)}{\sum_{R \in \mathcal{R}} \exp\left(Sigmoid\left(x_R W_{att} + b_{att}\right)\right)} \tag{10}$$

where W_{att} and b_{att} are the trainable parameters. We combine the output of corporate heterogeneous graph module with the output of the previous layers (the original embedding and the output of the feature graph module) to obtain the final fused representation of all levels, as shown in (11), where W_{out} and b_{out} are the trainable parameters, the final output $x_{out} \in \mathbf{R}^{d\_hid}$ contains hierarchical information.

$$x_{out} = Sigmoid\left(Concat(x, x_{fg}, x_{cg}) W_{out} + b_{out}\right). \tag{11}$$

3.6 Downstream task module

After the upstream module, we can obtain the labeled data embedding and the pseudo-labeled data embedding. The two types of embedding are fed into the labeled classifier and the pseudo-labeled classifier respectively to obtain the final rating prediction results, and the prediction results are compared with the true label and the pseudo-label respectively to obtain two types of loss. However, since the pseudo-labels are not completely accurate, we design an adversarial discriminator. Its role is to determine whether the current sample is a labeled sample or a pseudo-labeled sample. In other words, the adversarial discriminator is essentially a binary classifier, which is used to evaluate the credibility of the current sample. The input of the adversarial discriminator is the labeled data embedding and the pseudo-labeled data embedding which is output by the upstream module, and the output is 1 or 0, which means whether the sample is a pseudo-labeled sample or not. By adversarial discriminator, we obtain the third type of loss. With the training of HHGNN, it is more and more difficult for the adversarial discriminator to distinguish whether the sample is a pseudo-labeled sample or not, and the embedding of the pseudo-labeled sample is more accurate. The credibility of the pseudo-labels and the discrimination between real and pseudo-labels are achieved in the last module, and this information is fed back to the model by the adversarial discriminator. In other words, in addition to the labeled samples, the model also obtains relatively accurate labels for the unlabeled data. So even though the pseudo-labels are fake, the model still gets more information and this kind of data processing can improve model performance.

For convenience, the model output of labeled dataset C_L and pseudo-labeled dataset C_U are represented by C_{Lout} and C_{Uout} respectively. The final loss function $\mathcal{L}$ consists of three parts: the label classifier loss $\mathcal{L}_L$, the pseudo-label classifier loss $\mathcal{L}_U$, and the adversarial discriminator loss λ_{AD}, as shown in (12).

$$\mathcal{L} = \mathcal{L}_L + \lambda_1 \mathcal{L}_U + \lambda_2 \mathcal{L}_{AD}. \tag{12}$$

The regularization parameters λ_1 and λ_2 modulate the influence degree of the pseudo-label classifier and the adversarial discriminator. In addition, due to the class imbalance of corporate credit ratings, we use focal loss to alleviate such a situation.

Since the previous modules have embedded sufficient feature interactions and corporate interactions, we use a simple multilayer perceptron model (MLP) to implement the label classifier. The probability distribution of the classification is shown in (13).

$$\hat{Y}_L = Softmax(Label\ Classifier\ (C_{Lout})). \tag{13}$$

The label classifier loss function with focal loss is given in (14), where α and γ are the parameters of focal loss, Δ_L is the parameter of the label classifier, λ_L is its corresponding regularization parameter. The pseudo-label classifier is constructed in the same way as the label classifier.

$$\mathcal{L}_L = -\frac{1}{|C_L|}\sum_{i=1}^{|C_L|}\prod_{j=1}^{m}\alpha\,(1-Y_L(i,j))^{\gamma}\log\left(\hat{Y}_L(i,j)\right) + \lambda_L\|\Delta_L\|^2. \tag{14}$$

The adversarial discriminator $Dis(\cdot)$ is essentially a binary classifier. We use MLP with $Sigmoid$ as the activation function of the last layer. The adversarial loss function is as (15).

$$\mathcal{L}_{AD} = -\frac{1}{|C_L|}\sum_{C_L}\log\left(Dis(C_{Lout})\right) - \frac{1}{|C_U|}\sum_{C_U}\log\left(1-Dis(C_{Uout})\right) + \lambda_{AD}\|\Delta_{AD}\|^2. \tag{15}$$

HHGNN is trained by the final loss function $\mathcal{L}$. However, in the prediction stage, the final classification probability distribution is obtained from the output of the label classifier.

4 Experiment

In this section, we aim to answer the following five questions:

RQ1. Could HHGNN produce a higher accuracy of corporate credit rating compared with the previous approaches?

RQ2. What is the impact on HHGNN when using different PRMs to make pseudo-label?

RQ3. How does the amount of unlabeled data affect model performance?

RQ4. How do features interact with each other in the feature graph module?

RQ5. How the heterogeneous relations bring benefits to the model?

