Rising like a Phoenix: Emerging from the Pandemic and Reshaping Human Endeavors with Digital Technologies ICIS 2023

AI in Business and Society

Dec 11th, 12:00 AM

Leveraging Interactions for Stationary and Dynamic Financial Distress Prediction: A Spatio-Temporal Financial Graph Attention Network

Qi Qi
National University of Singapore, qi.q@u.nus.edu

Frank Xing
National University of Singapore, xing@nus.edu.sg

Leveraging Interactions for Stationary and Dynamic Financial Distress Prediction: A Spatio-Temporal Financial Graph Attention Network

Completed Research Paper

Qi Qi
National University of Singapore
Singapore,117418
qi.q@u.nus.edu

Frank Xing
National University of Singapore
Singapore,117418
xing@nus.edu.sg

Abstract

Financial distress prediction is a prominent research topic in information systems, with two primary modelling categories: stationary and dynamic modelling. Recent stationary modelling works have leveraged company interactions to improve prediction performance, considering the heterogeneity of interactions while ignoring the dynamicity. However, few dynamic modelling works utilized interactions. To address the inconsistency and limitation of stationary and dynamic modelling works in leveraging interactions, we propose the Spatio-Temporal Financial Graph Attention Network with Meta-learning (STFGAN-Meta). STFGAN-Meta leverages interactions' spatial heterogeneity via the Spatial Aggregation module and temporal dynamicity via the Temporal Aggregation module. STFGAN-Meta introduces the Meta-learning Optimization module to unify stationary and dynamic modelling. Our experimental evaluation demonstrates that leveraging dynamicity and heterogeneity of interactions outperforms leveraging dynamicity or heterogeneity alone. Meta-learning succeeds in providing a generalized approach between stationary and dynamic modelling. STFGAN-Meta can be a promising risk assessment and decision-making tool in the financial industry.

Keywords: Financial distress prediction, interactions, dynamicity and heterogeneity, spatio-temporal aggregation, meta-learning

Introduction

Financial Distress Prediction (FDP) has been a prominent research topic in the field of information systems for several decades (Óskarsdóttir & Bravo, 2021; Wang et al., 2021) because of the adverse impact of financial distress risk (Chen et al., 2016; Karanikolos et al., 2013) and the significant benefits of company risk management (Wang et al., 2021). Financial distress risks of companies can lead to significant economic losses, reduced revenues, and increased costs (Karanikolos et al., 2013), posing threats to companies, investors, and industry regulators (Lin et al., 2011). As a useful tool, FDP models can provide early warning signals and support decision-making to mitigate potential losses. FDP research can be broadly categorized into two types based on whether the distress data changes dynamically: stationary FDP (SFDP) modelling and dynamic FDP (DFDP) modelling (Sun et al., 2020).

Traditional FDP research has focused on SFDP modelling, using the distress data in a certain period. These works have mainly focused on combining heterogeneous characteristics or constructing adaptive classifiers.

To obtain company financial characteristics, researchers have collected comprehensive public information about companies, such as their financial ratios (Du Jardin, 2015; Du Jardin, 2018; Geng et al., 2015; Sun & Li, 2012; Wang et al., 2019), lexical and sentimental characteristics (Li et al., 2016; Wang et al., 2021). Statistical and machine learning (ML) methods have been commonly used as classifiers for SFDP tasks (Chen et al., 2016), such as the support vector machine (SVM) (Sun & Li, 2012), artificial neural network (ANN) (Du Jardin, 2015; Geng et al., 2015), and ensemble learning methods (Du Jardin, 2015; Du Jardin, 2018; Wang et al., 2021). These statistical and ML methods have mainly focused on leveraging the endogenous characteristics of companies and ignored company interactions and the accompanying risk spillover effects among companies. With the rise of graph data, some works have emphasized the importance of leveraging interactions and graph-based models to supplement SFDP tasks (Jiang et al., 2022; Óskarsdóttir & Bravo, 2021). However, existing SFDP works only collect heterogeneous interactions in the stationary setting as a snapshot of company interactions.

The SFDP research relies on stationary prediction, which is ill-suited to predicting financial distress in a dynamic economic scenario with financial distress concept drift (Sun & Li, 2011). Motivated by this issue, the DFDP task has been proposed as an advanced FDP task for companies' dynamic operational environments. The DFDP task employs historical prediction experiences to predict the financial distress risk of companies in a specified period, leading to the financial distress concept drift between the distress data. Some work (Sun et al., 2017; Sun & Li, 2011; Sun et al., 2013; Sun et al., 2020) has indicated the superiority of DFDP models over SFDP models. These models tackle the issue of dynamic data distribution by incorporating dynamic classifiers, instance selection (Sun & Li, 2011), timing-weighting (Sun et al., 2017; Sun et al., 2020), and ensemble learning methods (Sun et al., 2017; Sun et al., 2013; Sun et al., 2020). However, existing DFDP research has only focused on limited characteristics, such as financial ratios, which deviates from the progress of mainstream research in SFDP tasks.

Both SFDP and DFDP tasks fail to fully acquire the company interactions due to their high costs and complexity. In reality, company interactions can be intricate, heterogeneous, and dynamic. The two primary challenges in modelling company interactions are dynamicity and heterogeneity. Additionally, there is a synchronicity between dynamicity and heterogeneity in company interactions, which further adds to the challenge of addressing interactions in FDP tasks. Consequently, existing DFDP and SFDP research neglect company interactions or only collect one snapshot of dynamic and heterogeneous interactions. Existing FDP research leads to the following problems: Firstly, one snapshot of interactions heavily relies on the snapshot timing and may introduce sampling bias, resulting in an incomplete usage or misuse of interaction information. Moreover, biased interactions cause limited generalizability and compromise the performance of the FDP tasks. Secondly, the research gap between DFDP and SFDP modelling forces users to trade off between leveraging interaction information and historical prediction experience. The inconsistency of the FDP model compels FDP researchers to adapt their FDP methodologies to distinct prediction scenarios repeatedly. This necessity amplifies the maintenance overhead and the application complexity of FDP models, severely limiting the popularity and generalization of FDP tasks.

To mitigate the modelling burdens on FDP researchers, alleviate the complexities in maintaining FDP services, and enhance the precision of FDP results for end-users, our research objective is to introduce a general FDP framework that can leverage interaction information to get better performance in both SFDP and DFDP tasks. This raises three research questions:

1. *Do dynamicity and heterogeneity of company interactions have an effect on the predictive performance in both SFDP and DFDP tasks?*
2. *How can we leverage company interactions better to improve performance in SFDP and DFDP tasks?*
3. *Can a general FDP framework alleviate concerns from financial distress concept drifts?*

To answer these questions, we propose a novel Spatio-Temporal Financial Graph Attention Network with Meta-leaning (*STFGAN-Meta*) method that can effectively leverage company interactions in stationary and dynamic scenarios. Our proposed *STFGAN-Meta* consists of three main components: the Spatial Aggregation module, which addresses heterogeneity in spatial interactions; the Temporal Aggregation module, which addresses dynamicity in temporal interactions; and the Meta-learning Optimization module, which tackles potential distress concept drifts and optimize prediction services for both SFDP and DFDP tasks. Our proposed method can effectively address the dynamicity and heterogeneity of interactions and potential distress concept drifts.

We empirically evaluate our proposed method using publicly accessible Chinese listed companies' data from January 1^{st} 2015 to December 31^{st} 2020. Our experimental results reveal that interaction information plays a critical role in both SFDP and DFDP tasks, and their dynamicity and heterogeneity significantly impact FDP tasks. Firstly, GNNs-based methods often outperform traditional ML methods, showing interactions' effectiveness in both FDP tasks. Secondly, leveraging heterogeneity of interactions can improve prediction performance in both FDP tasks. Thirdly, leveraging dynamic historical interactions helps exploit valuable financial distress risk information, while stacking dynamic historical interactions fails to exploit their distress information. Fourthly, interactions exhibit synchronous dynamic and heterogeneous properties, significantly impacting FDP tasks. Our proposed method can effectively address synchronicity between dynamicity and heterogeneity in SFDP and DFDP tasks. Finally, meta-learning can address the problem of distress concept drift in DFDP tasks and applies to SFDP tasks.

Our findings carry significant implications for companies, investors, and industry regulators and contribute to the field of FDP research in several ways. Firstly, we introduce the valuable characteristic of dynamicity in interactions and emphasize the advantage of leveraging the synchronicity of heterogeneity and dynamicity in company interactions, providing insights to address sampling bias and misuse of interactions in FDP tasks. Secondly, our study pioneers the integration of meta-learning into FDP research, demonstrating the potential of meta-learning to design generalized FDP models in ever-changing financial scenarios. Our study unifies the SFDP and DFDP models, mitigating the modelling burdens on FDP researchers and alleviating the complexities in maintaining FDP services. Finally, our proposed *STFGN-Meta* model outperforms several state-of-the-art (SOTA) FDP methods on SFDP and DFDP tasks, providing a powerful tool for risk assessment and decision-making in the financial industry.

