

多模块注意力协同的异构图神经网络产业链风险识别模型

蓝仲舒¹, 应时², 李宁², 田相波², 李田港²

¹(武汉大学 国家网络安全学院, 武汉 430072)

²(武汉大学 计算机学院, 武汉 430072)

E-mail: yingshi@whu.edu.cn

摘要: 风险识别通过精准定位产业链的脆弱节点以阻断风险传导, 对于维护产业链完整性至关重要. 然而, 现有方法缺乏对产业链节点进行静态指标级、动态时序级和结构空间级的协同建模. 因此, 本文提出一种多模块注意力协同的异构图神经网络产业链风险识别模型(MGRI). MGRI首先根据企业所属行业动态学习并分配各属性维度的权重, 以突出关键指标的重要性; 同时捕捉低频财务时间序列蕴含的风险演变信息, 形成全局动态时序表征; 最后以动静融合特征作为初始表示聚合关系感知的邻域信息, 生成包含产业链结构空间依赖的最终嵌入并用于风险识别, 实现了跨维度多层次特征的协同建模. 实验验证了MGRI在多个真实产业链数据集上准确识别风险企业方面优于最先进的方法.

关键词: 产业链; 风险识别; 异构图神经网络; 注意力机制; 多模块协同

中图分类号: TP391

文献标识码: A

文章编号: 1000-1220(2026)05-1032-09

Industrial Chain Risk Identification Model Based on Heterogeneous Graph Neural Network with Multi-module Attention Collaboration

LAN Zhongshu¹, YING Shi², LI Ning², TIAN Xiangbo², LI Tiangang²

¹(School of Cyber Science and Engineering, Wuhan University, Wuhan 430072, China)

²(School of Computer Science, Wuhan University, Wuhan 430072, China)

Abstract: Risk identification is essential for maintaining the integrity of industry chains by precisely locating vulnerable nodes to interrupt risk propagation. However, existing methods lack collaborative modeling of industry nodes across static-indicator level, dynamic-time-series level, and structural-spatial level representations. Therefore, this paper proposes an industrial chain risk identification model based on a heterogeneous graph neural network with multi-module attention collaboration (MGRI). MGRI first dynamically learns and assigns weights to attribute dimensions based on a company's industry to emphasize the importance of key indicators; it simultaneously captures the risk evolution information embedded in low-frequency financial time series to form a global dynamic temporal representation; finally, using the fused static and dynamic features as initial representations, it aggregates relation-aware neighborhood information to generate the final embeddings that contain the structural-spatial dependencies of the industry chain for risk identification, thereby achieving collaborative modeling of cross-dimensional, multi-level features. Experiments on two real-world industry chain datasets demonstrate that MGRI outperforms state-of-the-art methods in accurately identifying at-risk enterprises.

Keywords: industry chain; risk identification; heterogeneous graph neural network; attention mechanism; multi-module collaboration

0 引言

在全球价值链深度重构与数字化转型加速的双重背景下, 产业链网络的风险传导机制呈现出前所未有的复杂性^[1,2]. 企业风险识别通过精准定位产业链网络拓扑中的脆弱性节点, 阻断风险的跨层传导, 成为维护产业链完整性、预防系统性经济风险的核心机制. 在当前数字化转型加速的时代背景下, 数据驱动决策已成为企业风险识别的核心手段^[3]. 在企业风险识别中, 所需数据主要涵盖企业内部及外部多个层面的信息. 内部数据主要包括企业财务指标、生产能力、销售库存数据等多维度信息, 这些数据能够直接反映企业

的偿债能力和经营稳定性, 成为风险评估的重要依据; 而外部数据则涵盖产业链上下游关系、行业市场波动、新闻舆情以及宏观经济指标等, 这些数据能够反映企业所处产业链的整体风险传导效应.

但由于商业机密保护和竞争壁垒, 产业链中存在数据敏感性和信息不透明性的问题, 生产能力、销售库存数据等内部数据难以获取^[4], 基于经过审计公开的公司财务数据进行风险识别是一直以来可行高和主流的研究主题^[5-7]. 然而, 现有方法在处理财务数据时, 常将其视为扁平化特征, 忽略了其内在的结构性、以及不同指标在特定情境下的差异化重要性, 使得关键指标的表达不够精准. 此外, 公司财务数据审计公开的

收稿日期: 2025-06-11 收修改稿日期: 2025-07-31 基金项目: 国家重点研发计划项目(2022YFB3304300)资助; 国家自然科学基金项目(62472329, 62072342)资助. 作者简介: 蓝仲舒, 男, 2001年生, 硕士研究生, 研究方向为人工智能、图神经网络; 应时(通信作者), 男, 1965年生, 博士, 教授, CCF会员, 研究方向为人工智能、软件工程、知识图谱; 李宁, 男, 1997年生, 博士研究生, 研究方向为图神经网络、无监督学习、产业链分析; 田相波, 男, 1995年生, 博士研究生, 研究方向为软件智能化运维、图神经网络、因果推断; 李田港, 男, 1995年生, 博士研究生, CCF会员, 研究方向为人工智能、软件工程、强化学习.

时间粒度均为年度或者季度,传统时序模型在处理低频发布的时间序列财务数据时,难以有效捕捉其动态演变规律和关键风险突变时间点并形成综合表征^[8,9]。更重要的是,如何将从不同维度独立提取的信息进行有效的协同整合,使它们相互增强而非相互干扰,以全面提升风险识别的准确性,仍然是一个亟待解决的关键挑战。

针对上述挑战,本文提出一种多模块注意力协同的异构图神经网络产业链风险识别模型(MGRI),旨在深度融合企业的多粒度时空特征。首先,静态指标行业感知增强模块使用注意力机制根据企业所属行业自适应地为各项财务指标分配权重并在节点的所有时间步中共享,以强化关键指标表征,此举突破了传统方法对财务特征扁平化及情境无差别处理的局限;同时,低频时序动态特征学习模块以Transformer编码器为核心,凭借其自注意力机制有效学习财务序列中的长期依赖关系和复杂演变模式,并由终端时间注意力层进行聚合,提炼出对风险识别至关重要的全局时序特征。然后,产业链异构图信息聚合模块是元路径引导的层次注意力异构图神经网络,将处理后的关键指标强化表征与提取的动态聚合时序特征共同作为增强输入,利用图注意力网络进行节点邻域信息的聚合,随后通过元路径间注意力机制评估不同元路径所捕获企业关系信息的相对重要性并加权融合这些多角度的节点表示。

为了清晰划分各模块在整体模型中的功能定位与协同关系,本文将静态指标行业感知增强模块、低频时序动态特征学习模块及其融合表征的输出层共同构成企业动静双维度特征编码器。该编码器输出的综合特征,将作为产业链异构图信息聚合模块的输入,用于进行复杂的结构关系建模。这种设计使得MGRI能够同时感知行业差异化的关键财务特征和企业风险演变信息,并通过图网络在产业链的异构图结构中进行传播与推理,深入挖掘产业链中企业间复杂的结构依赖和风险传导路径。本文的主要贡献如下:

1) 提出了一种针对企业财务数据设计的动静双维度特征编码器,该编码器集成了静态指标行业感知增强模块和低频时序动态特征学习模块,能够有效学习财务数据中各指标的内在重要性及其动态时序特征,突破了传统财务数据处理

的局限。

2) 构建了一个关系感知的产业链异构图信息聚合模块,该模块以企业动静双维度特征编码器输出的融合特征为输入,并通过元路径引导的异构图神经网络实现对邻域信息的关系感知聚合,能够精准捕获并建模产业链中企业间的多维结构依赖与潜在风险传导路径,实现了多模块的协同表征学习。

3) 在多个真实的产业链数据集及公开财务数据上进行的实验验证了MGRI模型的优越性,MGRI相比SOTA方法在召回率、F1分数和AUC分别提高了2.76%、7.90%和2.27%。此外,消融实验进一步验证了各个模块的有效性及其协同作用。