4.1 Experimental configurations

Dataset. We collect corporate annual report data from China stock market & accounting research database (CSMAR) as the financial features of corporate rating database. The dataset contains 39 features and most of them are financial data, which can be divided into six aspects: profitability (e.g., net income), operating ability (e.g., inventory turnover), growth ability (e.g., relative assets), debt serviceability (e.g., assets and liabilities), cash flow ability (e.g., interest) and Dupont index (e.g., Dupont return on equity). Non-financial features include tax credit rating and so on. The rating label is obtained from CCXI, China lianhe credit rating (CLCR), etc. There are 9 labels as the rating results: AAA, AA, A, BBB, BB, B, CCC, CC, C, and their grades are decreasing. The final distribution of the label is shown in Fig. 2.

Obviously, the distribution of rating data is unbal-

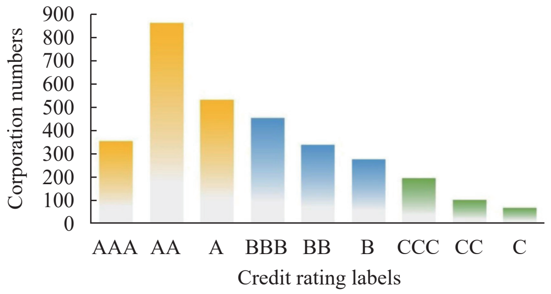

Fig. 2 Data distribution of credit rating

anced. Previous researchers used SMOTE[45] method to augment data[26], and some researchers proposed comparative self-supervised methods[46] for such problems. This study uses focal loss, which is effective and simple in industrial practice, and the specific implementation is presented in Section 3.6.

Hyperparameter setup. We split 80% of the labeled dataset C_L as the training set and 20% as the test set. The selection of the amount of unlabeled data C_U is controlled by the hyperparameter γ. We set the dimension of the hidden layer $d\_hid = 64$ and use the Xavier uniform distribution for parameter initialization[47]. The loss function of the model is optimized using the Adam optimizer of which the initial learning rate is set to 0.001, and the decay rate of the learning rate is set to 0.000 01. The regularization parameter for the pseudo-label classifier and the adversarial discriminator is set to 0.5 and 0.3 respectively.

Baselines. We evaluate the models on seven baselines. The machine learning models include logistic regression (LR), support vector machine (SVM), multi-layer perceptron (MLP), XGBoost[48]. The four deep learning credit rating models are described as follows:

1) CCR-CNN[26]: A corporate credit rating model based on convolutional neural network.

2) CCR-GNN[41]: A corporate credit rating model based on basic GNN.

3) ASSL4CCR[49]: A corporate credit rating model based on semi-supervision learning with an adversarial method, which aims to utilize unlabeled data.

4) HGT[17]: A heterogeneous graph model with a heterogeneous attention mechanism which does not require manually designed meta-paths.

Three evaluation metrics accuracy, recall, and F1-score are used to evaluate model performance.

4.2 Improvement with baseline methods (RQ1)

The overall performance of experiments is displayed in Table 1. All deep learning models perform better than machine learning models. HHGNN proposed in our work significantly exceeds the previous deep learning models in recall, accuracy, and F1-score, with the results all greater

Table 1 The performance of HHGNN compared with other baselines

Model type	Model name	Recall	Accuracy	F1-score
Machine learning	LR	0.762 50	0.809 70	0.819 46
	SVM	0.837 50	0.892 47	0.889 61
	MLP	0.914 06	0.935 68	0.932 54
	<u>XGBoost</u>	<u>0.923 43</u>	<u>0.942 25</u>	<u>0.941 33</u>
Deep learning	HGT	0.926 56	0.946 80	0.941 46
	CCR-CNN	0.928 12	0.952 53	0.945 18
	CCR-GNN	0.934 37	0.950 12	0.951 77
	<u>ASSL4CCR</u>	<u>0.953 21</u>	<u>0.961 15</u>	<u>0.962 52</u>
	HHGNN_Base	0.954 69	0.961 88	0.960 99
	HHGNN	**0.970 31**	**0.975 57**	**0.978 82**

than 0.97. One possible reason is that HHGNN not only integrates the feature graph network and the massive unlabeled data but also creatively constructs heterogeneous relationship graphs of corporations supplemented by an adaptive graph. The model performance improvement reflects the powerful feature extraction and representation ability of HHGNN. In addition, the parameter quantity of HHGNN model is only 22 M, which is roughly the same as HGT, 1/12 of CCR-CNN and 1/4 of CCR-GNN. The attention mechanism brings strong interpretability to HHGNN, which is lacking in ASSL4CCR.

4.3 Impact of different PRMs (RQ2)

Any rating model can be used as a PRM in theory. However, we should not choose a model with high complexity in the application which will make the training time too long. Vice versa, we should not choose a model that has fast inference speed but poor performance. Therefore, we select three relatively simple models that are less difficult to train but have good performance: GBDT, MLP, and XGBoost, and then do experiments for detailed analysis. The results are shown in Table 2.