Literature Review

FDP has garnered significant attention due to its vital implications and insights for key business decision-makers, such as shareholders, financial institutions, investors, and regulators (Alfaro et al., 2008). FDP research can be categorized into two types: SFDP modelling, which assumes that all distress data follows the independent and identical distribution (IID) (Hoadley, 1971); and DFDP modelling, which considers the financial distress concept drift in the distress data (Sun et al., 2020).

Stationary Financial Distress Prediction (SFDP)

Traditional FDP research has mainly concentrated on the SFDP task. Researchers have extensively utilized corporate disclosed data from public information platforms to extract financial features. Financial ratios have been consistently verified as the dominant features in SFDP research (Du Jardin, 2015; Du Jardin, 2018; Geng et al., 2015; Sun & Li, 2012; Wang et al., 2019). In addition to accounting data, previous SFDP research has sought to enhance performance by leveraging lexical or sentimental information from annual reports, financial news, and other online media (Li et al., 2016; Wang et al., 2021). However, these works have primarily focused on endogenous corporate characteristics, overlooking the potential risk spillover effects (Lang & Stulz, 1992). Recent works (Bi et al., 2022; Kou et al., 2021; Tobback et al., 2017; Yang et al., 2021) have addressed this issue by incorporating company interactions with heterogeneous financial graphs. These works have employed various methods, such as network centrality (Bi et al., 2022), relational statistics (Tobback et al., 2017), and heterogeneous interactions (Kou et al., 2021; Yang et al., 2021), to extract financial distress risk spillovers along these interactions. However, existing SFDP research only collects heterogeneous and stationary interactions by intercepting interactions as one snapshot.

In terms of methodology, traditional SFDP tasks have commonly used statistical and ML methods, such as logistic regression (LR) (Du Jardin, 2015; Du Jardin, 2018), SVM (Sun & Li, 2012), and ANN (Du Jardin, 2015; Geng et al., 2015). Recently, ML methods (Chen et al., 2016; Li et al., 2016) have shown overwhelming evidence of outperforming statistical methods. Ensemble learning methods have demonstrated superior performance over traditional ML methods, particularly for heterogeneous data (Du Jardin, 2015; Du Jardin, 2018; Wang et al., 2021). With the introduction of financial graphs, existing methods have utilized personalized PageRank (Óskarsdóttir & Bravo, 2021), the weighted-vote relational neighbour (wvRN) classifier (Tobback et al., 2017), and graph neural networks (GNNs) (Bi et al., 2022; Jiang et al., 2022) to mine distress information in interactions. These methods have shown promise in capturing the heterogeneity of financial graphs and facilitating more accurate SFDP.

However, a common drawback of SFDP research is that they all focus on stationary modelling for prediction with sample data in a certain period (Sun & Li, 2011). This SFDP research keeps the assumption that sample volume never changes in FDP and all distress data follow the IID. In the changing real world, new financially distressed companies gradually emerge to form sample data flow, changing the company operational environment and introducing the financial distress concept drift (Schlimmer & Granger, 1986; Sun et al., 2013; Sun et al., 2020). As time passes, stationary models can not effectively forecast financial distress in the changing economic environment. Consequently, research on DFDP modelling should be conducted regarding the financial distress concept drift to fit companies' dynamic operational environments over time.

Dynamic Financial Distress Prediction (DFDP)

Considering the financial distress concept drift, the DFDP task has been introduced to leverage dynamic financial distress data in a certain period. Existing DFDP works similarly focus on the adaptive data or classifiers to address this problem. To extract adaptive data, sample selection (Sun & Li, 2011) and timing-weighting (Sun et al., 2017; Sun et al., 2020) are proposed, such as the fixed-width time window method (Mitchell et al., 1994), and similar data selection methods (Sun & Li, 2011). To construct adaptive classifiers, ensemble learning (Sun et al., 2017; Sun et al., 2013; Sun et al., 2020) algorithms are also a critical approach to dealing with distress concept drift. Moreover, most classifier ensemble approach combines with the data ensemble approach, such as data selection (Almeida et al., 2018) and timing-weighting strategy (Sun et al., 2020).

Given the remarkable performance of interaction data in SFDP tasks, exploring its applicability in DFDP tasks is natural. However, in DFDP tasks, dynamic extraction of interactions faces significant challenges due to the complexities, variabilities, and high extraction costs of interactions. Consequently, most DFDP studies only use financial ratio data, which deviates from the prevailing research trend in SFDP tasks. Therefore, novel work is urgently needed to fill this research gap. With the development of dynamic prediction tasks (Lan et al., 2010; Zhang et al., 2020), meta-learning (Vilalta & Drissi, 2002) has shown promising results in addressing the challenge of dynamic drifts. Inspired by the success of meta-learning in addressing dynamic tasks, meta-learning may provide a solution to the challenges associated with DFDP tasks.

Graph Representation Learning

Graph representation learning is becoming increasingly popular in FDP research due to its ability to capture complex company interactions. There are two graph embedding techniques based on whether the graph data is dynamically generated: stationary graph embedding, which considers the graph data as a single snapshot, and dynamic graph embedding, which considers the graph data as a set of multiple snapshots.

For stationary graph embedding, GNNs and advanced GNNs-based methods(Velickovic et al., 2017) have shown promising results for modelling interaction data as a feature aggregation framework. However, traditional GNN-based methods need to be better equipped for heterogeneous graphs. To tackle this limitation, some research (Dong et al., 2017; Perozzi et al., 2014; Wang et al., 2019) models homogeneous representations and preserves the heterogeneous structures with different semantics. These works can be categorized into two categories: meta-path-based methods (Dong et al., 2017) and deep neural network methods. Meta-path-based methods utilize the random walk strategy (Perozzi et al., 2014) along the meta-path—— the compound interactions between nodes to explore the semantics of heterogeneous graphs. Deep neural network-based models aggregate meta-path information by neighbours' information (Zhang et al., 2019) or bi-molecular subgraph's information (Wang et al., 2019) to explore the semantics of heterogeneous graphs. Nevertheless, stationary graph models overlook temporal information and fail to capture the dynamicity of graphs.

Dynamic graph embedding techniques have been proposed to capture graphs' semantic features and dynamicity. The mainstream of existing dynamic graph embedding methods is snapshot-based methods (Du et al., 2018; Zhang et al., 2020), which treat the dynamic graph as a sequence of stationary graphs. These methods provide a coarse-grained view of the graph dynamicity, emphasizing global changes and effectively capturing long-term dynamic features. However, due to the synchronous dynamicity and heterogeneity, the snapshot-based heterogeneous graph method is usually more complex than the heterogeneous graph method. Some dynamic heterogeneous graph methods (Fan et al., 2022; Zhang et al., 2020) adopt spatio-temporal aggregation mechanisms to address this synchronicity between compound spatial and temporal interactions.

Financial systems in the real world are often characterized by a wide range of entities, interactions, and temporal information, representing dynamic heterogeneous graphs. Mathematically, company interactions can be represented as dynamic heterogeneous graphs with a set of nodes (companies and related entities, such as executives and shareholders) and edges (various relationships between nodes, such as supply chains, investment, and employment relationships). GNNs have shown great promise in financial applications, given their ability to capture heterogeneous interactions (Bi et al., 2022) and dynamic evolution (Yang et al., 2021). However, current GNNs-based models for FDP tasks ignore the temporal aspect of financial graphs. Therefore, there is a pressing need to explore the role of interaction dynamicity through dynamic graph embedding methods, which may help comprehend how financial risks propagate and lead to more accurate and effective FDP models.

Meta-learning

Meta-learning (Vilalta & Drissi, 2002), known as "learning to learn", aims to enable ML algorithms to learn from historical experience and generalize to new and unseen tasks. Meta-learning imitates the data generation method of the original task to generate multiple meta-tasks from the training set. It improves the models' generalization ability by learning from these meta-tasks. Consequently, meta-learning has a wide range of generalized implications (Lan et al., 2010; Zhang et al., 2020). Some works (Lan et al., 2010; Zhang et al., 2020) have explored the potential of meta-learning in handling dynamic concept drift in dynamic prediction tasks. For example, Lan et al. (2010) proposed a dynamic meta-learning framework for dynamic failure prediction, obtaining failure patterns from a changing training set. Despite these advancements, a research gap exists in the literature on combining meta-learning and DFDP tasks, leaving ample space for exploration.

Methodology

This section presents a comprehensive overview of our proposed method: the Spatio-Temporal Financial Graph Attention Network with Meta-leaning (*STFGAN-Meta*). Our model consists of three main components: the Spatial Aggregation module, which addresses heterogeneity in spatial interactions; the Temporal Aggregation module, which addresses dynamicity in temporal interactions; and the Meta-learning Optimization module, which tackles potential distress concept drifts and optimizes prediction services for both SFDP and DFDP tasks. Figure 1 provides an overview of the proposed method's architecture.