1 相关工作

1.1 图神经网络

图通常可以被分类为同构图和异构图^[10]。同构图神经网络

主要处理节点和边类型单一的图结构,它们通过迭代地聚合邻居节点的信息来学习目标节点的有效表征。例如,图卷积网络(GCN)^[11]利用谱卷积的一阶近似推导层间传播规则,有效融合图拓扑与节点特征。图注意力网络(GAT)^[12]则通过引入注意力机制,为不同邻居节点分配不同的权重,从而更精细地捕捉节点间的关系。然而,现实中的产业链和金融网络往往具有显著的异构性,同构图模型难以充分表达这种复杂的特性。

异构图神经网络能够有效处理包含多种类型节点和边的复杂网络数据,并通过学习不同类型对象及其关系的特定语义信息,从而更精确地建模和理解网络中的多元化交互模式与高阶依赖^[13]。例如HetCAN^[14]通过重启随机游走采样邻居,并设计专门的内容编码与类型感知的邻居聚合模块来学习节点在异构图中的综合嵌入。HAN^[15]通过节点级和语义级的层级注意力机制,分别学习元路径引导下邻居节点的重要性的不同元路径的语义贡献。HGT^[16]采用Transformer架构,并针对异构图的元关系设计参数化的注意力和消息传递机制,以高效处理大规模数据并学习节点特定类型的表示。

1.2 企业风险识别

早期的企业风险识别主要依赖统计模型与专家规则。其中,财务指标模型(如Altman Z-Score^[17]、Ohlson O-Score^[18])通过线性组合关键财务比率来预测企业破产风险,但这类模型依赖复杂的人工特征工程,且难以捕捉数据间的非线性关系。随后出现的基于机器学习的模型(如支持向量机SVM^[19,20]、逻辑回归^[20])相较于传统统计模型,能够更好地处理非线性关系和高维数据;然而,它们通常将每个企业视为独立的实体进行分析,因而难以有效捕捉企业间的复杂关联以及风险在产业链网络中的传导效应。

图神经网络因其强大的图结构数据表征学习能力,在社交网络分析、推荐系统、生物信息学等领域取得了显著成功,并逐渐被应用于金融风控和企业风险识别领域^[10,21,22]。目前大多数的企业风险评估方法都是基于异构图神经网络构建,这些方法能够区分并利用不同类型节点和边的特定语义信息,从而更全面地理解企业间的复杂依赖关系和风险传导路径。例如,HetCAN^[23]利用协同元路径,通过由两个注意力分数和成对重要性学习组成的协同注意力机制来提取有效的风险特征。Wei等人^[24]使用由超图神经网络和异构图神经网络组成的传染风险编码器,分别从超边和企业知识图谱之间复杂的异构关系这两个方面来建模传染风险。ICRE-AC^[25]通过图加权融合解决产业链关系复杂性,并利用注意力机制补全缺失属性,实现了对产业链异质图节点风险的高效评估。

2 预备知识

2.1 理论基础

定义1. 异构图(Heterogeneous Graph)

异构图定义为 $G=(V, E)$,其中 V 代表节点的集合, E 代表边的集合。其中包含两个映射函数:一个节点类型映射函数 $\varphi_v: V \rightarrow \mathcal{A}$ 和一个链接类型映射函数 $\psi_e: E \rightarrow \mathcal{R}$ 。其中 \mathcal{A} 和 \mathcal{B} 表示预定义的对象类型和链接类型的集合,并且 $|\mathcal{A}| + |\mathcal{R}| > 2$ 。

基于异构图的一般定义,本文研究的产业链场景可以建

模为一个具体的异构图 $G_{ic} = (V, E)$. 在此图中, V 代表节点的集合, 包含了产业链中的不同实体, 即企业节点和行业类别节点. 具体地, $V = \{v_1, v_2, \dots, v_n\}$. E 代表边的集合, 描述了这些节点之间的关系, 例如供应关系、投资关系、竞争关系、行业归属关系等. 此外, 将产业链异构图包含节点属性信息定义为 F . 特别地, 对于图中的企业类型节点 $v \in V$ (即 $f_n(v) = \text{企业}$), 其关联的属性 $f_v \in F$ 由从公开渠道获取的多时间步财务数据构成. 该属性可以表示为一个时间序列 $f_v = (x_v^{(1)}, x_v^{(2)}, \dots, x_v^{(k)})$, 其中 K 是观测的时间步总数, $x_v^{(k)} \in \mathbb{R}^D$ 是企业 v 在第 k 个时间步的 D 维财务特征向量.

定义 2. 元路径 (Meta-Path)

元路径 Φ 是在异构图上定义的一条路径, 它表示为 $A_1 \xrightarrow{R_1} A_2 \xrightarrow{R_2} \dots \xrightarrow{R_i} A_{i+1}$ 的形式, 其中 $A_i \in \mathcal{A}$ 是通过节点类型映射函数 ψ_v 得到的节点类型, $R_j \in \mathcal{R}$ 是通过边类型映射函数 ψ_e 得到的关系类型. 这条路径定义了起始节点类型 A_1 和结束节点类型 A_{i+1} 之间的一个复合关系 $R_{\text{composite}} = R_1 \circ R_2 \circ \dots \circ R_i$, 其中 \circ 表示关系的组合运算符.

对于产业链知识图谱 G_{ic} 中的一条元路径 Φ , 元路径实例 Φ 是 G_{ic} 中遵循元路径 Φ 所定义的模式的一个具体节点序列

$\phi = (v_1, v_2, \dots, v_{n+1})$. 基于元路径的邻居 N_v^ϕ 是指通过元路径 ϕ 的至少一个实例与节点 v 相连接的节点集合. 元路径可达图 G^ϕ 是原始产业链知识图谱 G_{ic} 中指定元路径 ϕ 的子图, 包含遵循 ϕ 的所有元路径实例.

2.2 问题形式化

问题 1. 产业链风险识别

本文的研究目标是识别产业链中的企业在某一时间节点是否存在风险. 具体来说, 给定一个产业链异构图 $G_{ic} = (V, E)$, 对于每个企业节点 $v \in V$, 其风险状态 $y_v \in \{0, 1\}$ (其中 0 代表正常, 1 代表存在风险), 任务是基于从图 G_{ic} 中提取的节点自身特征和节点间的交互信息, 来预测产业链中每一个企业 v 存在风险的概率 \hat{y}_v .

3 方法

本文提出的 MGRI 模型, 其整体框架如图 1 所示. 该模型主要包含 3 个核心模块, 具体如下: (a) 静态指标行业感知增强模块, 其任务是根据企业的行业归属动态学习各属性维度的重要性, 强化关键指标的表征; (b) 低频时序动态特征学习模块, 负责捕捉企业财务数据序列中所蕴含的演化规律与长

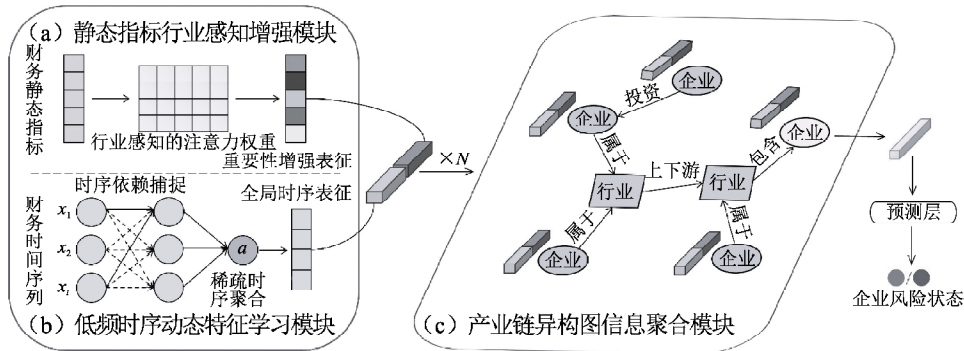


图 1 MGRI 整体架构

Fig. 1 Architecture of MGRI

程依赖特性, 形成全局动态时序视图; (c) 产业链异构图信息聚合模块, 该模块将前序模块提取的特征作为增强输入, 在元路径引导的产业链异构图谱上聚合邻域信息并融合多语义关系下的节点表征, 以学习企业间的复杂依赖并生成最终用于下游风险判别任务的节点嵌入. 其中 (a) 和 (b) 在逻辑上同属于企业动静双维度特征编码器, 与 (c) 协同实现产业链风险识别.