Table 2 The influence of different PRMs on HHGNN

PRM	Recall	Accuracy	F1-score
MLP	0.964 06	0.969 52	0.973 24
XGBoost	0.968 75	0.978 13	0.974 72
GBDT	0.970 31	0.975 57	0.978 82

The following conclusions can be drawn from Table 2. First, the performance of HHGNN is enhanced with either PRM. This fully reflects the auxiliary effect of unlabeled data for labeled supervised model training. Second, among the three PRMs, HHGNN constructed by XGBoost as PRM achieves the best accuracy. However, the HHGNN constructed by GBDT as PRM achieves the best results in recall and F1-score. Since XGBoost algorithm has been further improved compared with GBDT and the model is more complex, GBDT is subsequently used as the PRM to make pseudo-labels according to the principle that training could be easy but the model performance is supposed to be good.

4.4 Amount of unlabeled data (RQ3)

To measure the effect of unlabeled data, only labeled data was used to train HHGNN_base in the ablation experiment. The comparison results are shown in Table 1. We can infer that, only using the labeled data, HHGNN_base has already achieved the best performance in accuracy and recall and is the second to ASSLCCR model in F1-score.

The experiment also tests the impact of different amounts of unlabeled data. The parameter γ is to control the proportion of unlabeled data which is defined as $|C_U| = \gamma|C_L|$. We feed HHGNN with various amounts of unlabeled data and all labeled data. The logarithmic line charts shown in the Fig. 3 illustrate the variation of recall, accuracy, and F1-score with the parameter γ.

The addition of unlabeled data makes the distribution of data closer to the real world and improves the performance of the model. The overall impact of the amount of unlabeled data on the performance of HHGNN shows an "inverted V" shape. When $\gamma \approx 1$, the model performance reaches the optimal state. When the amount of unlabeled data is much larger than labeled data, i.e., $\gamma \geq 10$, some of the pseudo-labels will be wrong, which increases the training difficulty.

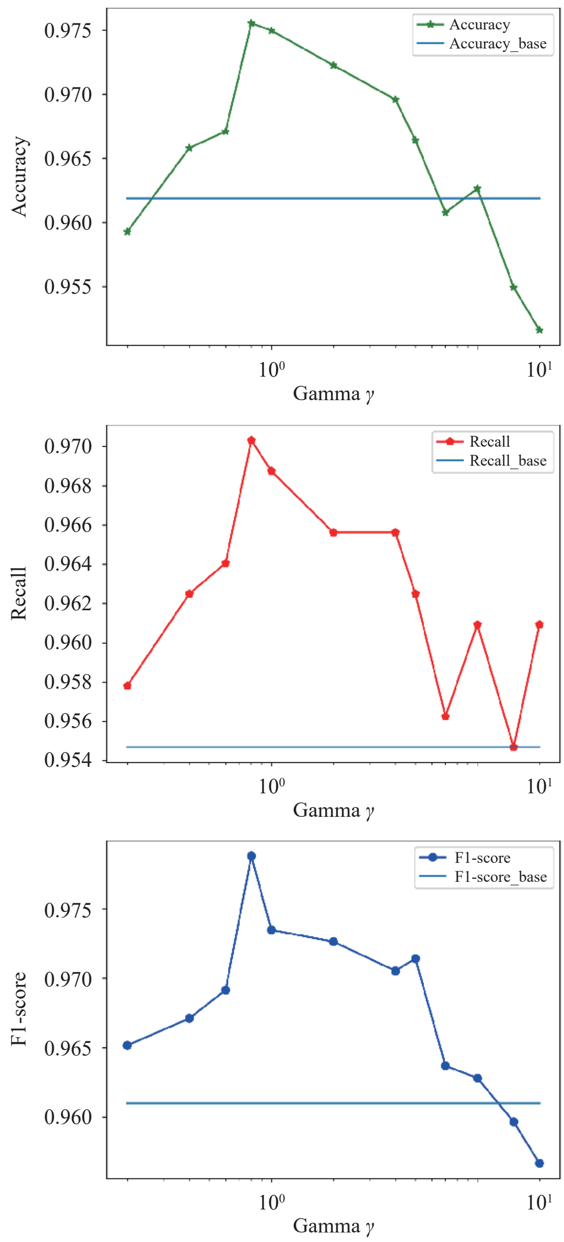

Fig. 3 Influence of the amount of unlabeled data

4.5 Interaction of feature graph networks (RQ4)

In this section, we analyze the interaction process at the feature level by visualizing specific examples, which brings a degree of interpretability to HHGNN. We feed the test dataset into HHGNN for prediction and observe its similarity matrix and adjacency matrix during feature-level inference, i.e., in feature graph module. According to the different interaction ways of features, corporations can be roughly divided into the following three types: Radial type, grid type, and hybrid type, as shown in Fig. 4.

Radial type. The similarity matrix and interaction matrix of such enterprises are divided into four parts: Upper left, upper right, bottom left, and bottom right. From the upper left and bottom right parts, we can see that the features of the radial type corporations are divided into two types. They have high internal similarity respectively, so the graph network is relatively dense. These two types of features interact internally first, and then the high-level interaction takes place between them subsequently (shown as the bottom left and bottom right part). From the analysis of the similarity matrix and the interaction matrix, we can see that HHGNN fully models the interactive characteristics of the radial type corporations' features.