Problem Formulation

The SFDP task contains an annual FDP task for one benchmark year. In contrast, the DFDP task contains a set of annual FDP tasks with varying benchmark years and uses previous annual FDP tasks to predict the latest annual FDP task. Thus, SFDP and DFDP tasks are collections of annual FDP tasks. We define the uniform FDP task that applies to dynamic and stationary scenarios, with the benchmark year span as $\mathcal{T}$ years, $\tau \in \mathcal{T}$. For the SFDP task, $\mathcal{T} = 1$. Conversely, $\mathcal{T}$ is greater than 1 for the DFDP task. The goal of uniform FDP task is to complete $\mathcal{T}$ annual FDP tasks $\{FDP_\tau\}_{\tau=1}^{\mathcal{T}}$, specifying some annual FDP tasks from $\{FDP_\tau\}_{\tau=1}^{\mathcal{T}}$ as the training data. The summary of defined parameters is shown in Table 1.

For the annual FDP task FDP_τ, we define the time window of feature collection as T years. The FDP_τ uses the historical data in T years since the benchmark year τ, the period $[\tau + 1 - T, \tau]$, to predict the future

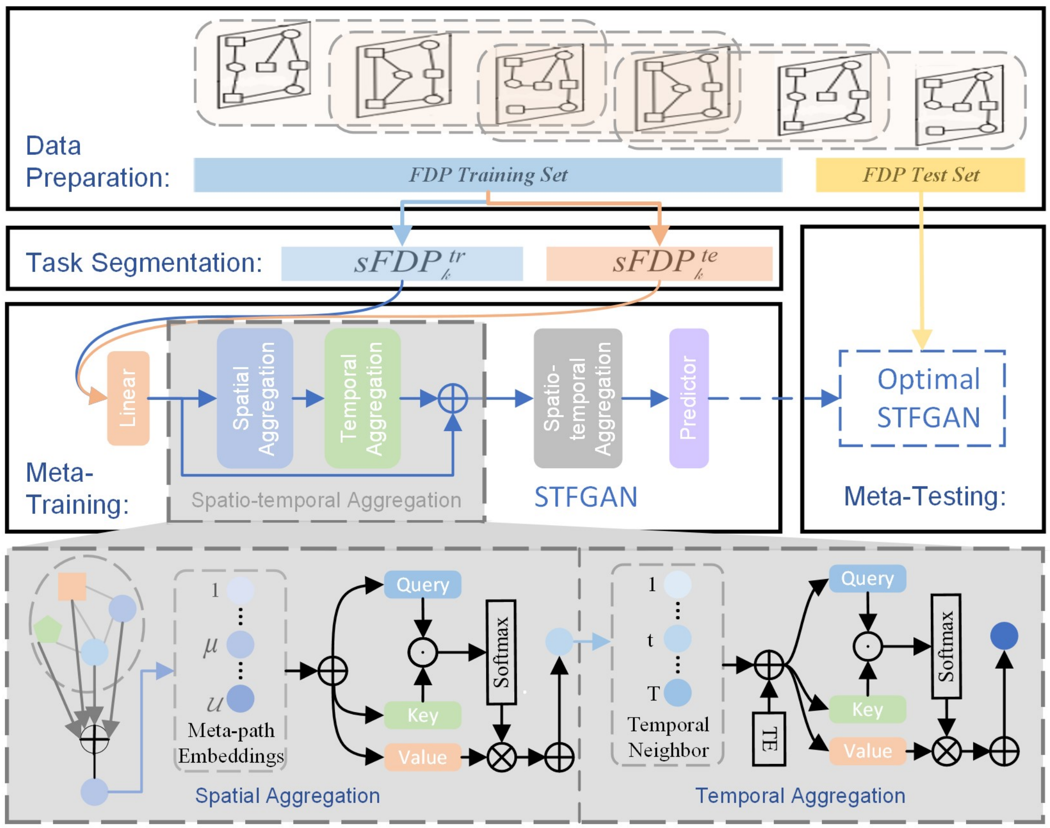

Figure 1. The overall framework of the proposed *STFGAN-Meta*.

Parameter	Description	Parameter	Description
$\mathcal{T}$	Number of years in the benchmark year span, $\tau \in \mathcal{T}$	$\mathcal{M}_{\mathcal{V}}^{\tau}$	Node-specific projection matrix for node set $\mathcal{V}$
FDP_τ	Annual FDP task for the benchmark year τ	$\mathcal{P}_\theta$	Predictor with parameter θ
T	Time windows of distress feature	$\mathcal{F}_\psi$	Feature extractor with parameter ψ
$\mathcal{G}_\tau$	ST-Graph for the FDP_τ	$\mathcal{K}$	Number of sub-FDP tasks
ϱ_τ^t	Snapshot of the ST-Graph at timestamp t in FDP_τ	α	Learning-rate hyperparameter for internal optimization
$\mathcal{N}_{v_t}^S$	Spatial neighbour set of node v_t in snapshot ϱ_τ^t	β	Learning-rate hyperparameter for cross-task meta-optimization
Φ_{v_t}	Meta-paths connecting node v_t to its spatial neighbours	L	Number of spatio-temporal aggregation layers in the framework
$\mathcal{N}_v$	Temporal neighbour set of node v in FDP_τ	$\eta_{n_{v_t}}^{\phi_\mu}$	Weight coefficient for spatial neighbour n_{v_t} in meta-path ϕ_μ
$\mathcal{V}$	Node set, including company and individual nodes	w_{ϕ_μ}	Weight coefficient of meta-path ϕ_μ
$\mathcal{E}_\tau$	Edge set, which contains all interaction snapshots ϵ_t at timestamp t	r_{v_t}	Weight coefficient of temporal neighbour v_t

Table 1. Summary of defined parameters.

financial distress situation for target companies. The T years historical data can be constructed as a spatio-temporal graph (ST-Graph) $\mathcal{G}_\tau = \{\varrho_\tau^t\}_{t=1}^T = \{(\mathcal{V}, \mathcal{E}_\tau, \mathcal{X}_\tau, \mathcal{Y}_\tau)\}$, where $t \in [1, T]$. Here, $\mathcal{V}$ represents the node set, including both target nodes (the listed company set $C = \{c_n\}_{n=1}^N$) and non-target nodes (the related company set $R = \{r_m\}_{m=1}^M$ and individual set $I = \{i_j\}_{j=1}^J$). The edge set $\mathcal{E}_\tau = \{\epsilon_t\}_{t=1}^T$, where ϵ_t represents the interaction snapshot at timestamp t. The feature set $\mathcal{X}_\tau = \{X_{v,t}\}_{t=1}^T$, where $v \in \mathcal{V}$ and $X_{v,t}$ represents the input feature of the node v at timestamp t. The label set $\mathcal{Y}_\tau$ represents listed companies' financial distress risk label in the FDP_τ task, $\mathcal{Y}_\tau = \{y_{c_n,\tau}\}_{c_n \in C}$. Here, $y_{c_n,\tau} = 1$ denotes financial distress for the listed company c_n in the FDP_τ task. In sum, each ST-graph consists of T snapshots, and each snapshot contains the interaction set and node feature set for the corresponding observation timestamp t, $\varrho_\tau^t = \{(\mathcal{V}, \epsilon_t, X_t)\}$.

To describe the spatio-temporal financial graphs, we illustrate the definition from ST-Graph $\mathcal{G}_\tau = \{\varrho_\tau^t\}_{t=1}^T$. For the target node v, we define the temporal neighbour set of node v as $\mathcal{N}_v = \{v_t\}_{t=1}^T$ in T snapshots $\{\varrho_\tau^t\}_{t=1}^T$. For each ϱ_τ^t, node v_t have interactions with its spatial neighbour $n_{v_t} \in \mathcal{N}_{v_t}^S$, where $\mathcal{N}_{v_t}^S$ is node v_t 's spatial neighbour set in ϱ_τ^t. The node v_t can contact its spatial neighbour n_{v_t} via multiple meta-paths $\Phi_{v_t} = \{\phi_1, ..., \phi_\mu, ..., \phi_\mathcal{U}\}$. For instance, take node c_n as the target node, c_n is the client of c_m, meanwhile, c_n and c_m have shared shareholder i_j in ϱ_τ^t. Thus, $\Phi_{c_n}^t = \{\phi_1 = (c_n, \text{'}client\text{'}, c_m), \phi_2 = (c_n, \text{'}shareholder\text{'}, i_j, \text{'}shareho-lder\text{'}, c_m)\}$, and $\mathcal{N}_{c_n,t}^S = \{c_n, c_m, i_j\}$.