3.1 企业动静双维度特征编码器

企业动静双维度特征编码器在设计上集成了两个并行的核心子模块——静态指标行业感知增强模块和低频时序动态特征学习模块, 这两个模块分别从静态属性和动态演变两个维度对节点特征进行深度挖掘, 其输出最终被动静表征联合输出层融合, 为产业链异构图中的每个企业节点生成一个能全面反映其内在经营状况和风险动态的初始特征表示.

3.1.1 静态指标行业感知增强模块

企业财务数据包含多维度指标, 这些指标在不同行业背景下的重要性存在显著差异. 例如, 在集成电路产业链中, 对于上游的材料供应企业, 其库存周转率可能更为关键; 而对于下游的组装和销售企业, 则更侧重于应收账款周转率、营业收

入增长率等指标. 因此, 设计了一个静态指标行业感知增强模块, 该模块能够根据企业所属行业, 使用注意力机制动态学习并分配各财务指标的权重, 从而强化对风险识别至关重要的关键财务指标的表征, 提升模型对企业基本面风险的感知能力. 模块架构如图 2 所示.

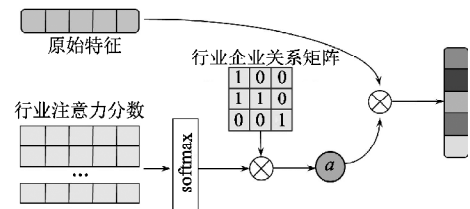


图 2 静态指标行业感知增强模块

Fig. 2 Indicator enhancement module

假设输入的企业财务特征为 $\mathbf{X} \in \mathbb{R}^{N \times T \times D}$, 其中 N 表示企业数量, T 表示时间步数, D 表示财务指标维度. 另有企业-行业映射矩阵 $\mathbf{M} \in \mathbb{R}^{N \times M}$, 其中 M 是行业总数, $M_{i,j} = 1$ 表示企业 i 属于行业 j , 否则为 0, 其中 $i \in \{1, 2, \dots, N\}$, $j \in \{1, 2, \dots, M\}$.

为了捕捉各行业不同财务指标的相对重要性,定义行业注意力参数矩阵 $\mathbf{A} \in \mathbb{R}^{M \times D}$, 其中 \mathbf{A}_j 表示行业 j 的指标权重分布. 该矩阵通过模型训练学习得到, 并通过 softmax 函数进行归一化处理:

$$\mathbf{w}_j = \text{softmax}(\mathbf{A}_j) = \frac{\exp(\mathbf{A}_j)}{\sum_{d=1}^D \exp(\mathbf{A}_{j,d})} \quad (1)$$

其中 $\mathbf{w}_j \in \mathbb{R}^D$ 是行业 j 的归一化特征权重向量, 满足 $\sum_{d=1}^D \mathbf{w}_{j,d} = 1$, 此权重向量反映了在行业 j 的背景下, 各个财务特征维度 d 对于风险评估的相对贡献度. 然后, 基于企业-行业映射关系, 为企业节点计算其特征权重向量 α_i . 鉴于单个企业节点可能同时关联至数量不定且组合各异的行业类别, 采用均值加权聚合相关行业的注意力权重, 以获得节点层面的最终注意力权重:

$$\alpha_i = \frac{\sum_{j=1}^M \mathbf{M}_{i,j} \mathbf{w}_j}{\sum_{j=1}^M \mathbf{M}_{i,j}} \quad (2)$$

其中 $\alpha_i \in \mathbb{R}^D$. 另外, 由于产业链异构图中存在的企业投资关系, 导致某些企业节点 i 不属于产业链异构图中任何行业类别, 对这些企业节点共用一套独立的注意力权重 α_0 .

最后, 对节点 i 的原始财务特征 $\mathbf{X}_i \in \mathbb{R}^{T \times D}$ 进行动态加权:

$$\hat{\mathbf{X}}_i = \mathbf{X}_i \odot \alpha_i \quad (3)$$

其中 $\mathbf{X}_i \in \mathbb{R}^{T \times D}$ 是企业 i 的加权特征表示, \odot 表示广播乘法操作, 得到加权后的特征张量 $\hat{\mathbf{X}}_i \in \mathbb{R}^{N \times T \times D}$. 对于任意节点 i , 其特征注意力权重 α_i 在所有时间步 T 上是共享的, 从而使模型能够基于企业节点的静态行业属性持久且自适应地关注对其风险状况最具判别力的特征维度.

3.1.2 低频时序动态特征学习模块

在产业链风险评估中, 其风险特征往往随着时间的推移展现出复杂的动态模式和演变趋势. 例如, 持续的盈利能力下滑、偿债指标的逐步恶化, 或是关键财务比率的突变, 都可能预示着潜在风险的累积或爆发. 因此, 利用 Transformer 编码器来捕捉序列中的长短期依赖关系, 并结合时间步注意力层来生成一个能够概括整个时间序列信息的固定维度节点表示, 从而解决长期依赖建模失效和短期波动过拟合等传统时序建模难题. 模块架构如图 3 所示.

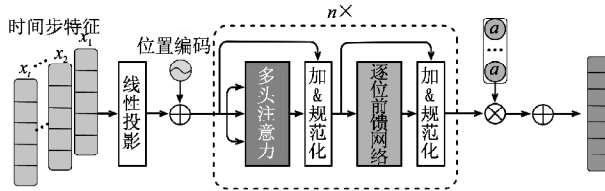


图3 低频时序动态特征学习模块

Fig. 3 Temporal feature learning module

企业财务数据包含成长能力相关指标, 如营业收入增长率、净利润增长率、总资产增长率等, 为避免静态指标行业感知增强模块可能弱化这些关键属性和引入噪声, 使用未经处理的原始企业财务特征 $\mathbf{X} \in \mathbb{R}^{N \times T \times D}$ 作为低频时序动态特征学习模块的输入.

1) 特征适配与时间位置注入

为使输入特征符合 Transformer 模型的内部处理维度, 首先采用线性投影层对原始维特征进行维度变换, 将其映射至

D' 维:

$$\mathbf{X}_{\text{proj}} = \text{Linear}(\mathbf{X}) = \mathbf{X}\mathbf{W}_{\text{proj}} + \mathbf{b}_{\text{proj}} \quad (4)$$

其中, $\mathbf{W}_{\text{proj}} \in \mathbb{R}^{D \times D'}$ 是投影权重矩阵, $\mathbf{b}_{\text{proj}} \in \mathbb{R}^{D'}$ 是偏置向量. 投影后的特征张量为 $\mathbf{X}_{\text{proj}} \in \mathbb{R}^{N \times T \times D'}$.

由于 Transformer 的自注意力机制本身不感知序列中元素的顺序, 需要引入位置编码来注入时序信息. 对于序列中的每个时间步位置 $pos \in \{0, \dots, T-1\}$ 和投影后特征的每个维度 $J \in \{0, \dots, D'/2-1\}$, 位置编码 $\mathbf{PE} \in \mathbb{R}^{T \times D'}$ 的计算方式如下:

$$\mathbf{PE}(pos, k) = \begin{cases} \sin\left(\frac{pos}{10000^{\frac{k}{D'}}}\right) & \text{if } k = 2j \\ \cos\left(\frac{pos}{10000^{\frac{k-1}{D'}}}\right) & \text{if } k = 2j + 1 \end{cases} \quad (5)$$

然后将学习到的位置编码 \mathbf{PE} 与投影后的特征 \mathbf{X}_{proj} 逐元素相加, 得到注入位置编码后的特征表示 \mathbf{X}_{pe} :

$$\mathbf{X}_{\text{pe}} = \mathbf{X}_{\text{proj}} + \mathbf{PE} \quad (6)$$

2) 基于自注意力的时序依赖捕捉

在时序依赖建模中, 首先将融合了位置信息的特征 \mathbf{X}_{pe} (亦表示为 $\mathbf{H}^{(0)}$) 送入一个由 L 层相同结构堆叠而成的 Transformer 编码器模块. 对于编码器中的第 l 层, 其输入为上一层的输出 $\mathbf{H}^{(l-1)}$, 其中 $l \in \{1, 2, \dots, L\}$. 每一编码层的核心计算均由多头自注意力 (Multi-Head Self-Attention, MHSA) 机制和基于位置的前馈神经网络 (Position-wise Feed-Forward Network, FFN) 构成.