Grid type. Compared with the radial type corporations, the similarity and interaction matrix of this kind of corporations are divided into small squares by multiple lines, and each small square forms its own grid area. This means that HHGNN learns the features of such corporations are formed by multiple groups. The feature similarity within each group is high and the feature network's interaction is dense as well. Therefore, HHGNN models such corporations in a grid-like way, simulating the interaction of features within the group and the interaction between different groups.

Hybrid type. In contrast to radial and grid corporations, hybrid type of corporations have no focused interaction features and need to be analyzed separately.

4.6 Gain brought by heterogeneous relations (RQ5)

We conduct experiments on the corporate heterogeneous graph module to analyze the interaction between corporations under the same relationship and how the different relationships are fused together. In addition, we also analyze the contribution of heterogeneous relationships to the final predicted rating results.

First, the interaction matrices of heterogeneous relationships in the corporate heterogeneous graph module in the test dataset are shown in Fig. 5.

The interaction characteristics of the main business interaction chart and the fund position interaction chart both present a grid shape. This is because many funds choose their positions according to the company′s main business. For example, liqueur funds hold shares of corporations that manufacture liqueur. However, the concept interaction chart and the industry interaction chart are both diagonal-centered and the interaction mostly exists between scatter points. From the perspective of attention value, the interaction of concept chart and industry chart is more intensive than the other two, so HHGNN pays more attention to them. The interaction chart of the adaptive graph is shown in Fig. 4. Its interaction is grid style similar to main business chart and fund holdings chart but with the lowest attention value.

The fusion mode of heterogeneous relations is shown in Fig. 6. Taking corporation SZ.002416 as an example, the contribution ratios of industry, main business, concept, fund position, and adaptive graph to the rating result BBB are 0.46, 0.05, 0.41, 0.04, and 0.04 respectively. This paper also lists the contribution of different heterogeneous relations to HHGNN, as shown in Table 3.

From Table 3, we can draw conclusions that the contribution of these five relationship graph networks ranked from high to low is: industry > concept > main business > fund position > adaptive relationship. The industry relationship is the most important between corporations, of which the contribution degree is more than 40%. The contribution of concept relationship is about 30% to 50%. However, the main business, fund position, and adaptive relationship are less than 10%.

5 Conclusions

In this work, we propose a novel framework named hierarchical heterogeneous graph neural network