In the general FDP task, we aim to learn a time-independent financial distress probability function $\hat{\mathcal{Y}} = \mathcal{P}_\theta(\mathcal{F}_\psi(\mathcal{X}))$ for listed companies. Given the label space $\mathcal{Y}_{c_n \in C} = [\mathcal{Y}_{c_n,1}, ..., \mathcal{Y}_{c_n,\tau}, ..., \mathcal{Y}_{c_n,\mathcal{T}}]$ and its feature space $\mathcal{X}_{c_n \in C} = [\mathcal{X}_{c_n,1}, ..., \mathcal{X}_{c_n,\tau}, ..., \mathcal{X}_{c_n,\mathcal{T}}]$, the objective is to learn the feature extractor $\mathcal{F}_\psi$ with the parameter ψ that maps the input feature vectors to their corresponding embeddings, and the predictor $\mathcal{P}_\theta$ with the parameter θ that maps node embeddings to the probability of financial distress risk. The general FDP task can be divided into several annual FDP tasks, where each annual task has a shared framework to learn a time-independent financial distress probability function.

Framework

For each annual FDP task FDP_τ, we firstly uniform the feature spaces for heterogeneous nodes. Since different types of nodes have various features, dealing with nodes with distinct feature spaces becomes challenging. Node-specific linear transformation matrixes can project different types of nodes to a uniform feature space. The projection process can be formulated as follows:

$$h_{v_t}^0 = \mathcal{M}_\mathcal{V}^\tau \cdot X_{v_t}, \tag{1}$$

where X_{v_t} is the feature vector of node v_t at the snapshot ϱ_τ^t, $\mathcal{M}_\mathcal{V}^\tau$ is the node-specific projection matrix for the FDP_τ task, and $h_{v_t}^0$ is the transformed feature vector for node v_t. After this transformation, a spatio-temporal aggregation block is applied to capture spatio-temporal dependencies among each snapshot. This block comprises L spatio-temporal aggregation layers, each including a spatial aggregation module to address spatial heterogeneity and a temporal aggregation module to address temporal dependence. Once the final spatio-temporal embeddings of the listed company set are obtained, they are fed into a predictor to perform the final binary classification task. In this work, a fully connected network (FCN) (LeCun et al., 1998) module is used as the predictor $\mathcal{P}_\theta$, and the prediction result for each listed company c_n is given by:

$$\hat{y}_{c_n} = \mathcal{P}_\theta(\sum_{t=1}^T h_{c_n,t}^L), c_n \in C, \tag{2}$$

where $h_{c_n,t}^L$ is the final spatio-temporal embedding of listed company c_n at the timestamp t, and $\hat{y}_{c_n}$ is the corresponding prediction result in the FDP_τ task.

Spatial Aggregation

Spatial aggregation modules aim to depict spatial heterogeneity by aggregating node embeddings of the target node v_t's spatial neighbours in the corresponding snapshot ϱ_τ^t. The formalization of this process is as follows:

$$h_{v_t}^{l,S} = Aggregation_{spa}(h_{v_t}^{l-1}, \varrho_\tau^t | \psi_{spa}^l), l \in [1, L], \tag{3}$$

where $h_{v_t}^{l-1}$ is the output embedding of node v_t from the $l-1$ spatio-temporal aggregation layer at snapshot ϱ_τ^t, and ψ_{spa}^l represents the trainable parameters of the spatial aggregation in the l spatio-temporal aggregation layer. The output $h_{v_t}^{l,S}$ represents the spatial embedding of node v_t at snapshot ϱ_τ^t. Considering the heterogeneity in financial graphs, the proposed approach employs a heterogeneous graph neural network that utilizes both a neighbour-level attention mechanism and a meta-path level attention mechanism. These attention mechanisms account for spatial neighbour nodes' varying roles and importance in the prediction task.

The neighbour-level attention mechanism aims to learn the importance of neighbours from the same meta-path in ϱ_τ^t. We employ self-attention (Vaswani et al., 2017) to learn the importance of various neighbours n_v from the subset of neighbours $\mathcal{N}_{v_t}^{S,\phi_\mu}$ with the same meta-path ϕ_μ to the target node v_t. Specifically, the neighbour-level attention score $e_{n_{v_t}}^{\phi_\mu}$ for neighbour n_{v_t} to the target node v_t is obtained by calculating following:

$$e_{n_{v_t}}^{\phi_\mu} = attention_N(h_{n_{v_t}}^{l-1}, h_{v_t}^{l-1}|\phi_\mu), n_{v_t} \in \mathcal{N}_{v_t}^{S,\phi_\mu}, \tag{4}$$

where the function $attention_N$ is a shared deep neural network responsible for performing the neighbour-level attention for the meta-path ϕ_μ. Then, we apply the softmax function, also known as normalized exponential function (Bishop & Nasrabadi, 2006), to normalize the attention score $e_{n_{v_t}}^{\phi_\mu}$ of neighbour n_{v_t} and obtain the corresponding weight coefficient $\eta_{n_{v_t}}^{\phi_\mu}$:

$$\eta_{n_{v_t}}^{\phi_\mu} = softmax(e_{n_{v_t}}^{\phi_\mu}) = \frac{e_{n_{v_t}}^{\phi_\mu}}{\sum_{n'_{v_t} \in \mathcal{N}_{v_t}^{S,\phi_\mu}} e_{n'_{v_t}}^{\phi_\mu}}. \tag{5}$$

The meta-path-specific embedding of node v_t for the meta-path ϕ_μ can be aggregated by the neighbour's projected features with the corresponding coefficients $\eta_{n_{v_t}}^{\phi_\mu}$ as follows:

$$h_{v_t,\phi_\mu}^l = \sigma(\sum_{n_{v_t} \in \mathcal{N}_{v_t}^{S,\phi_\mu}} \eta_{n_{v_t}}^{\phi_\mu} \cdot h_{n_{v_t}}^{l-1}). \tag{6}$$

Applying a meta-path attention mechanism enables learning the importance of each meta-path in the FDP task, accounting for their varying influence on the target node. To achieve this, all meta-path-specific embeddings for the target node v_t, denoted by $\{h_{v_t,\phi_\mu}^l, \phi_\mu \in \Phi_{v_t}\}$, are utilized by the meta-path attention layer to learn the importance of each meta-path w_{ϕ_μ}, as expressed below:

$$w_{\phi_\mu} = attention_M(h_{v_t,\phi_1}^l, ..., h_{v_t,\phi_\mu}^l, ..., h_{v_t,\phi_\mathcal{U}}^l|\Phi_{v_t}), \tag{7}$$

where $attention_M$ performs the meta-path attention mechanism and captures various meta-path information. We use the transformer attention mechanism (Vaswani et al., 2017) to aggregate the meta-path-specific embeddings of target nodes. Firstly, we transform the target node's meta-path-specific embeddings into spatial Query vector q_{v_t,ϕ_μ}^S, Key vector k_{v_t,ϕ_μ}^S, and Value vector v_{v_t,ϕ_μ}^S:

$$q_{v_t,\phi_\mu}^S = W_q^S \cdot h_{v_t,\phi_\mu}^l, \quad k_{v_t,\phi_\mu}^S = W_k^S \cdot h_{v_t,\phi_\mu}^l, \quad v_{v_t,\phi_\mu}^S = W_v^S \cdot h_{v_t,\phi_\mu}^l; \tag{8}$$

where $W_q^S, W_k^S, W_v^S \in R^{d \times d}$ are trainable transformation matrices for Query, Key, and Value, respectively.

Then, we calculate the dot product of the Query vector and Key vector, resulting in the attention coefficient for each meta-path-specific embedding. The calculation can be expressed as follows:

$$w_{v_t}^{\phi_\mu} = \frac{\exp([q_{v_t,\phi_\mu}^S] \cdot [k_{v_t,\phi_\mu}^S])}{\sum_{\mu'=1}^{\mathcal{U}} \exp([q_{v_t,\phi_{\mu'}}^S] \cdot [k_{v_t,\phi_{\mu'}}^S])} \tag{9}$$

The final spatial embedding of node v_t in the snapshot ϱ_τ^t is a linear combination of its Value vector and the calculated attention values:

$$h_{v_t}^{S,l} = \sigma(\sum_{\mu=1}^{\mathcal{U}} [w_{v_t}^{\phi_\mu}] \cdot [v_{v_t,\phi_\mu}^S]). \tag{10}$$

Temporal Aggregation

The Temporal aggregation module aims to capture the temporal dependencies among the target node v 's temporal neighbours v_t from T snapshots $\mathcal{G}_\tau = \{\varrho_\tau^t\}_{t=1}^T$. Mathematically, the Temporal aggregation module in the l spatio-temporal aggregation layer can be formalized as follows:

$$h_{v_t}^{ST,l} = Aggregation_{tem}(h_{v_1}^{S,l}, ..., h_{v_t}^{S,l}, ..., h_{v_T}^{S,l} | \psi_{tem}^l, \mathcal{G}_\tau). \tag{11}$$

For timestamp $t \in [1, T]$, $h_{v_t}^{S,l}$ represents the spatial embedding of the node v's temporal neighbour v^t in set $\mathcal{N}_v = \{v_t\}_{t=1}^T$. $h_{v_t}^{ST,l}$ is the spatial-temporal embedding of node v_t, and ψ_{tem}^l represents the trainable parameters of the temporal aggregation module.