具体而言, 输入 $\mathbf{H}^{(l-1)}$ 首先经过 MHSA 子层处理. 其输出随后与原始输入 $\mathbf{H}^{(l-1)}$ 进行残差连接, 并经过 Dropout 和层归一化操作, 得到中间表示 \mathbf{H}_{attn} :

$$\mathbf{H}_{\text{attn}} = \text{LayerNorm}(\mathbf{H}^{(l-1)} + \text{Dropout}(\text{MHSA}(\mathbf{H}^{(l-1)}))) \quad (7)$$

MHSA 的具体计算过程如下:

$$\begin{aligned} \text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) &= \mathbf{W}^Q [\text{head}_1 \parallel \text{head}_2 \parallel \dots \parallel \text{head}_n] \\ \text{head}_j &= f(\mathbf{Q}\mathbf{W}_j^Q, \mathbf{K}\mathbf{W}_j^K, \mathbf{V}\mathbf{W}_j^V) \\ f(\mathbf{Q}, \mathbf{K}, \mathbf{V}) &= \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \end{aligned} \quad (8)$$

其中, 查询矩阵 \mathbf{Q} 、键矩阵 \mathbf{K} 和值矩阵 \mathbf{V} 是由 $\mathbf{H}^{(l-1)}$ 经过不同的线性变换得到, $\mathbf{W}_j^Q, \mathbf{W}_j^K, \mathbf{W}_j^V$ 是特定于头 j 的可学习参数矩阵, \mathbf{W}^O 是用于整合所有头输出的参数矩阵, d_k 是键向量的维度, 用于缩放点积结果以稳定梯度.

随后, 中间表示 \mathbf{H}_{attn} 被送入 FFN 子层, 对序列中每个时间步的表示进行独立变换. FFN 由两层线性变换构成, 中间使用 ReLU 激活函数. 与 MHSA 子层类似, FFN 子层的输出也与 \mathbf{H}_{attn} 进行残差连接, 并同样应用 Dropout 和层归一化, 从而生成第 l 层的最终输出表示 $\mathbf{H}^{(l)}$:

$$\mathbf{H}^{(l)} = \text{LayerNorm}(\mathbf{H}_{\text{attn}} + \text{Dropout}(\text{FFN}(\mathbf{H}_{\text{attn}}))) \quad (9)$$

FFN 的具体计算过程如下:

$$\text{FFN}(\mathbf{H}_{\text{attn}}) = \text{ReLU}(\mathbf{H}_{\text{attn}}\mathbf{W}_1^{\text{FFN}} + \mathbf{b}_1^{\text{FFN}})\mathbf{W}_2^{\text{FFN}} + \mathbf{b}_2^{\text{FFN}} \quad (10)$$

其中, \mathbf{W}^{FFN} 和 \mathbf{b}^{FFN} 是 FFN 中线性变换层的权重矩阵和偏置向量.

为处理不等长的输入序列, 模型在自注意力计算中采用填充掩码机制. 经过 L 层这样的堆叠处理后, Transformer 编码器的最终输出为 $\mathbf{H}_l \in \mathbb{R}^{N \times T \times D'}$.

3) 稀疏时间步的聚合与表征

为凝练各企业节点的完整历史财务序列动态,并生成一个统一的、信息浓缩的节点表征,在 Transformer 编码器输出的序列特征 \mathbf{H}_t 基础上,引入时序注意力层. 该模块的核心目标在于对不同时间步的特征赋予动态权重,从而识别并凸显对节点整体时序状态最具判别力的时间片段. 旨在克服财务数据低频特性所引发的长期依赖建模困难与短期波动过拟合等挑战,通过聚焦于关键时序点,提炼出表征全局动态演化趋势的特征向量.

具体而言,假设 Transformer 编码器的输出为 $\mathbf{H}_t \in \mathbb{R}^{N \times T \times D'}$,其中 N 是节点数量, T 是时间步数量, D' 是编码后的特征维度. 对于第 i 个节点在时间步 t 的表示记为 $\mathbf{h}_{i,t} \in \mathbb{R}^{D'}$. 时序注意力机制首先计算各时间步的注意力得分 $e_{i,t}$,并进而通过 Softmax 函数归一化得到注意力权重 $\beta_{i,t}$:

$$\mathbf{e}_{i,t} = \mathbf{v}^T \tanh(\mathbf{W}_{\text{att}} \mathbf{h}_{i,t} + \mathbf{b}_{\text{att}}) \quad (11)$$

$$\beta_{i,t} = \frac{\exp(e_{i,t})}{\sum_{k=1}^T \exp(e_{i,k})} \quad (12)$$

其中 $\mathbf{W}_{\text{att}} \in \mathbb{R}^{D_a \times D'}$, $\mathbf{b}_{\text{att}} \in \mathbb{R}^{D_a}$ 和 $\mathbf{v} \in \mathbb{R}^{D_a}$ 是可学习的参数, D_a 为注意力机制的隐藏层维度. $\beta_{i,t}$ 是第 i 个节点在时间步 t 的注意力权重.

随后,通过这些注意力权重对节点 i 在所有时间步的表示进行加权求和,从而生成该节点的聚合时序特征表示 $\bar{\mathbf{x}}_i$:

$$\bar{\mathbf{x}}_i = \sum_{t=1}^T \beta_{i,t} \mathbf{h}_{i,t} \quad (13)$$

其中 $\bar{\mathbf{x}}_i \in \mathbb{R}^{D'}$ 是节点 i 聚合了其整个时序信息的固定维度表示. $\bar{\mathbf{X}} \in \mathbb{R}^{N \times D'}$ 为低频时序动态特征学习模块的最终输出,它为每个节点提供了一个固定维度的向量表示,该表示不仅内蕴了节点特征的时序演化全貌,更通过注意力机制的赋权,侧重于对该节点整体时序状态最具信息量的关键时间节点,从而有效地将整个序列信息浓缩为一个能表征全局动态的特征向量.

3.1.3 动静表征联合输出层

静态指标行业感知增强模块强化了关键财务指标的特征,而低频时序动态特征学习模块则捕捉了财务数据序列中的长短期依赖和动态模式. 为了让产业链异构图信息聚合模块能够基于一个全面且信息丰富的初始表示进行学习,首先需要将静态指标行业感知增强模块和低频时序动态特征学习模块的输出进行融合. 具体地,对于企业 i ,将其静态指标行业感知增强模块处理后的最后一个时间步(即风险标签所对应的时间步) $\hat{\mathbf{x}}_{\text{last},i}$ 与低频时序动态特征学习模块处理后的时序全局表示 $\bar{\mathbf{x}}_i$ 进行拼接,并通过线性变换投影到统一的特征空间:

$$\mathbf{h}_i^f = \text{ReLU}(\mathbf{W}_f [\hat{\mathbf{x}}_{\text{last},i} \parallel \bar{\mathbf{x}}_i] + \mathbf{b}_f) \quad (14)$$

其中 \mathbf{W}_f 和 \mathbf{b}_f 是可学习的参数, \parallel 表示拼接操作, \mathbf{h}_i^f 是融合后的节点表示.

