面向供应链金融的大语言模型行业语义增强企业信用风险评估方法

刘颖[1,2,3,4]，陈嘉茂[1]，孙楠[1]
1. 吉林财经大学 人工智能与计算机学院，长春 130117；
2. 吉林省商务大数据研究中心，长春 130117；
3. 大数据与交叉科学研究室，长春 130117；
4. 吉林省金融科技重点实验室，长春 130117

摘要

供应链金融中的企业信用风险评估受到行业异质性、财务指标语义差异等因素的共同影响，传统基于结构化财务数据的风险预测方法难以充分刻画财务指标背后的行业经营逻辑与跨期风险变化过程。针对上述问题，提出一种面向供应链金融的大语言模型行业语义增强企业信用风险评估方法ISemTAR。首先，将企业所属行业、核心财务指标及其年度变化特征构建为提示词，利用 GLM-4 生成行业感知的财务风险语义文本，将原始数值型财务表格转化为具有行业解释能力的语义信息。随后，采用预训练语言模型FinBERT2提取文本语义特征，并设计双阶段Transformer对企业财务面板数据进行编码。第一阶段学习同一年度内不同财务指标之间的交互关系，形成年度表示；第二阶段对各年度表示进行时序编码，刻画企业经营状态的跨年度变化。在此基础上，引入跨模态对齐学习机制，将文本语义表示与表格结构表示映射到统一语义空间，实现行业风险语义向结构化财务特征的迁移，从而增强模型对企业信用风险的识别能力。在供应链金融企业数据集上进行了大量的实验，实验结果表明 ISemTAR 在风险评估中显著优于基线模型。

关键词

供应链金融；信用风险评估；行业语义增强；大语言模型；Transformer

中图分类号：TP391　　文献标志码：A

Enterprise credit risk assessment via large language model-based industry semantic enhancement for supply chain finance

LIU Ying[1,2,3,4], CHEN Jiamao[1], SUN Nan[1]
1. Jilin University of Finance and Economics, Changchun 130117, China；
2. Jilin Provincial Research Center for Business Big Data, Changchun 130117, China；
3. Institute of Big Data and Cross Science, Changchun 130117, China；

4. Key Laboratory of Financial Technology in Jilin Province, Changchun 130117, China

Abstract

Supply chain finance risk assessment is influenced by factors such as industry heterogeneity and semantic differences in financial indicators; consequently, traditional risk prediction methods based on structured financial data struggle to fully capture the underlying business logic of financial indicators and the intertemporal evolution of risk. To address these issues, we propose ISemTAR, a supply chain finance risk assessment method that leverages large language models to enhance industry-specific semantics. First, we construct prompts based on the enterprise's industry, core financial indicators, and their annual variation characteristics. Using GLM-4, we generate industry-aware financial risk semantic text, transforming raw numerical financial tables into semantic information with industry-specific interpretability. Subsequently, FinBERT2 is employed to extract textual semantic features, and a two-stage Transformer is designed to encode corporate financial panel data. The first stage captures interactions among financial indicators within each year to obtain annual representations. The second stage encodes these representations chronologically to capture changes in corporate operating conditions across years. Building on this foundation, a cross-modal alignment learning mechanism is introduced to map text semantic representations and tabular structural representations into a unified semantic space. This enables the transfer of industry risk semantics into structured financial features, thereby enhancing the model's ability to identify corporate credit risk. Experiments are conducted using financial panel data from listed firms with supply chain relationships, with ST/*ST status serving as a proxy label for corporate credit risk. The results show that ISemTAR achieves better overall performance than the compared baseline models in predicting next-year corporate risk status.

Key words

supply chain finance, credit risk assessment, industry-specific semantic enhancement, large language model, transformer

1 引言

供应链金融是连接金融机构、核心企业与上下游中小企业的重要融资机制。它以真实交易和供应链协同为基础，将企业个体信用拓展为可验证、可传递的链式信用，因此信用风险管理是其研究与实践的核心[1]。真实业务通常横跨多个行业，而不同行业企业的资产结构、盈利模式、成本刚性、库存周转及现金流形成机制存在显著差异，行业不仅影响总体风险水平，也会改变财务指标的风险含义。已有研究表明，将行业特定的违约异质性纳入模型能够改善多期违约预测效果[2]，说明跨行业风险评估需要结合具体经营逻辑解释财务指标。