Similarly, we follow the transformer attention mechanism (Vaswani et al., 2017) to aggregate the spatial embeddings of temporal neighbours. The difference is that we additionally introduce a temporal embedding function (Fan et al., 2022), denoted by $TE(\cdot)$, for $h_{v_t}^{S,l}$ that incorporates timestamp-related factors:

$$TE(h_{v_t}^{S,l}) = \|_{i=1}^d \ (h_{v_t}^{S,l} + p(t,i)), \tag{12}$$

where i denotes the index of each element in the embedding of node v, which has a feature-length of d. The function $p(\cdot)$ is a timestamp-dependent sinusoid:

$$p(t, 2i) = sin(t/10000^{2i/d}), \quad p(t, 2i+1) = sin(t/10000^{(2i+1)/d}). \tag{13}$$

We introduce the $TE(\cdot)$ function to enable spatial embeddings to become discriminative about timestamps t. We then transform the target node's spatial embedding into Query vector q_{v_t}, its temporal neighbour's spatial embedding into Key vector k_{v_t}, and Value vector v_{v_t}:

$$q_{v_t} = W_q^{ST} \cdot TE(h_{v_t}^{S,l}), \quad k_{v_t} = W_k^{ST} \cdot TE(h_{v_t}^{S,l}), \quad v_{v_t} = W_v^{ST} \cdot TE(h_{v_t}^{S,l}); \tag{14}$$

where $W_q^{ST}, W_k^{ST}, W_v^{ST} \in R^{d \times d}$ are trainable temporal transformation matrices for Query, Key, and Value, respectively.

Then, we follow the meta-path attention and measure the importance of each temporal neighbour as the dot product of the Query vector and Key vector and combine its Value vector and the calculated attention values:

$$r_{v_t} = \frac{\exp([q_{v_t}] \cdot [k_{v_t}])}{\sum_{t'=1}^T \exp([q_{v_{t'}}] \cdot [k_{v_{t'}}])}, \tag{15}$$

$$h_{v_t}^{ST,l} = [r_{v_t}] \cdot [v_{v_t}]. \tag{16}$$

With the spatial-temporal embedding for each node, we design a gate mechanism for aggregating the features of the node itself and its neighbours:

$$h_{v_t}^l = \delta_t \cdot [h_{v_t}^{ST,l}] + (1 - \delta_t) \cdot [W \cdot h_{v_t}^{l-1}], \tag{17}$$

where $\delta_t \in R^1$ and $W_l \in R^{d \times d}$ are the trainable weight and transformation matrix.

Meta-learning Optimization

The learning algorithm of *STFGAN-Meta* is presented in Algorithm 1.

In the following, we detail the optimization process of the feature extractor $\mathcal{F}_\psi$ and prediction classifier $\mathcal{P}_\theta$. Given a general FDP task $\{FDP_\tau\}_{\tau=1}^{\mathcal{T}}$, we manually divide it into $\mathcal{K}$ sub-FDP tasks $\{sFDP_\kappa\}_{\kappa=1}^{\mathcal{K}}$, each subtask $sFDP_\kappa$ contains two splits, the training set $sFDP_\kappa^{tr}$ and test set $sFDP_\kappa^{te}$, respectively. For internal optimization on $sFDP_\kappa$, parameters ψ and θ are firstly updated from the $sFDP_\kappa$-specific supervised loss $\mathcal{L}$ (e.g. cross-entropy for classification):

$$(\psi'_\kappa, \theta'_\kappa) \leftarrow (\psi, \theta) - \alpha \nabla_{(\psi,\theta)} \mathcal{L}(sFDP_\kappa^{tr}; \psi, \theta), \tag{18}$$

Input: FDP set $\{FDP_\tau\}_{\tau=1}^{\mathcal{T}}$;
Hyperparameters $\alpha, \beta, \mathcal{K}$;
Maximum number of iterations I.
Output: Feature extractor $\mathcal{F}_\psi$;
Classifier network $\mathcal{C}_\theta$.
Randomly initialize θ, ψ;
Generate sub-FDP task lists: $\{sFDP_\kappa = (sFDP_\kappa^{tr}, sFDP_\kappa^{te})\}_{\kappa=1}^{\mathcal{K}}$;
for $i = 1 : I$ **do**
  for $\kappa = 1 : \mathcal{K}$ **do**
    $(\psi'_\kappa, \theta'_\kappa) \leftarrow (\psi, \theta) - \alpha \nabla_{(\psi,\theta)} \mathcal{L}(sFDP_\kappa^{tr}; \psi, \theta)$;
    Compute $\mathcal{L}(sFDP_\kappa^{te}; \psi'_\kappa, \theta'_\kappa)$;
  end
  $(\psi, \theta) \leftarrow (\psi, \theta) - \beta \nabla_{(\psi,\theta)} \sum_{\kappa=1}^{\mathcal{K}} \mathcal{L}(sFDP_\kappa^{te}; \psi'_\kappa, \theta'_\kappa)$;
end

Algorithm 1. The Spatio-Temporal Financial Graph Attention Network with Meta-leaning (*STFGAN-Meta*).

where α is a learning-rate hyperparameter.

After obtaining the internally optimized parameter set, we can apply external optimization to enable the model to obtain the attributes displayed on $sFDP_\kappa^{te}$. Specifically, based on the updated parameter $(\psi'_\kappa, \theta'_\kappa)$, we records the validation loss $\mathcal{L}(sFDP_\kappa^{te}; \psi'_\kappa, \theta'_\kappa)$ of the test set $sFDP_\kappa^{te}$. With all the loss values for each $sFDP_\kappa$ task, the cross-task meta-optimization is updated as follows:

$$(\psi, \theta) \leftarrow (\psi, \theta) - \beta \nabla_{(\psi,\theta)} \sum_{\kappa=1}^{\mathcal{K}} \mathcal{L}(sFDP_\kappa^{te}; \psi'_\kappa, \theta'_\kappa), \tag{19}$$

where β is a learning-rate hyperparameter.

Empirical Evaluation

We collect a real-world financial graph dataset to empirically evaluate the proposed *STFGAN-Meta* model for FDP tasks under both stationary and dynamic scenarios. This dataset comprises 3,469 Chinese listed companies and their interactions from January 1st 2015 to December 31st 2020.

Data

To fair comparison, this work focuses on listed companies from China's Shenzhen and Shanghai Stock Exchanges, following Wang et al. (2021) and Geng et al. (2015). The financial distress of Chinese companies has attracted increasing attention (Geng et al., 2015; Jiang et al., 2022; Sun & Li, 2011; Wang et al., 2021) due to China's significance as a key market for global investors. Therefore, developing an appropriate FDP model for Chinese companies is significant to global investors. Our dataset encompasses all listed companies from January 1st 2015 to December 31st 2020, ensuring the consistency and availability of their financial distress data. Specifically, we collected 3,469 Chinese companies listed throughout this six-year timeframe and their financial distress information. For financial distress labels, we adopted the special treatment (ST) warning mechanism (Geng et al., 2015; Wang et al., 2021) and used the ST as the annual financial distress label. For financial characters, we utilized the China Security Market Accounting Research (CSMAR, n.d.) database and selected 36 financial ratios following Alfaro et al. (2008) and Wang et al. (2021). For company interactions, we got the data from a leading Chinese data services company, ChinaScope (n.d.), and collected extensive, dynamic, and heterogeneous interactions. In total, we collected 332,611 related companies, 255,654 individuals, and 2,106,707 core interactions. Further information on the interaction data can be found in Table 2.

Year	Number of Entities			Number of Interactions				
	Listed Companies (C)	Related Companies (R)	Individuals (I)	C-C	C-R	C-I	R-I	I-I
2015	3,469	332,611	255,654	3,062	134,154	83,784	8,814	13,950
2016				4,614	183,429	107,549	11,777	12,632
2017				4,960	211,462	119,461	14,813	11,633
2018				4,361	227,003	119,770	18,222	12,261
2019				4,022	229,114	121,054	23,500	12,551
2020				4,048	236,969	123,070	32,209	12,459

Table 2. Companies' and Individual' Interaction Information.