3.2 产业链异构图信息聚合模块

在复杂的产业链异构图中,节点间的关系不仅仅是单一的直接连接,而是蕴含着丰富的多类型语义信息. 例如,企业之间可能存在上游供应、下游客户、市场竞争等多种复杂联系. 产业链异构图信息聚合模块的核心目标是有效学习节点在这些复杂结构关系中的表示. 为此,使用元路径以显式定义

并捕获不同类型的语义路径,并使用层级化注意力机制学习

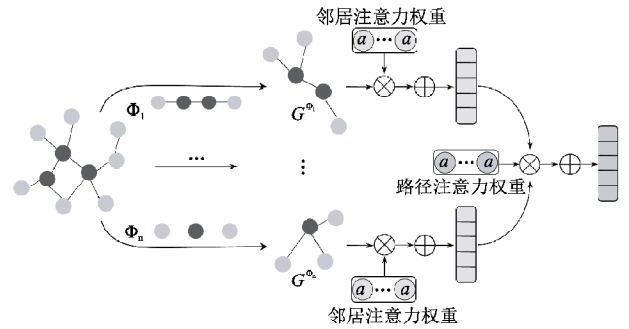


图4 产业链异构图信息聚合模块

Fig. 4 Graph info aggregation module

具备结构感知能力和丰富语义内涵的节点嵌入. 模块架构如图4所示.

3.2.1 关系路径引导的邻域聚合层

产业链网络中的企业间依赖关系是多样的,并不仅仅由直接连接定义. 元路径能够明确地定义和提取网络中连接两个实体的一系列特定类型的关系路径,从而帮助捕获超越直接邻居的、具有丰富语义的企业间高阶依赖关系. 给定产业链异构图 $G = (V, E)$ 和一组预定义的元路径 $\Phi = \{\phi_1, \phi_2, \dots, \phi_p\}$, 首先为每个元路径 ϕ_p 构建对应的元路径可达图 G^{ϕ_p} .

为了在每一种由元路径定义的特定语义关系下学习节点的表示,并且考虑到在特定语义路径下不同邻居节点对中心节点的影响程度可能不同,对每个元路径 $\phi_p \in \Phi$, 利用图注意力网络 (Graph Attention Network, GAT) 在其对应的可达图 G^{ϕ_p} 上聚合邻域信息,从而学习特定语义下的节点表示. GAT 的核心思想是为邻居节点分配不同的注意力权重,使得模型能够关注到对中心节点更重要的邻居. 具体来说,对于中心节点 i 和其在元路径 ϕ_p 下的邻居节点 $j \in \mathcal{N}_i^{\phi_p}$, 其注意力系数 $e_{ij}^{\phi_p}$ 计算如下:

$$e_{ij}^{\phi_p} = \text{LeakyReLU}(\mathbf{a}_{\phi_p}^T [\mathbf{W}_{\phi_p} \mathbf{h}_i^f \parallel \mathbf{W}_{\phi_p} \mathbf{h}_j^f]) \quad (15)$$

其中 \mathbf{W}_{ϕ_p} 是特定于元路径 ϕ_p 的可学习线性变换权重矩阵, \mathbf{a}_{ϕ_p} 是一个可学习的权重向量, \parallel 表示拼接操作. \mathbf{h}_i^f 和 \mathbf{h}_j^f 是节点 i 和 j 的输入特征.

为了使注意力系数在不同邻居间具有可比性,使用 softmax 函数进行归一化:

$$\alpha_{ij}^{\phi_p} = \text{softmax}_j(e_{ij}^{\phi_p}) = \frac{\exp(e_{ij}^{\phi_p})}{\sum_{k \in \mathcal{N}_i^{\phi_p}} \exp(e_{ik}^{\phi_p})} \quad (16)$$

节点 i 在元路径 ϕ_p 下聚合邻域信息后的表示 $\mathbf{z}_i^{\phi_p}$ 通过对邻居节点特征进行加权求和得到:

$$\mathbf{z}_i^{\phi_p} = \text{ELU}(\sum_{j \in \mathcal{N}_i^{\phi_p}} \alpha_{ij}^{\phi_p} \mathbf{W}_{\phi_p} \mathbf{h}_j^f) \quad (17)$$

此外,本文引入了多头注意力机制以增强模型的表达能力并稳定学习过程. 不同注意力头的输出被拼接起来,共同构成了最终的元路径特定节点表征 $\mathbf{z}_i^{\phi_p} \in \mathbb{R}^{D_{\text{gat}}}$, 其中 D_{gat} = 头数 \times GAT 输出维度. 完成此步骤后,对于每个节点 i ,获得了一组元路径特定的嵌入向量 $\{\mathbf{z}_i^{\phi_1}, \mathbf{z}_i^{\phi_2}, \dots, \mathbf{z}_i^{\phi_p}\}$.

3.2.2 多关系路径表示融合层

通过局部邻域聚合,获得了节点在不同元路径所定义的多重关系路径下的特定表示. 然而,这些不同的关系路径对于

理解企业的整体风险状况和其在产业链中的综合角色,其重要性各不相同,因此,本文引入了元路径间的注意力机制.全局关系融合层旨在学习不同元路径类型的全局重要性权重,并据此对各个元路径特定的节点表示进行加权聚合以形成一个统一且全面的节点表示,从而区分不同类型关系路径的贡献度,综合多角度的语义信息全面建模企业间的复杂邻域关系.

具体而言,对于节点 i 的一组元路径特定表示 $\{z_i^{\phi_1}, z_i^{\phi_2}, \dots, z_i^{\phi_p}\}$, 首先将它们堆叠起来形成一个张量 $\mathbf{Z}_i \in \mathbb{R}^{p \times D_{\text{gat}}}$. 然后,计算每条元路径 ϕ_p 的重要性得分 w_p . 这些得分在所有节点上计算并取均值,以学习元路径类型的全局重要性:

$$w_p = \frac{1}{N} \sum_{i=1}^N (\mathbf{q}^T \tanh(\mathbf{W} \mathbf{z}_i^{\phi_p} + \mathbf{b})) \quad (18)$$

其中, \mathbf{W} 和 \mathbf{b} 分别是 MLP 隐藏层的权重矩阵和偏置向量, \mathbf{q}^T 是输出层的权重向量, \tanh 为双曲正切激活函数. 这些参数在所有元路径和所有节点间共享. 这些原始得分通过 softmax 函数进行归一化,从而得到元路径 ϕ_p 的注意力权重 β_{ϕ_p} :

$$\beta_{\phi_p} = \frac{\exp(w_p)}{\sum_{k=1}^p \exp(w_k)} \quad (19)$$

这些权重 β_{ϕ_p} 反映了不同语义路径对于节点最终表示的相对贡献. 最后,节点 i 的最终结构感知语义表示 $\tilde{\mathbf{X}}_i$ 通过对所有元路径特定表示进行加权求和得到:

$$\tilde{\mathbf{X}}_i = \sum_{p=1}^p \beta_{\phi_p} \mathbf{z}_i^{\phi_p} \quad (20)$$

最终,得到产业链异构图信息聚合模块的输出 $\tilde{\mathbf{X}}$. 基于层次注意力机制的产业链异构图信息聚合模块可以堆叠多层,每一层的输出作为下一层的输入,旨在逐层抽象和学习图中更深层次、更复杂的结构特征与依赖关系.

3.3 风险判别与类别不平衡感知学习

MGR1 采用一个全连接线性层作为最终的预测层:

$$\hat{y} = \sigma(\tilde{\mathbf{X}} \mathbf{W}_{\text{pred}} + \mathbf{b}_{\text{pred}}) \quad (21)$$

其中 \mathbf{W}_{pred} 是预测层的权重矩阵, \mathbf{b}_{pred} 是偏置向量, σ 是 sigmoid 激活函数, \hat{y} 表示节点为风险企业的预测概率.

模型的训练目标是 minimize 预测结果与真实风险标签之间的差异. 考虑到在真实的金融市场中风险企业占比较少,导致数据存在显著的类别不平衡问题,采用焦点损失 (Focal Loss) 作为损失函数. 焦点损失通过引入调制因子来降低易分类样本的权重,促使模型更专注于学习那些难以区分的少数类 (风险) 样本,从而提升模型在不平衡数据集上的整体风险预测性能和泛化能力. 针对二分类问题,焦点损失的计算如下:

$$FL(\hat{y}) = \begin{cases} -\alpha(1-\hat{y})^\gamma \log(\hat{y}) & \text{if } y=1 \\ -\alpha\hat{y}^\gamma \log(1-\hat{y}) & \text{if } y=0 \end{cases} \quad (22)$$

其中, y 是企业的真实风险标签; $\gamma \geq 0$ 是聚焦参数,用于调节易分类样本权重降低的速率; α 是类别权重因子,用于平衡正负样本的重要性. 在 MGR1 中, α 值会依据训练集中各类别样本的比例进行设定,以进一步缓解类别不平衡问题.