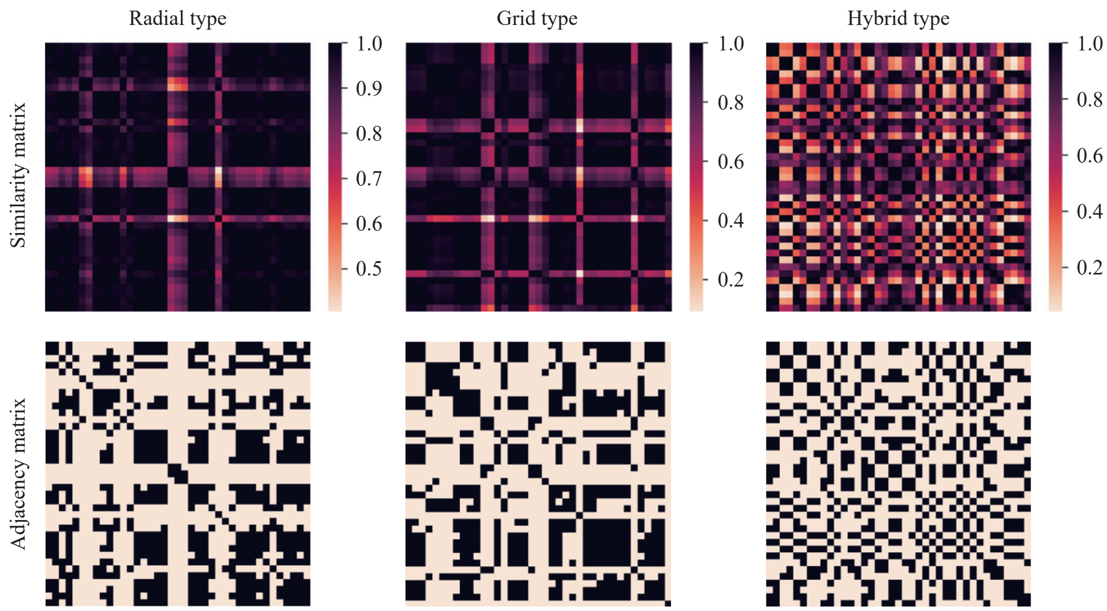

Fig. 4 Similarity and adjacency matrix of feature graph

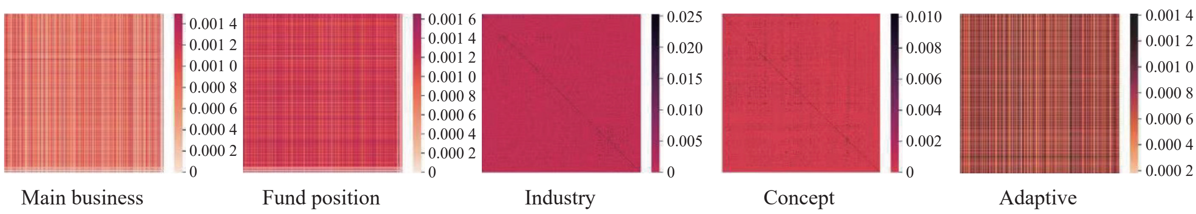

Fig. 5 Interaction matrix of heterogeneous and adaptive relationships

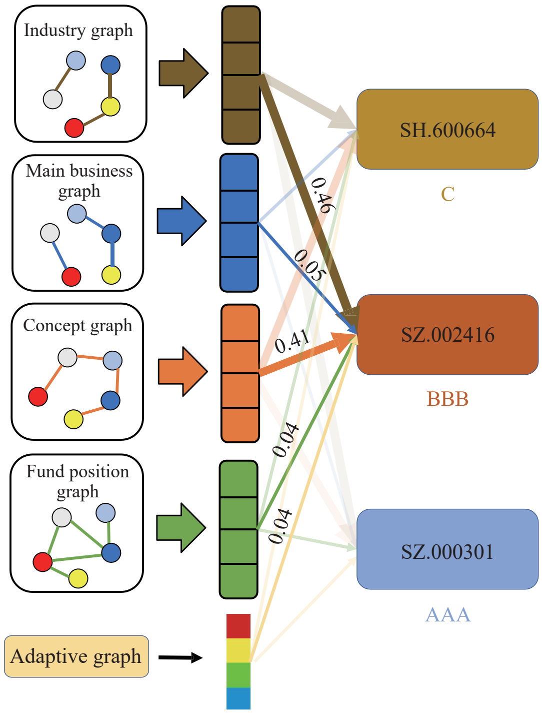

Fig. 6 Heterogeneous graph relational fusion

Table 3 Contribution of heterogeneous graphs to corporate credit ratings

Corporate code	Industry graph	Main business graph	Concept graph	Fund position graph	Adaptive graph
SZ.002416	0.46	0.05	0.41	0.04	0.04
SH.600741	0.45	0.04	0.44	0.03	0.04
SH.600160	0.62	0.06	0.23	0.04	0.05
SH.600380	0.53	0.05	0.34	0.04	0.05
SH.600664	0.53	0.05	0.33	0.04	0.05
SH.600050	0.62	0.06	0.23	0.04	0.05
SH.600500	0.46	0.04	0.44	0.03	0.03
SZ.002106	0.51	0.06	0.35	0.04	0.05
SH.600282	0.47	0.04	0.42	0.03	0.04
SZ.000301	0.45	0.04	0.44	0.03	0.04
SH.600066	0.51	0.04	0.38	0.03	0.04
SZ.002242	0.49	0.06	0.36	0.04	0.05
SH.600309	0.55	0.05	0.33	0.03	0.04
SH.601268	0.55	0.05	0.31	0.04	0.05
SZ.000528	0.53	0.05	0.34	0.04	0.04
SH.600079	0.55	0.05	0.32	0.04	0.05
SH.600839	0.52	0.06	0.32	0.04	0.05
SH.601003	0.48	0.04	0.41	0.03	0.04
SZ.000059	0.45	0.04	0.43	0.03	0.04
SH.600797	0.67	0.07	0.15	0.05	0.06
SZ.000968	0.45	0.04	0.43	0.03	0.04
SH.600188	0.56	0.05	0.30	0.04	0.05
SZ.002624	0.50	0.04	0.41	0.03	0.03
SH.600460	0.31	0.03	0.60	0.02	0.03

(HHGNN), which can fully model the hierarchy of corporate features and the heterogeneity of inter-corporate relations. In the downstream task, we use adversarial learning to make use of large amounts of unlabeled data, which further improves the accuracy and generalization ability of the model. Extensive experiments demonstrate that HHGNN achieves SOTA compared with the baseline methods. The analysis of ablation experiments shows HHGNN with attention mechanism has great interpretability. HHGNN is capable of intelligently interpreting financial data and has inestimable application value for reducing financial risk through corporate credit rating.

Declarations of conflict of interest

The authors declared that they have no conflicts of interest to this work.

References

[1] X. Dastile, T. Celik, M. Potsane. Statistical and machine learning models in credit scoring: A systematic literature survey. *Applied Soft Computing*, vol. 91, Article number 106263, 2020. DOI: 10.1016/j.asoc.2020.106263.

[2] P. Golbayani, D. Wang, I. Florescu. Application of deep neural networks to assess corporate credit rating, [Online], Available: https://arxiv.org/abs/2003.02334, 2020.