To generate ST-graphs for annual FDP tasks, we use financial ratio data from three years ($T = 3$) prior to the benchmark years (Geng et al., 2015; Wang et al., 2021). For instance, financial ratio data from 2018 to 2020 can be used as the hard feature of FDP_{2021}. For interactions, we collect historical data from $\Gamma = [0, 1, 2, 3]$ years prior to the benchmark years, where $\Gamma = 0$ indicates the use of only financial ratios without considering interactions and $\Gamma = 3$ indicates the use of ST-graphs. However, due to the limited availability of interaction information in our dataset, only benchmark year data from 2018 to 2021 is used.

Baselines

We compare our proposed model with four traditional methods to evaluate the effectiveness of our proposed method and investigate the impact of interaction dynamicity and heterogeneity on the SFDP and DFDP tasks. These included:

(1) ML-based methods that only utilize financial ratios:
- Random Forest (RF) (Breiman, 2001): a traditional ML technique that utilizes ensemble learning with decision trees to provide solutions to complex problems.
- Multi-Layer Perceptron (MLP) (Marini et al., 2007): a fully connected feedforward ANN with two fully connected layers and one classifier.

(2) The homogeneous GNNs-based methods that utilize financial ratios and homogeneous interactions:
- Financial Graph Attention Networks (FGAN): an advanced extension of Graph Attention Networks (Velickovic et al., 2017) that utilizes masked self-attention layers to aggregate neighbours' embeddings with different weights.

(3) The heterogeneous GNNs-based methods that utilize financial ratios and heterogeneous interactions:
- STFGAN without Temporal Aggregation (SFGAN): a simplified extension of our proposed model without the Temporal Aggregation module, designed for heterogeneous graph data.

(4) The temporal homogeneous GNNs-based methods utilize financial ratios and homogeneous interactions with temporal information:
- STFGAN without Spatial Aggregation (TFGAN): an advanced extension of FGAN with the Temporal Aggregation module, designed for dynamic and homogeneous graph data.

Experimental Procedure

The proposed model and all baselines are implemented using official codes in PyTorch and Deep Graph Library (DGL) in Python 3.7. The Adam optimizer (Kingma & Ba, 2015) is used with a learning rate of $\alpha = 0.05, \beta = 0.005$. AUC (Area Under Curve) and F1-score (Huang et al., 2015) are used as aggregate performance measures. The feature extractor $\mathcal{F}_{\psi}$ has an input dimension of 36 and an output dimension of 18, while the financial distress predictor $\mathcal{P}_{\theta}$ has an input dimension of 18 and an output dimension of 2. All methods are run for 3000 epochs, with models being updated based on improvements in both AUC and F1-score on the training set. In sub-experiments in the SFDP task, we estimate prediction performance using 5-fold cross-validation, with each testing fold providing one performance estimate. However, in the

Dynamicity & Heterogeneity	Method	Data Γ	Metric	Benchmark Year 2018	2019	2020	2021	Average
\	RF-Meta	0	AUC	84.20 (3.32)	85.10 (2.15)	86.30 (2.65)	77.56 (4.77)	83.29
			F1	37.12 (2.69)	39.89 (5.38)	42.69 (5.70)	32.52 (6.80)	38.06
	MLP-Meta		AUC	84.16 (2.61)	86.75 (3.52)	90.70 (1.53)	88.14 (2.89)	87.44
			F1	46.46 (3.60)	48.91 (5.32)	57.07 (4.99)	56.59 (3.74)	52.26
	FGAN-Meta	1	AUC	79.28. (5.18)	85.23 (2.22)	88.93 (2.93)	86.36 (1.69)	84.95
			F1	46.81 (3.66)	46.38 (6.12)	55.96 (3.01)	48.57 (5.10)	49.43
		2	AUC	81.26 (5.10)	84.34 (5.17)	90.06 (2.15)	85.26 (2.37)	85.23
			F1	44.59 (6.64)	43.95 (6.01)	55.13 (3.54)	51.45 (3.40)	48.78
		3	AUC	82.89 (5.12)	83.92 (4.30)	88.61 (1.54)	87.43 (2.30)	85.71
			F1	46.42 (5.43)	47.03 (6.29)	56.79 (3.83)	50.78 (2.66)	50.26
Dynamics	TFGAN-Meta		AUC	87.67 (2.13)	86.83 (3.99)	90.22 (1.63)	89.07 (2.64)	88.45
			F1	**59.25 (2.94)**	47.56 (5.71)	60.10 (3.00)	52.62 (4.23)	54.88
Heterogeneity	SFGAN-Meta	1	AUC	81.36 (5.96)	84.72 (6.45)	90.13 (2.19)	86.76 (3.96)	85.74
			F1	52.11 (5.35)	52.60 (5.82)	61.48 (3.65)	56.72 (2.32)	55.73
		2	AUC	85.97 (4.38)	**87.92 (6.54)**	**90.72 (2.53)**	86.88 (5.52)	87.87
			F1	52.08 (4.37)	52.65 (5.41)	61.22 (3.21)	**58.65 (3.57)**	56.15
		3	AUC	80.98 (5.59)	85.79 (5.33)	**90.72 (3.47)**	**90.11 (2.80)**	86.90
			F1	51.40 (5.48)	50.67 (6.51)	60.36 (4.07)	57.36 (4.63)	54.95
Synchronicity & Dynamicity	*STFGAN-Meta*		AUC	**90.23 (4.12)**	87.63 (5.33)	90.34 (3.65)	89.42 (5.72)	**89.41**
			F1	57.32 (3.98)	**52.78 (3.16)**	**62.23 (4.34)**	58.53 (5.21)	**57.72**

Table 3. SFDP Results: Mean (Standard Deviation) of AUC (%) and F1-score (%).

DFDP task sub-experiment, we cannot apply K-fold cross-validation to ensure the generalization of results because the training set and test set are not randomly partitioned. Therefore, we resort to conducting the DFDP experiments in five trials and averaging the results as an alternative approach.

Our dataset contains the annual FDP task set $\{FDP_{2018}, FDP_{2019}, FDP_{2020}, FDP_{2021}\}$ with benchmark years from 2018 to 2021. For the SFDP task, we use the annual FDP tasks from 2018 to 2021 as the corresponding SFDP task, following the IID assumption. For the DFDP task, we choose 2020 and 2021 as the benchmark years and use the historical FDP experiences as the training set. Thus, the DFDP task $DFDP_{2020}$ has the training set $\{FDP_{2018}, FDP_{2019}\}$, and the test set $\{FDP_{2020}\}$. We set $\mathcal{K} = 2$, where $(sFDP^{tr}_{2020}, sFDP^{te}_{2020}) = \{(FDP_{2018}, FDP_{2019}), (FDP_{2019}, FDP_{2018})\}$. The DFDP task $DFDP_{2021}$ has the training set $\{FDP_{2018}, FDP_{2019}, FDP_{2020}\}$, and the test set $\{FDP_{2021}\}$. We set $\mathcal{K} = 6$ in $DFDP_{2021}$, where $(sFDP^{tr}_{2021}, sFDP^{te}_{2021}) = \{(FDP_{2018}, FDP_{2019}), (FDP_{2019}, FDP_{2018}), (FDP_{2019}, FDP_{2020}), (FDP_{2020}, FDP_{2019}), (FDP_{2018}, FDP_{2020}), (FDP_{2020}, FDP_{2018})\}$.

Results

In this section, we conduct a series of experiments to evaluate the effectiveness of our proposed method.

Evaluation in SFDP task

We conducted a sub-experiment to validate the effectiveness of our proposed method in the SFDP task. Table 3 presents the experimental results, highlighting the best results for each annual SFDP task. Our findings suggest that interactions significantly impact the SFDP task, and it is necessary to handle the heterogeneity and dynamicity of interactions carefully.

Firstly, we found that graph-based methods generally outperform ML-based methods when they address the dynamicity or heterogeneity of interactions. *STFGAN-Meta* achieves the highest average F1-score (57.72%), 10.45% higher than the SOTA MLP-Meta model, and the highest AUC (89.41%), which is 2.25% higher than the MLP-Meta model. This result shows that interactions significantly impact improving the performance of the SFDP task. Only the FGAN-Meta method performs worse than MLP-Meta, indicating that ignoring the dynamicity and heterogeneity of interactions can lead to misuse and harm the performance of graph-based models. Secondly, the SFGAN-Meta models generally outperform the FGAN-Meta methods, indicating the significant role of heterogeneity of interactions in the SFDP task. Adding historical interactions to SFGAN-Meta and FGAN-Meta methods does not lead to a significant performance gain for the SFDP task, indicating