4 实验

4.1 数据集

为了验证方法的有效性,本文在两个真实的数据集 ICD1 和 ICD2 上进行实验,均为国内某大型制造集团提供的真实产业链数据. 针对产业链的企业风险识别任务具体而言,数据

集包含产业链异构图、节点属性和风险标签 3 个部分. 数据集统计信息详见表 1.

表 1 数据集统计信息

Table 1 Dataset statistics

名称	类型	ICD1	ICD2	解释
节点	企业	1972	1501	企业节点
	行业	430	294	行业节点
	上下游	1884	1100	行业和行业上下游关系
边	投资	699	113	企业和企业投资关系
	包含	3003	2936	行业和企业所属关系
标签	0	14931	9257	无风险
	1	1645	1595	有风险

ICD1 代表了受宏观政策影响显著的资本密集型产业. 该数据集包含行业节点 430 个,企业节点 1972 个. 数据集 ICD2 则代表了由市场竞争主导的消费驱动型产业,包含行业节点 294 个,企业节点 1501 个. 通过在这两个特性不同的数据集上进行验证,可以更全面地检验模型的泛化能力.

两个数据集中的企业属性原始财务数据均来源于 CSMAR 数据库. 经过筛选匹配与数据预处理,最终获得了包含 21 个关键指标的财务面板数据,其中时间跨度为 2000 年至 2023 年,时间间隔为 1 年. 财务指标信息详见表 2.

表 2 财务指标定义表

Table 2 Definitions of financial indicators

资产结构	总资产对数	总资产的自然对数
资产结构	有形资产比率	(资产总计 - 无形资产净额 - 商誉净额)/资产总计
	流动资产比率	流动资产/总资产
	非债务税盾	(固定资产折旧 + 油气资产折耗 + 生产性生物资产折旧)/总资产
账面市值比	股东权益/公司市值	
偿债能力	经营性净现金流比率	经营性净现金流/总负债
	流动负债比率	流动负债/总负债
	金融负债比率	(短期借款 + 一年内到期的非流动负债 + 长期借款 + 应付债券)/总资产
	营运资金比率	(流动资产 * 流动负债)/总资产
	资产负债率	总负债/总资产
	流动比率	流动资产/流动负债
营运能力	短期借款依赖度	(短期借款 + 一年到期的长期借款)/总资产
	现金流波动性	(现金流/总资产)三年标准差
	营业收入比率	营业收入/总资产
	现金及现金等价物周转率	营业收入/期末现金及现金等价物余额
	存货周转率	营业成本/存货净额期末余额
盈利能力	总资产净利润率	净利润/资产总计
	留存收益比率	(盈余公积 + 未分配利润)/总资产
	盈利波动性	(息税前利润/总资产)三年标准差
成长能力	总资产增长率	(本期末总资产 - 上期末总资产)/上期末总资产
	营业收入增长率	(本季营业收入 - 上季营业收入)/上季营业收入

企业风险标签的生成首先以 CSMAR 数据库中多种财务困境模型的量化评分为基础,并参照公认的行业风险标准与公开披露的重大事实信息进行综合判定,最终生成二分类标签.

4.2 基线方法

为了验证本文所提方法 MGRI 的有效性,本文将与多种类型最先进的基线方法进行比较,包括风险评估领域的经典机器学习模型和前沿的神经网络模型,具体如下:

PCA-LR:一种两阶段评价模型,首先通过主成分分析(PCA)对大量相关指标进行降维以提取关键特征,然后利用这些特征作为逻辑回归(LR)模型的输入,LR通过对输入特征进行线性加权并使用 Sigmoid 函数输出概率。

XGBoost^[26]:一种基于梯度提升树的集成学习方法,能够高效处理结构化数据。该方法通过集成多个决策树,逐步优化损失函数,并利用特征拼接与欠采样策略提升模型对类别不平衡问题的鲁棒性。

MLP:MLP 是一种基础的前馈神经网络,由多个全连接层组成,通过非线性激活函数进行特征转换和模式学习。

GCN:GCN 是一种直接在图结构数据上进行半监督学习的卷积神经网络。它通过定义图上的卷积操作来聚合邻居节点的特征信息,从而学习节点的有效表示。每一层 GCN 通过对邻域节点特征进行加权平均,并结合非线性激活函数,实现节点信息的传播和转换。

GAT:GAT 是一种在图结构数据上进行节点表示学习的图神经网络。其核心机制在于引入了自注意力机制,使得节点在聚合邻居信息时可以为不同的邻居分配不同的学习权重,从而更关注重要的邻居节点。

HAN:HAN 是一种基于层级注意力机制的半监督图神经网络。首先,它采用节点级注意力学习一个节点与其基于元路径的邻居之间的重要性。然后,使用语义级注意力学习不同元路径的重要性,并据此聚合通过不同元路径学习到的节点嵌入,形成最终的节点表示。

Simple-HGN^[27]:Simple-HGN 是一种以 GAT 为骨干构建的简单而有效的异构图神经网络模型。它通过整合可学习的边类型嵌入、残差连接和输出嵌入的 L2 归一化 3 种关键技术来增强 GAT 在异构图上的建模能力。

4.3 实验设置

MGRI 模型及所有基线方法均基于 Python 3.9、PyTorch 2.5.1、DGL 1.1.2 和 CUDA 11.8 实现。实验硬件环境为配备 Intel Core i5-13600KF(3.50 GHz)、32GB 内存、NVIDIA GeForce RTX 4070 Ti 显卡(12GB 显存)以及操作系统为 Windows 11(版本 23H2)的计算机上进行。

训练过程中采用早停策略,耐心值设置为 30,最大训练轮次设置为 300。将数据集按照时间进行划分,使用 2022 年的数据进行训练和验证,并使用 2023 年的数据进行测试。在训练集和验证集的划分上,按照 7:3 的比例进行分层随机抽样,以确保类别分布的平衡。现实产业链数据在风险识别任务中天然存在正负样本不均衡的问题,为了全面且准确地评估 MGRI 模型的性能,本文采用召回率(Recall)、F1 分数(F1-Score)和 AUC 作为评价指标。

4.4 结果分析

本文在 ICD1 和 ICD2 上对 MGRI 进行了严格评估,并与多种代表性的基线方法进行了比较。对于所有方法,使用不同的随机种子重复实验 7 次,按 F1 分数取剔除最大值和最小值后的平均得分作为最终结果。表 3 详细列出了各模型在 F1 分

数(F1-Score)、召回率(Recall)和 AUC 3 个关键指标上的性能表现。

表 3 对比实验结果

Table 3 Results of comparative experiment

数据集 评价指标	ICD1			ICD2		
	Recall	F1-Score	AUC	Recall	F1-Score	AUC
PCA-LR	0.8261	0.5714	0.8233	0.7543	0.6281	0.8284
XGBoost	0.8009	0.6275	0.8407	0.7326	0.6598	0.8310
MLP	0.7011	0.4902	0.7773	0.6675	0.6117	0.7742
GCN	0.8246	0.5798	0.8068	0.7396	0.6530	0.8447
GAT	0.8276	0.5783	0.8060	0.7082	0.6343	0.8542
HAN	0.8963	0.6536	0.9282	0.7607	0.6656	0.8778
Simple-HGN	0.9090	0.6694	0.9300	0.7886	0.7165	0.8757
MGRI	0.9366	0.7484	0.9527	0.8952	0.7180	0.9458

从表 3 数据可以看出, MGRI 在两个特性不同的数据集上,其 F1 分数、召回率和 AUC 3 个评价指标上均显著优于所有对比的基线方法,这充分证明了 MGRI 在产业链风险识别任务中的有效性和优越性。具体而言:

传统的机器学习和统计方法 PCA-LR、XGBoost 以及基础的神经网络模型 MLP,由于将每个企业视为独立的实体进行分析,其各项指标普遍低于图神经网络方法。这表明,忽略企业间的关联和产业链的结构信息,难以有效捕捉复杂的系统性风险模式。值得注意的是,XGBoost 作为强大的集成学习模型,在某些指标上表现尚可,但仍不及能够建模异构关系的图神经网络。

同构图神经网络方法 GCN 和 GAT 相较于非图方法,在各项指标上均有明显提升。例如,在 ICD1 数据集上,GCN 相较于 MLP 在召回率、F1 分数和 AUC 上分别提升了 12.35%、8.81% 和 2.87%。召回率的显著提高意味着模型能够识别出更多实际存在风险的企业。这证明了利用图结构聚合邻域信息对于捕捉风险传导模式的有效性。然而,异构图神经网络方法 HAN 和 Simple-HGN 的表现则更为出色,相较于同构图方法,在 ICD1 上 AUC 提升超过 12.14%,在 ICD2 上也有一致的优势。这证明了在产业链场景下,区分不同类型的关系并赋予不同重要性,对于精准建模企业间的复杂联系至关重要。

MGRI 的各项指标全面超越了表现最好的基线模型 Simple-HGN,其优越性在不同类型的产业链数据中均得到验证。在 ICD1 上, MGRI 的召回率、F1 分数和 AUC 相较于 Simple-HGN 分别提升了 2.76%、7.90% 和 2.27%。其中,F1 分数的提升尤为显著,这表明在该数据集代表的风险模式下, MGRI 能更精准地权衡错报与漏报的代价,实现对风险企业更均衡、可靠的判定。而在 ICD2 上, MGRI 同样表现出最佳性能,其召回率和 AUC 分别取得了高达 10.66% 和 7.01% 的显著提升。召回率和 AUC 的大幅领先,证明了在该场景中 MGRI 能更敏锐地捕捉到那些被传统模型忽略的、更隐蔽的风险信号,显著提升了风险识别的全面性和覆盖广度。这种在不同特性数据集上性能增益的差异化表现,凸显了 MGRI 的强大适应能力。充分证明了 MGRI 能够自适应地应对不同产业的风险特性,对节点自身属性进行更深层次的挖掘与增强,从而实现了更为全面且卓越的风险判别能力。

4.5 消融实验

为进一步验证 MGRI 中各个核心模块的有效性及其对整

体性能贡献,本文设计 3 种变体进行消融实验,具体如下:

- $MGRI_{VF}$:从 $MGRI$ 中移除静态指标行业感知增强模块.
- $MGRI_{VT}$:从 $MGRI$ 中移除低频时序动态特征学习模块.
- $MGRI_{IH}$:将 $MGRI$ 中产业链异构图信息聚合模块的异构图神经网络替换为同构图神经网络.

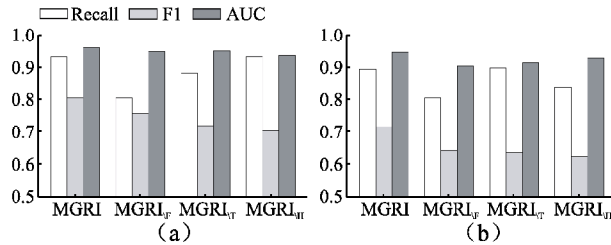


图 5 消融实验结果
Fig. 5 Results of ablation study

实验结果如图 5 所示,图 5(a)和图 5(b)分别对应 ICD1 和 ICD2 数据集.在两个数据集上,所有变体在召回率、F1 分数和 AUC 指标上均表现出不同程度的性能下降:1) $MGRI_{VF}$ 的召回率降幅最为显著,F1 分数也大幅下滑,这说明行业感知的特征强化对全面识别潜在风险企业及提升模型判别能力的重要性;2) $MGRI_{VT}$ 的召回率和 F1 分数明显降低凸显了对离散稀疏的财务数据时间动态性建模、捕捉其长期依赖与演变模式的必要性;3) $MGRI_{IH}$ 的 F1 分数出现了最大幅度的下降,这充分表明了有效捕获和区分产业链异构网络中复杂企业依赖关系对于模型精准评估风险的关键作用.

4.6 参数敏感性分析

为了进一步探究 $MGRI$ 模型对关键超参数变化的敏感程度,并验证实验设置中所选参数的合理性,本文针对各模块进行了一系列的参数敏感性分析,实验结果如图 6 所示,图 6(a)和图 6(b)分别对应 ICD1 和 ICD2 数据集.

实验结果表明,模型性能并非随参数值的单向增长而线性提升,而是常常在特定范围内达到峰值后出现下降趋势.在低频时序动态特征学习模块中,通过对两个数据集的综合分析,Transformer 的层数设置为 2、注意力头数为 16、FFN 维度为 64 时,模型展现出整体最高的 F1 值.这表明过少(的参数)则不足以捕捉财务序列中的复杂动态依赖,过多则可能导致模型对训练数据中的噪声过于敏感,反而降低了泛化能力.同时,低频时序动态特征学习模块处理的时间序列长度在 9 个时间步、输出维度为 128 时,性能达到最佳.这意味着在产业链风险预测的场景下,适度的历史窗口足以捕捉关键的风险演化趋势,并有效浓缩时序信息,为后续的图神经网络模块提供精炼且信息量充足的输入.

对于产业链异构图信息聚合模块的 GNN,其最优超参数在不同类型的数据集上呈现出显著差异.在易受宏观政策影响较大的资本密集型产业 ICD1 中,当隐藏层维度设置为 128、注意力头数为 8 时,模型性能最佳,较大的隐藏层维度有助于学习和表征其结构中复杂的深层语义依赖;而在由市场竞争主导的消费驱动型产业 ICD2 中,当隐藏层维度降低至 64、注意力头数增加至 16 时,效果最优,更多的注意力头数则赋予模型从多个子空间捕捉异构邻域信息的能力,以应对更动态和多样化的风险传导路径,同时较小的隐藏层维度有助

于提升模型的泛化能力并防止在复杂信号中过拟合.

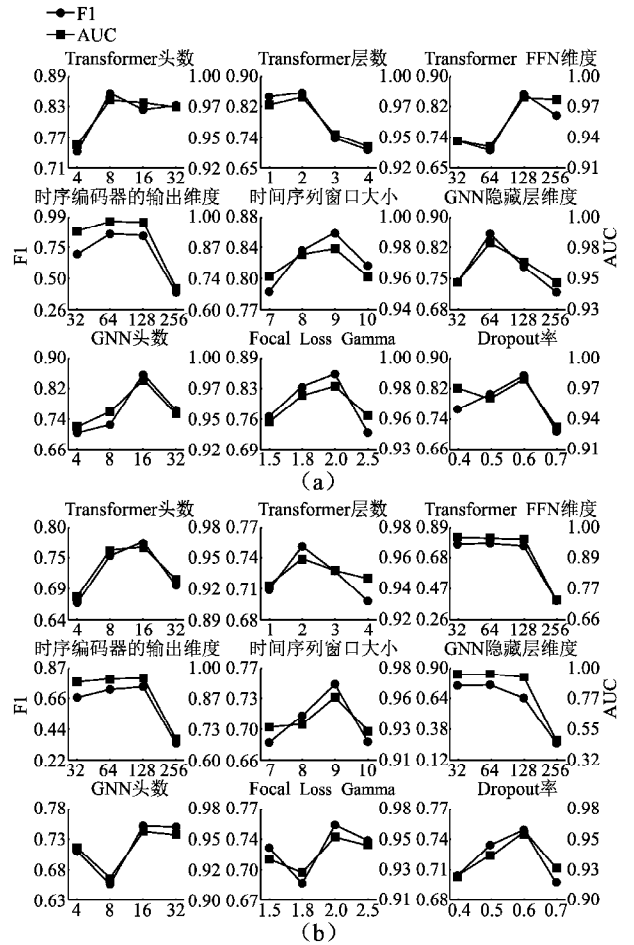


图 6 参数敏感性分析结果
Fig. 6 Results of parameter sensitivity analysis

最后,在两个数据集上,Dropout 率均为 0.6 以及 Focal Loss 的 gamma 值在 2.0 附近时模型性能达到最佳,这凸显了在处理真实世界中普遍存在类别不平衡的金融风险数据时,强化模型正则化、并迫使模型关注少数难分类风险样本的共同重要性.