[3] B. F. Shi, G. T. Chi, W. P. Li. Exploring the mismatch between credit ratings and loss-given-default: A credit risk approach. *Economic Modelling*, vol. 85, pp. 420–428, 2020. DOI: 10.1016/j.econmod.2019.11.032.

[4] Y. D. Lu, M. Su. Asset allocation model across business cycle. In *Proceedings of International Conference on Business Management and Electronic Information*, Guangzhou, China, pp. 327–330, 2011. DOI: 10.1109/ICBMEI.2011.5917913.

[5] R. Vedala, B. R. Kumar. An application of naive bayes classification for credit scoring in e-lending platform. In *Proceedings of International Conference on Data Science & Engineering*, Cochin, India, pp. 81–84, 2012. DOI: 10.1109/ICDSE.2012.6282321.

[6] F. Shen, X. C. Zhao, G. Kou. Three-stage reject inference learning framework for credit scoring using unsupervised transfer learning and three-way decision theory. *Decision Support Systems*, vol. 137, Article number 113366, 2020. DOI: 10.1016/j.dss.2020.113366.

[7] P. Golbayani, I. Florescu, R. Chatterjee. A comparative study of forecasting corporate credit ratings using neural networks, support vector machines, and decision trees. *The North American Journal of Economics and Finance*, vol. 54, Article number 101251, 2020. DOI: 10.1016/j.najef.2020.101251.

[8] H. A. Abdou, J. Pointon. Credit scoring, statistical techniques and evaluation criteria: A review of the literature. *Intelligent Systems in Accounting, Finance and Management*, vol. 18, no. 2–3, pp. 59–88, 2011. DOI: 10.1002/isaf.325.

[9] R. Florez-Lopez, J. M. Ramon-Jeronimo. Enhancing accuracy and interpretability of ensemble strategies in credit risk assessment. A correlated-adjusted decision forest proposal. *Expert Systems with Applications*, vol. 42, no. 13, pp. 5737–5753, 2015. DOI: 10.1016/j.eswa.2015.02.042.

[10] J. Abellán, J. G. Castellano. A comparative study on base classifiers in ensemble methods for credit scoring. *Expert Systems with Applications*, vol. 73, pp. 1–10, 2017. DOI: 10.1016/j.eswa.2016.12.020.

[11] M. F. Wang, H. Ku. Utilizing historical data for corporate credit rating assessment. *Expert Systems with Applications*, vol. 165, Article number 113925, 2021. DOI: 10.1016/j.eswa.2020.113925.

[12] A. M. Ozbayoglu, M. U. Gudelek, O. B. Sezer. Deep learning for financial applications: A survey. *Applied Soft Computing*, vol. 93, Article number 106384, 2020. DOI: 10.1016/j.asoc.2020.106384.

[13] X. Wang, G. Chen, G. Qian, P. Gao, X. Y. Wei, Y. Wang, Y. Tian, W. Gao. Large-scale multi-modal pre-trained models: A comprehensive survey. *Machine Intelligence Research*, vol. 20, no. 4, pp. 447–482, 2023. DOI: 10.1007/s11633-022-1410-8.

[14] X. Y. Fu, J. N. Zhang, Z. Q. Meng, I. King. MAGNN: Metapath aggregated graph neural network for heterogeneous graph embedding. In *Proceedings of Web Conference 2020*, Taipei, China, pp. 2331–2341, 2020. DOI: 10.1145/3366423.3380297.

[15] Y. W. Fu, Y. Xiong, P. S. Yu, T. Y. Tao, Y. Y. Zhu. Metapath enhanced graph attention encoder for HINs representation learning. In *Proceedings of IEEE International Conference on Big Data*, Los Angeles, USA, pp. 1103–1110, 2019. DOI: 10.1109/BigData47090.2019.9006097.

[16] S. Yun, M. Jeong, R. Kim, J. Kang, H. J. Kim. Graph transformer networks. In *33rd Conference on Neural Information Processing Systems*, Vancouver, Canada, pp. 11960–11970, 2019.

[17] Z. N. Hu, Y. X. Dong, K. S. Wang, Y. Z. Sun. Heterogeneous graph transformer. In *Proceedings of Web Conference*, Taipei, China, pp. 2704–2710, 2020. DOI: 10.1145/3366423.3380027.

[18] H. T. Hong, H. T. Guo, Y. C. Lin, X. Q. Yang, Z. Li, J. P. Ye. An attention-based graph neural network for heterogeneous structural learning. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, New York, USA, pp. 4132–4139, 2020. DOI: 10.1609/aaai.v34i04.5833.

[19] C. Bravo, L. C. Thomas, R. Weber. Improving credit scoring by differentiating defaulter behaviour. *Journal of the Operational Research Society*, vol. 66, no. 5, pp. 771–781, 2015. DOI: 10.1057/jors.2014.50.

[20] F. Louzada, A. Ara, G. B. Fernandes. Classification methods applied to credit scoring: Systematic review and overall comparison. *Surveys in Operations Research and Management Science*, vol. 21, no. 2, pp. 117–134, 2016. DOI: 10.1016/j.sorms.2016.10.001.