Dynamicity & Heterogeneity	Method	Data Γ	Metric	Benchmark Year		Average
				2020	2021	
\	RF-Meta	0	AUC	**88.85 (1.58)**	87.39 (1.32)	88.12
			F1	39.84 (2.27)	43.58 (1.06)	41.71
	MLP-Meta		AUC	86.84 (2.26)	**89.30 (1.10)**	88.07
			F1	50.29 (1.84)	50.44 (2.40)	50.37
	FGAN-Meta	1	AUC	83.43 (3.68)	86.53 (1.25)	84.98
			F1	40.37 (3.56)	44.16 (1.83)	42.27
		2	AUC	86.72 (2.45)	87.24 (5.38)	86.98
			F1	42.03 (5.27)	44.78 (3.25)	43.41
		3	AUC	84.32 (4.59)	85.36 (2.30)	84.84
			F1	41.68 (4.18)	42.84 (4.37)	42.26
Dynamicity	TFGAN-Meta	3	AUC	86.38 (4.73)	88.14 (5.36)	87.26
			F1	50.88 (5.28)	50.96 (3.43)	50.92
Heterogeneity	SFGAN-Meta	1	AUC	89.12 (1.00)	88.52 (3.01)	88.82
			F1	52.61 (2.26)	51.77 (3.83)	52.19
		2	AUC	88.38 (4.24)	86.26 (5.32)	87.32
			F1	51.52 (5.12)	**53.06 (3.62)**	52.29
		3	AUC	86.85 (4.57)	86.89 (1.35)	86.87
			F1	51.82 (5.21)	51.58 (3.29)	51.70
Heterogeneity & Dynamicity	*STFGAN-Meta*	3	AUC	88.79 (3.27)	89.26 (1.80)	**89.03**
			F1	**52.73 (4.68)**	52.65 (4.68)	**52.69**

Table 4. DFDP Results: Mean (Standard Deviation) of AUC (%) and F1-score (%).

that simply stacking historical interactions does not provide noticeable benefits. Thirdly, the TFGAN-Meta method outperforms all FGAN-Meta methods, demonstrating the importance of the dynamicity of interactions for the SFDP task. Leveraging information from historical interactions requires temporal aggregation rather than simple stacking. Finally, our proposed method outperforms all other graph-based methods, demonstrating the importance of synchronously addressing the heterogeneity and dynamicity of interactions.

Evaluations in DFDP task

The experimental results of the DFDP task are presented in Table 4. Our study on the DFDP task produced results similar to those observed in the SFDP task. Although interaction information remains beneficial for the DFDP task, addressing the heterogeneity and dynamicity of interactions is crucial for leveraging complicated interactions. The results demonstrate that graph-based methods outperform ML-based methods in the DFDP task when capturing the dynamics or heterogeneity of interactions. However, the performance gain from interactions information in the DFDP task is lower than in the SFDP task. The best-performing graph model, *STFGAN-Meta*, outperforms the best-performing ML-based model, MLP-Meta, by 4.61% in F1-score. In contrast, the top-performing graph model, *STFGAN-Meta*, in the SFDP task outperforms the top-performing ML-based model, MLP-Meta, by 10.45% in F1-score. Furthermore, the SFGAN-Meta and TFGAN-Meta models outperform the FGAN-Meta model, indicating the significance of interactions' heterogeneity and dynamics in the DFDP task. Our proposed method outperforms existing methods, achieving the best average performance, with an AUC of 89.03% and an F1-score of 52.69%. It demonstrates the superiority of our model in handling interaction information in the DFDP task.

Both results in the SFDP and DFDP tasks suggest that our model effectively captures the heterogeneous and dynamic interactions in the DFDP and SFDP tasks and can serve as a promising solution to the challenges associated with interaction modelling in financial domains.

Task	Method	Metric	Benchmark Year				Average
			2017	2018	2019	2020	
SFDP	*STFGAN-Meta*	AUC	90.23 (4.12)	87.63 (5.33)	90.34 (3.65)	89.42 (5.72)	89.41
		F1	57.32 (3.98)	52.78 (3.16)	62.23 (4.34)	58.53 (5.21)	57.72
	STFGAN	AUC	89.71 (4.63)	86.52 (4.27)	88.48 (3.13)	87.23 (5.10)	87.99
		F1	56.34 (4.18)	51.87 (4.65)	61.75 (3.85)	57.34 (3.32)	56.83
DFDP	*STFGAN-Meta*	AUC	\	\	88.79 (3.27)	89.26 (1.80)	89.03
		F1	\	\	52.73 (4.68)	52.65 (4.68)	52.69
	STFGAN	AUC	\	\	81.34 (5.24)	83.59 (3.72)	82.47
		F1	\	\	42.24 (6.73)	44.23 (5.35)	43.24

Table 5. Ablation Experiments to Validate Meta-learning Optimization.

Ablation Experiments

We conducted ablation experiments to verify the effectiveness of our proposed module, Meta-learning Optimization. Specifically, we compared our proposed method, *STFGAN-Meta*, with the STFGAN method without the Meta-learning Optimization module on both SFDP and DFDP tasks. The experimental results are presented in Table 5. Our experiments show that using meta-optimization can improve the performance of models in both DFDP and SFDP tasks. In the SFDP task, applying meta-optimization to STFGAN leads to a significant performance gain, increasing its AUC from 87.99% to 89.41% and its F1-score from 56.83% to 57.72%. In the DFDP task, the performance gain from meta-optimization is even more pronounced. Applying meta-optimization increases the AUC value from 82.47% to 89.03% and the F1-score from 43.24% to 52.69%. Our findings indicate that Meta-learning Optimization can improve performance for both tasks and is particularly noticeable in DFDP tasks that involve distress concept shift.

Discussion

Contributions and Implications

This study makes several significant contributions to the field of FDP.

Firstly, our work contributes to the FDP study by introducing the second valuable characteristic of interactions, dynamicity, and emphasizing the advantage of leveraging the synchronicity of heterogeneity and dynamicity in company interactions. Traditional FDP methods have often ignored the interactions (Sun & Li, 2012; Sun et al., 2013; Wang et al., 2021) or focused solely on their heterogeneity (Kou et al., 2021; Tobback et al., 2017; Yang et al., 2021), which can introduce sampling bias and lead to misuse of interaction information. In contrast, our work introduces dynamicity and emphasizes the synchronicity of heterogeneity and dynamicity in company interactions. Our findings demonstrate the advantage of leveraging the synchronicity of dynamicity and heterogeneity of interactions in the FDP study. It provides valuable insights for future FDP tasks by addressing sampling bias or misuse of interactions.

Secondly, our study represents a pioneering attempt to bridge the gap between SFDP and DFDP studies and unify both approaches by linking FDP studies with meta-learning literature. Previous work has to trade-off between leveraging interaction information or historical prediction experience due to the inconsistency of the SFDP and DFDP models (Kou et al., 2021; Sun et al., 2018; Wang et al., 2021), which limits the popularity and generalization of FDP tasks. In contrast, our work complements existing methods by providing a meta-learning mechanism to adapt FDP models to varying FDP scenarios. To the best of our knowledge, this is the first attempt to introduce meta-learning into the FDP study. We demonstrate the potential of meta-learning on FDP tasks, making FDP models more generalizable in ever-changing financial scenarios. Furthermore, our study resolves the inconsistency of the SFDP and DFDP models, which mitigates the modelling burdens on FDP researchers and alleviates the complexities in maintaining FDP services.

Finally, our study contributes to the design science paradigm of IS by proposing a novel FDP method, the *STFGN-Meta*, which effectively leverages interactions in both SFDP and DFDP tasks. Our proposed model utilizes a Spatial Aggregation module and a Temporal Aggregation module to handle interaction heterogeneity and dynamicity, allowing for more accurate predictions. Moreover, the proposed Meta-learning Optimization module helps to optimize the prediction services by imitating the prediction experience and enduring simulated distress concept drifts. With promising prediction performance, our proposed method can be a powerful tool for risk assessment and decision-making in the financial industry.

Limitations and Future Research

While our work provides valuable contributions, it is essential to acknowledge its limitations. Firstly, our dataset restricts the time frame from January 2015 to December 2020, covering 5 SFDP tasks and 2 DFDP tasks. Only 2 DFDP tasks limit our exploration of the model's generalization ability on DFDP tasks. Future research can consider expanding the dataset to a broader time frame. Secondly, our study only collected single-source financial characteristics and interactions to support FDP tasks. While our work primarily focuses on the heterogeneity of interactions, it overlooks the heterogeneity of financial characteristics (Wang et al., 2021). Future research can explore leveraging multi-source heterogeneous financial characteristics in general FDP research. Thirdly, the ST-warning mechanism limits the availability of real-time FDP studies. In future work, we can replace the annually updated labels with a practical and real-time index of financial distress risks for companies, providing stakeholders with more timely distress warnings. Finally, our experiments fully compared the performance of GNNs-based models using company interactions considering the promising performance of GNNs on graph data. However, some non-GNNs-based methods, such as wvRN, can be included for comparison. Future work will further explore the effectiveness of non-GNNs-based work in leveraging company interactions.