综上所述, $MGRI$ 对这些超参数的调整表现出一定的敏感性,且最优参数会随数据集特性的不同而变化.这既验证了本文所选参数的合理性,也强调了针对具体应用场景进行细致调优以激发模型最大潜能的必要性.

5 结论

本文提出了一种多模块注意力协同的异构图神经网络产业链风险识别模型($MGRI$),有效应对了现有方法在处理低频、少量财务时间序列数据时长期依赖建模能力不足的问题,并解决了财务指标在不同语境下重要性差异及特征间复杂结构关系被忽视等挑战.该模型由静态指标行业感知增强模块、低频时序动态特征学习模块及产业链异构图信息聚合模块组成,分别从行业敏感性、时间动态性和关系结构性 3 个维度建模企业风险特征,并通过动静双维度编码器与元路径引导的异构图神经网络实现信息融合.在真实的产业链数据集上的

实验结果表明, MGRI 的所有指标均优于表现最好的基线方法, 更重要的是, 通过在不同特性产业链场景上的验证, 揭示了模型对不同风险模式的强大适应能力与泛化性能。

尽管如此, MGRI 仍有进一步拓展的空间。目前模型主要基于结构化的财务指标和产业链图谱构建特征, 未来可考虑引入更多维度的非结构化外部信息, 例如企业相关的宏观经济指标、行业政策、新闻舆情、专利技术等, 以期从更广阔的视角捕捉可能影响企业风险的因素, 进一步提升风险识别的全面性和前瞻性。

References:

- [1] GUO Z M, QIU Y. Reconstruction of the global value chain in the digital economy era: typical facts, theoretical mechanisms and Chinese strategies [J]. *Reform*, 2020, (10): 73-85, doi: 10. 1109/ECIT52743. 2021. 00046.
- [2] CHEN X D, YANG X X. Does digital transformation improve the autonomous and controllable capabilities of the industrial chain? [J]. *Business Management Journal*, 2022, 44(8): 23-39.
- [3] Wen C, Yang J, Gan L, et al. Big data driven internet of things for credit evaluation and early warning in finance [J]. *Future Generation Computer Systems*, 2021, 124: 295-307, doi: 10. 1016/j. future. 2021. 06. 003.
- [4] Hewage C, Yasakethu L, Jayakody D N K. Data protection: the wake of ai and machine learning [M]. Cham: Springer Nature Switzerland, 2024.
- [5] Bi W, Xu B, Sun X, et al. Company-as-tribe: company financial risk assessment on tribe-style graph with hierarchical graph neural networks [C] // *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2022: 2712-2720.
- [6] Song L, Li H, Tan Y, et al. Enhancing enterprise credit risk assessment with cascaded multi-level graph representation learning [J]. *Neural Networks*, 2024, 169: 475-484.
- [7] Zhou Z, Bi K, Zhong Y, et al. HKTGNN: hierarchical knowledge transferable graph neural network-based supply chain risk assessment [C] // *Parallel Distributed Processing with Applications, Big Data Cloud Computing, Sustainable Computing Communications, Social Computing Networking*, 2023: 772-782.
- [8] Mahmoud A, Mohammed A. A survey on deep learning for time-series forecasting [M]. *Machine Learning and Big Data Analytics Paradigms: Analysis, Applications and Challenges*, 2021: 365-392.
- [9] DONG S Y. The research of the sparse time series prediction based on empirical mode decomposition and attention mechanism [D]. Changchun: Jilin University, 2024.
- [10] Zhou J, Cui G, Hu S, et al. Graph neural networks: a review of methods and applications [J]. *AI Open*, 2020, 1: 57-81, doi: 10. 1016/j. aiopen. 2021. 01. 001.
- [11] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks [J]. *arXiv*: 1609. 02907, 2017, doi: 10. 48550/arXiv. 1609. 02907.
- [12] Veličković P, Cucurull G, Casanova A, et al. Graph attention networks [C] // *International Conference on Learning Representations*, 2018, doi: 10. 48550/arXiv. 1710. 10903.
- [13] Bing R, Yuan G, Zhu M, et al. Heterogeneous graph neural networks analysis: a survey of techniques, evaluations and applications [J]. *Artificial Intelligence Review*, 2023, 56(8): 8003-8042.
- [14] Zhang C, Song D, Huang C, et al. Heterogeneous graph neural network [C] // *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019: 793-803.
- [15] Wang X, Ji H, Shi C, et al. Heterogeneous graph attention network [C] // *The World Wide Web Conference*, 2019: 2022-2032.
- [16] Hu Z, Dong Y, Wang K, et al. Heterogeneous graph transformer [C] // *Proceedings of The Web Conference*, 2020: 2704-2710.
- [17] Altman E I. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy [J]. *The Journal of Finance*, 1968, 23(4): 589-609.
- [18] Ohlson J A. Financial ratios and the probabilistic prediction of bankruptcy [J]. *Journal of Accounting Research*, 1980, 18(1): 109-131.
- [19] Lee Y C. Application of support vector machines to corporate credit rating prediction [J]. *Expert Systems with Applications*, 2007, 33(1): 67-74.
- [20] Hua Z, Wang Y, Xu X, et al. Predicting corporate financial distress based on integration of support vector machine and logistic regression [J]. *Expert Systems with Applications*, 2007, 33(2): 434-440.
- [21] Xu K, Hu W, Leskovec J, et al. How powerful are graph neural networks? [C] // *International Conference on Learning Representations*, 2019, doi: 10. 48550/arXiv. 1810. 00826.
- [22] Wu Z, Pan S, Chen F, et al. A comprehensive survey on graph neural networks [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 32(1): 4-24.
- [23] Zhang Z, Ji Y, Shen J, et al. Collaborative metapath enhanced corporate default risk assessment on heterogeneous graph [C] // *Proceedings of the ACM Web Conference, Association for Computing Machinery*, 2024: 446-456.
- [24] Wei S, Lv J, Guo Y, et al. Combining intra-risk and contagion risk for enterprise bankruptcy prediction using graph neural networks [J]. *Information Sciences*, 2024, 659: 120081, doi: 10. 1016/j. ins. 2023. 120081.
- [25] Bi Kejun, Sun Pengzhao, Wang Ruijin, et al. Industry chain risk assessment model combining graph fusion and attribute completion [J/OL]. *Journal of Chinese Computer Systems*, 2025: 1-11, <http://kns.cnki.net/kcms/detail/21. 1106. tp. 20240731. 1001. 002. html>.
- [26] Chang Y C, Chang K H, Wu G J. Application of eXtreme gradient boosting trees in the construction of credit risk assessment models for financial institutions [J]. *Applied Soft Computing*, 2018, 73: 914-920, doi: 10. 1016/j. asoc. 2018. 09. 029.
- [27] Lv Q, Ding M, Liu Q, et al. Are we really making much progress? Revisiting, benchmarking and refining heterogeneous graph neural networks [C] // *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 2021: 1150-1160.

附中文参考文献:

- [1] 郭周明, 裴莹. 数字经济时代全球价值链的重构: 典型事实、理论机制与中国策略 [J]. *改革*, 2020, (10): 73-85.
- [2] 陈晓东, 杨晓霞. 数字化转型是否提升了产业链自主可控能力? [J]. *经济管理*, 2022, 44(8): 23-39.
- [9] 董思逾. 基于经验模态分解和注意力机制的稀疏时间序列预测研究 [D]. 长春: 吉林大学, 2024.
- [25] 毕可骏, 孙鹏钊, 王瑞锦, 等. 结合图融合和属性补全的产业链风险评估模型 [J/OL]. *小型微型计算机系统*, 2025: 1-11, <http://kns.cnki.net/kcms/detail/21. 1106. tp. 20240731. 1001. 002. html>.