[21] D. Wang, Z. Q. Zhang, R. Q. Bai, Y. N. Mao. A hybrid system with filter approach and multiple population genetic algorithm for feature selection in credit scoring. *Journal of Computational and Applied Mathematics*, vol. 329, pp. 307–321, 2018. DOI: 10.1016/j.cam.2017.04.036.

[22] A. C. Bahnsen, D. Aouada, B. Ottersten. Example-dependent cost-sensitive decision trees. *Expert Systems with Applications*, vol. 42, no. 19, pp. 6609–6619, 2015. DOI: 10.1016/j.eswa.2015.04.042.

[23] F. Shen, X. C. Zhao, G. Kou, F. E. Alsaadi. A new deep learning ensemble credit risk evaluation model with an improved synthetic minority oversampling technique. *Applied Soft Computing*, vol. 98, Article number 106852, 2021. DOI: 10.1016/j.asoc.2020.106852.

[24] S. Hamori, M. Kawai, T. Kume, Y. Murakami, C. Watanabe. Ensemble learning or deep learning? Application to default risk analysis. *Journal of Risk and Financial Management*, vol. 11, no. 1, Article number 12, 2018. DOI: 10.3390/jrfm11010012.

[25] H. T. Zhang, H. L. He, W. Y. Zhang. Classifier selection and clustering with fuzzy assignment in ensemble model for credit scoring. *Neurocomputing*, vol. 316, pp. 210–221, 2018. DOI: 10.1016/j.neucom.2018.07.070.

[26] B. J. Feng, W. F. Xue, B. D. Xue, Z. Y. Liu. Every corporation owns its image: Corporate credit ratings via convolutional neural networks. In *Proceedings of the 6th IEEE International Conference on Computer and Communications*, Chengdu, China, pp. 1578–1583, 2020. DOI: 10.1109/ICCC51575.2020.9344973.

[27] J. H. Dahooie, S. H. R. Hajiagha, S. Farazmehr, E. K. Zavadskas, J. Antucheviciene. A novel dynamic credit risk evaluation method using data envelopment analysis with common weights and combination of multi-attribute decision-making methods. *Computers & Operations Research*, vol. 129, Article number 105223, 2021. DOI: 10.1016/j.cor.2021.105223.

[28] K. Tran, T. Duong, Q. Ho. Credit scoring model: A combination of genetic programming and deep learning. In *Proceedings of Future Technologies Conference*, San Francisco, USA, pp. 145–149, 2016. DOI: 10.1109/FTC.2016.7821603.

[29] L. Wang, Y. G. Chen, H. Jiang, J. R. Yao. Imbalanced credit risk evaluation based on multiple sampling, multiple kernel fuzzy self-organizing map and local accuracy ensemble. *Applied Soft Computing*, vol. 91, Article number 106262, 2020. DOI: 10.1016/j.asoc.2020.106262.

[30] A. Namvar, M. Siami, F. Rabhi, M. Naderpour. Credit risk prediction in an imbalanced social lending environment. *International Journal of Computational Intelligence Systems*, vol. 11, no. 1, pp. 925–935, 2018. DOI: 10.2991/ijcis.11.1.70.

[31] H. L. He, W. Y. Zhang, S. Zhang. A novel ensemble method for credit scoring: Adaption of different imbalance ratios. *Expert Systems with Applications*, vol. 98, pp. 105–117, 2018. DOI: 10.1016/j.eswa.2018.01.012.

[32] C. C. Luo, D. S. Wu, D. X. Wu. A deep learning approach for credit scoring using credit default swaps. *Engineering*

Applications of Artificial Intelligence, vol. 65, pp. 465–470, 2017. DOI: 10.1016/j.engappai.2016.12.002.

[33] V. E. Neagoe, A. D. Ciotec, G. S Cucu. Deep convolutional neural networks versus multilayer perceptron for financial prediction. In *Proceedings of International Conference on Communications*, Bucharest, Romania, pp. 201–206, 2018. DOI: 10.1109/ICComm.2018.8484751.

[34] S. H. Yeh, C. J. Wang, M. F. Tsai. Deep belief networks for predicting corporate defaults. In *Proceedings of the 24th Wireless and Optical Communication Conference*, Taipei, China, pp. 159–163, 2015. DOI: 10.1109/WOCC.2015.7346197.

[35] A. Barja, A. Martínez, A. Arenas, P. Fleurquin, J. Nin, J. J. Ramasco, E. TomEás. Assessing the risk of default propagation in interconnected sectoral financial networks. *EPJ Data Science*, vol. 8, no. 1, Article number 32, 2019. DOI: 10.1140/epjds/s13688-019-0211-y.

[36] A. Khazane, J. Rider, M. Serpe, A. Gogoglou, K. Hines, C. B. Bruss, R. Serpe. DeepTrax: Embedding graphs of financial transactions. In *Proceedings of the 18th IEEE International Conference on Machine Learning and Applications*, Boca Raton, USA, pp, 126–133, 2019. DOI: 10.1109/ICMLA.2019.00028.