Conclusions

This work proposes a novel approach that unifies FDP tasks in stationary and dynamic scenarios by leveraging interaction information. Specifically, we address the synchronicity of heterogeneity and dynamicity in interactions by integrating the Spatial Aggregation and Temporal Aggregation modules. Moreover, the Meta-learning Optimization module enables our model to learn from historical prediction experiences and tackle potential distress concept drifts. We empirically evaluate our proposed method using publicly accessible Chinese listed companies' interaction data in 2015-2020. Experimental evaluation demonstrates that our approach not only outperforms several FDP methods but also provides a comprehensive and accurate understanding of the role of interactions in FDP tasks. Our study will significantly impact future works in FDP research, providing suggestions to make better use of interaction information and design more generalizable FDP models in ever-changing financial scenarios.

References

Alfaro, E., García, N., Gámez, M. & Elizondo, D. (2008). Bankruptcy forecasting: An empirical comparison of AdaBoost and neural networks. *Decision Support Systems, 45(1)*, 110–122.

Almeida, P. R., Oliveira, L. S., Britto, A. S. & Sabourin, R. (2018). Adapting dynamic classifier selection for concept drift. *Expert Systems with Applications, 104*, 67–85.

Bi, W., Xu, B., Sun, X., Wang, Z., Shen, H. & Cheng, X. (2022). Company-as-tribe: Company financial risk assessment on tribe-style graph with hierarchical graph neural networks. *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining,* 2712–2720.

Bishop, C. M. & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning,* vol. 4. 4. Springer.

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32.

Chen, N., Ribeiro, B. & Chen, A. (2016). Financial credit risk assessment: A recent review. *Artificial Intelligence Review, 45(1)*, 1–23.

ChinaScope. **https://www.chinascope.com/**.

CSMAR. **https://cn.gtadata.com/**.

Dong, Y., Chawla, N. V. & Swami, A. (2017). Metapath2vec: Scalable representation learning for heterogeneous networks. *Proceedings of the 23rd ACM SIGKDD Conference on Knowledge Discovery and Data Mining,* 135–144.

Du, L., Wang, Y., Song, G., Lu, Z. & Wang, J. (2018). Dynamic network embedding: An extended approach for skip-gram based network embedding. *Proceedings of 27th International Joint Conferences on Artificial Intelligence,* 2086–2092.

Du Jardin, P. (2015). Bankruptcy prediction using terminal failure processes. *European Journal of Operational Research, 242(1)*, 286–303.

Du Jardin, P. (2018). Failure pattern-based ensembles applied to bankruptcy forecasting. *Decision Support Systems, 107*, 64–77.

Fan, Y., Ju, M., Zhang, C. & Ye, Y. (2022). Heterogeneous temporal graph neural network. *Proceedings of the 2022 SIAM International Conference on Data Mining,* 657–665.

Geng, R., Bose, I. & Chen, X. (2015). Prediction of financial distress: An empirical study of listed Chinese companies using data mining. *European Journal of Operational Research, 241(1)*, 236–247.

Hoadley, B. (1971). Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case. *The Annals of Mathematical Statistics, 42(6)*, 1977–1991.

Huang, H., Xu, H., Wang, X. & Silamu, W. (2015). Maximum F1-score discriminative training criterion for automatic mispronunciation detection. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 23(4)*, 787–797.

Jiang, C., Wang, Z., Liu, X., et al. (2022). Capturing heterogeneous interactions for financial risk prediction of SMEs. *Proceedings of the 26th Pacific Asia Conference on Information Systems,* 56, 1361–1369.

Karanikolos, M., Mladovsky, P., Cylus, J., Thomson, S., Basu, S., Stuckler, D., Mackenbach, J. P. & McKee, M. (2013). Financial crisis, austerity, and health in Europe. *The Lancet, 381(9874)*, 1323–1331.

Kingma, D. P. & Ba, J. (2015). Adam: A method for stochastic optimization. *Proceedings of the 3rd International Conference on Learning Representations,* 1–15.

Kou, G., Xu, Y., Peng, Y., Shen, F., Chen, Y., Chang, K. & Kou, S. (2021). Bankruptcy prediction for SMEs using transactional data and two-stage multiobjective feature selection. *Decision Support Systems, 140(113429)*, 1–14.

Lan, Z., Gu, J., Zheng, Z., Thakur, R. & Coghlan, S. (2010). A study of dynamic meta-learning for failure prediction in large-scale systems. *Journal of Parallel and Distributed Computing, 70(6)*, 630–643.

Lang, L. H. & Stulz, R. (1992). Contagion and competitive intra-industry effects of bankruptcy announcements: An empirical analysis. *Journal of Financial Economics, 32(1)*, 45–60.

LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE, 86(11)*, 2278–2324.

Li, Q., Chen, Y., Jiang, L. L., Li, P. & Chen, H. (2016). A tensor-based information framework for predicting the stock market. *ACM Transactions on Information Systems, 34(2)*, 1–30.

Lin, W.-Y., Hu, Y.-H. & Tsai, C.-F. (2011). Machine learning in financial crisis prediction: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42(4)*, 421–436.

Marini, F., Magrì, A. L. & Bucci, R. (2007). Multilayer feed-forward artificial neural networks for class modeling. *Chemometrics and Intelligent Laboratory Systems, 88(1)*, 118–124.

Mitchell, T. M., Caruana, R., Freitag, D., McDermott, J., Zabowski, D., et al. (1994). Experience with a learning personal assistant. *Communications of the ACM, 37(7)*, 80–91.

Óskarsdóttir, M. & Bravo, C. (2021). Multilayer network analysis for improved credit risk prediction. *Omega, 105(102520)*, 1–13.

Perozzi, B., Al-Rfou, R. & Skiena, S. (2014). Deepwalk: Online learning of social representations. *Proceedings of the 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining,* 701–710.

Schlimmer, J. C. & Granger, R. H. (1986). Incremental learning from noisy data. *Machine Learning, 1*, 317–354.

Sun, J., Fujita, H., Chen, P. & Li, H. (2017). Dynamic financial distress prediction with concept drift based on time weighting combined with Adaboost support vector machine ensemble. *Knowledge-Based Systems, 120*, 4–14.

Sun, J., Lang, J., Fujita, H. & Li, H. (2018). Imbalanced enterprise credit evaluation with DTE-SBD: Decision tree ensemble based on SMOTE and bagging with differentiated sampling rates. *Information Sciences, 425*, 76–91.

Sun, J. & Li, H. (2011). Dynamic financial distress prediction using instance selection for the disposal of concept drift. *Expert Systems with Applications, 38(3)*, 2566–2576.

Sun, J. & Li, H. (2012). Financial distress prediction using support vector machines: Ensemble vs. individual. *Applied Soft Computing, 12(8)*, 2254–2265.

Sun, J., Li, H. & Adeli, H. (2013). Concept drift-oriented adaptive and dynamic support vector machine ensemble with time window in corporate financial risk prediction. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 43(4)*, 801–813.

Sun, J., Li, H., Fujita, H., Fu, B. & Ai, W. (2020). Class-imbalanced dynamic financial distress prediction based on Adaboost-SVM ensemble combined with SMOTE and time weighting. *Information Fusion, 54*, 128–144.

Tobback, E., Bellotti, T., Moeyersoms, J., Stankova, M. & Martens, D. (2017). Bankruptcy prediction for SMEs using relational data. *Decision Support Systems, 102*, 69–81.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. & Polosukhin, I. (2017). Attention is all you need. *Proceedings of the 31st Neural Information Processing Systems,* 30, 1–15.

Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P. & Bengio, Y. (2017). Graph attention networks. *Stat, 1050*, 1–12.

Vilalta, R. & Drissi, Y. (2002). A perspective view and survey of meta-learning. *Artificial Intelligence Review, 18*, 77–95.

Wang, G., Chen, G., Zhao, H., Zhang, F., Yang, S. & Lu, T. (2021). Leveraging multisource heterogeneous data for financial risk prediction: A novel hybrid-strategy-based self-adaptive method. *MIS Quarterly, 45(4)*, 1949–1998.

Wang, X., Ji, H., Shi, C., Wang, B., Ye, Y., Cui, P. & Yu, P. S. (2019). Heterogeneous graph attention network. *Proceedings of the 30th the World Wide Web Conference,* 2022–2032.

Yang, S., Zhang, Z., Zhou, J., Wang, Y., Sun, W., Zhong, X., Fang, Y., Yu, Q. & Qi, Y. (2021). Financial risk analysis for SMEs with graph-based supply chain mining. *Proceedings of the 29th International Conference on International Joint Conferences on Artificial Intelligence,* 4661–4667.

Zhang, C., Song, D., Huang, C., Swami, A. & Chawla, N. V. (2019). Heterogeneous graph neural network. *Proceedings of the 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining,* 793–803.

Zhang, Y., Li, Y., Zhou, X. & Luo, J. (2020). cST-ML: Continuous spatial-temporal meta-learning for traffic dynamics prediction. *Proceedings of 20th IEEE International Conference on Data Mining,* 1418–1423.