[37] D. W. Cheng, Z. B. Niu, Y. Y. Zhang. Contagious chain risk rating for networked-guarantee loans. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ACM, pp. 2715–2723, 2020. DOI: 10.1145/3394486.3403322.

[38] D. W. Cheng, Y. Tu, Z. W. Ma, Z. B. Niu, L. Q. Zhang. Risk assessment for networked-guarantee loans using high-order graph attention representation. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, Macao, China, pp. 5822–5828, 2019. DOI: 10.24963/ijcai.2019/807.

[39] D. W. Cheng, S. Xiang, C. C. Shang, Y. Y. Zhang, F. Z. Yang, L. Q. Zhang. Spatio-temporal attention-based neural network for credit card fraud detection. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, New York, USA, pp. 362–369, 2020. DOI: 10.1609/aaai.v34i01.5371.

[40] D. W. Cheng, Y. Y. Zhang, F. Z. Yang, Y. Tu, Z. B. Niu, L. Q. Zhang. A dynamic default prediction framework for networked-guarantee loans. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, Beijing, China, pp. 2547–2555, 2019. DOI: 10.1145/3357384.3357804.

[41] B. J. Feng, H. N. Xu, W. F. Xue, B. D. Xue. Every corporation owns its structure: Corporate credit ratings via graph neural networks, [Online], Available: https://arxiv.org/abs/2012.01933, 2020.

[42] A. Beutel, P. Covington, S. Jain, C. Xu, J. Li, V. Gatto, E. H. Chi. Latent cross: Making use of context in recurrent recommender systems. In *Proceedings of the 11th ACM International Conference on Web Search and Data Mining*, Marina Del Rey, USA, pp. 46–54, 2018. DOI: 10.1145/3159652.3159727.

[43] X. N. He, X. Y. Du, X. Wang, F. Tian, J. H. Tang, T. S. Chua. Outer product-based neural collaborative filtering. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, Stockholm, Sweden, pp. 2227–2233, 2018. DOI: 10.5555/3304889.3304969.

[44] X. N. He, L. Z. Liao, H. W. Zhang, L. Q. Nie, X. Hu, T. S. Chua. Neural collaborative filtering. In *Proceedings of the 26th International Conference on World Wide Web*, Perth, Australia, pp. 173–182, 2017. DOI: 10.1145/3038912.3052569.

[45] N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002. DOI: 10.1613/jair.953.

[46] B. J. Feng, W. F. Xue. Contrastive pre-training for imbalanced corporate credit ratings. In *Proceedings of the 14th International Conference on Machine Learning and Computing*, Guangzhou, China, pp, 293–297, 2021. DOI: 10.1145/3529836.3529911.

[47] X. Glorot, Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, Sardinia, Italy, pp. 249–256, 2010.

[48] T. Chen, C. Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, USA, pp. 785–794, 2016. DOI: 10.1145/2939672.2939785.

[49] B. J. Feng, W. F. Xue. Adversarial semi-supervised learning for corporate credit ratings. *Journal of Software*, vol. 16, no. 6, pp. 259–266, 2021. DOI: 10.17706/jsw.16.6.259-266.

Bo-Jing Feng received the M. Sc. degree in computer science in Center for Research on Intelligent Perception and Computing (CRIPAC), National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences (CASIA), China in 2022. Currently, he is an algorithm engineer at Meituan Inc., China.

His research interests include data mining, financial risk, machine learning and deep learning in finance, especially in corporate credit rating.

E-mail: bojing.feng@cripac.ia.ac.cn

ORCID iD: 0000-0002-8330-6333

Xi Cheng received the B. Sc. degree in information and computing science from University of Science and Technology Beijing, China in 2021. Currently, she is a master student in computer science in Center for Research on Intelligent Perception and Computing (CRIPAC), National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences (CASIA), China.

Her research interests include quantitative investment, financial risk, reinforcement learning and deep learning in stock trend forecasting.

E-mail: xi.cheng@cripac.ia.ac.cn
ORCID iD: 0000-0002-8289-0208

Hao-Nan Xu received the B. Sc. degree in information and computing science from Beijing Jiaotong University, China in 2020. Currently, he is a master student in computer science in Center for Research on Intelligent Perception and Computing (CRIPAC), National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences (CASIA), China.

His research interests include data mining, financial risk, machine learning and deep learning in finance, especially in stock trend forecasting.

E-mail: haonan.xu@cripac.ia.ac.cn

Wen-Fang Xue received the Ph. D. degree in Beihang University, China in 2003. Now he is an associate professor in Center for Research on Intelligent Perception and Computing (CRIPAC), Institute of Automation, Chinese Academy of Sciences (CASIA), China, and an executive director in Tianjin Academy for Intelligent Recognition Technologies, China.

His research interests include predictive stability, interpretability, and reasoning cognitive ability of intelligent risk control systems.

E-mail: wenfang.xue@ia.ac.cn (Corresponding author)
ORCID iD: 0000-0002-9646-914X