近年来，大语言模型的发展为金融风险研究提供了新的切入点，BloombergGPT、FinGPT 和 PIXIU 分别从金融语料预训练、开放式数据治理与轻量微调、指令数据及评测基准等方面，验证了大模型在金融情感分析、问答和信息抽取等任务中的领域理解能力[3-5]。程大伟等[6]提出知识增强的继续预训练与监督微调方法，提升中文金融大模型在专业领域任务中的表现。王敏等[7]结合知识图谱和

大语言模型增强用户与商品特征，并将辅助信息表征与图神经网络获得的ID表征进行对齐，为语义信息与结构化表示的融合提供了方法参考。在企业信用风险评估中，还需要进一步结合行业背景，挖掘结构化财务指标所包含的风险语义。财务表格并非孤立数值的集合。例如，零售企业通常具有较快的存货周转速度和较高的现金交易比例，即使其流动比率与重资产制造企业相同，二者所对应的短期偿债风险也可能不同。由此，有必要将行业背景、财务指标及其变化转化为可学习的风险语义。

企业信用风险还具有明显的时间依赖和阶段演化特征。经营波动、回款迟滞、库存积压、融资成本变化及供应链风险传导会随时间逐步累积[8-10]。Wang表明多年度财务信息能够支持更长期的企业财务困境预测[11]。Pellegrino 等基于多年会计时间序列构建多头 LSTM 模型，发现按财务变量分别建模时间序列有助于提升企业破产预测效果[12]。然而，现有时序模型往往只是将多个年份的财务指标整体输入网络，难以同时刻画指标间结构关系和风险演化路径。

针对上述问题，本文提出大语言模型行业语义增强的供应链金融风险评估方法ISemTAR（Industry Semantic-enhanced Transformer with Alignment Learning for Risk Assessment）。首先，将企业所属行业、核心财务指标及其年度变化组织为提示词，由GLM-4生成行业感知的财务风险文本，并利用FinBERT2提取语义表示，使结构化财务数据获得可解释的行业风险语义。其次，按照“年度内 — 跨年度”的顺序对表格数据进行编码：先由指标Transformer聚合同一年度内各项指标的组合关系，形成年度表示；再将各年度表示按时间顺序输入年份Transformer，刻画企业经营状态及风险累积的跨期演化。最后，通过跨模态对齐使同一企业的文本语义表示与表格结构表示在统一空间中相互接近，将行业风险语义迁移至结构化财务特征并用于风险预测。

本文研究的主要创新点包括：

(1) 将供应链企业所属行业、财务指标等组合成大模型提示词，利用大语言模型生成行业感知的财务风险文本，使模型能够引入行业知识，提升对企业风险状态的语义理解能力。

(2) 设计双阶段 Transformer 编码结构，首先学习同一年度内财务指标之间的组合关系，随后对年度表示进行时序编码，联合建模财务指标交互与跨年度依赖，增强对企业经营状态变化的刻画能力。

(3) 本文通过对齐学习拉近同一企业的文本语义表示和表格结构表示，使大语言模型生成的行业风险语义能够迁移到结构化表格分支中，在预测阶段增强模型对供应链金融企业信用风险的识别能力。

2 相关工作

2.1 信用风险的行业异质性

企业信用风险的形成及传导具有明显的行业差异。不同行业在资产结构、盈利模式、现金流机制及外部冲击暴露程度等方面存在差异。Huang和Lee从产业网络视角发现，行业连接位置及尾部风险外溢会影响企业违约概率[13]；Spatareanu等进一步表明，银行风险可沿上下游及同行业联系传导至非金融企业[14]。因此，企业风险不仅取决于自身财务状况，也受到行业网络和供应链关系的影响。

跨行业预测主要通过行业类别和外部风险信号引入行业信息。Sigrist 和

Leuenberger将行业类别与财务、市场变量共同输入非线性模型[15]；Bitetto等利用同行业上市企业的市场信号改善非上市中小企业的违约预测[16]；农业企业网络研究则揭示了供应链、投资和信息网络中的差异化风险传导[17]。

总体而言，现有研究多将行业作为控制变量、分组标签或外部环境变量纳入模型，虽能缓解行业样本差异并改善预测，却较少挖掘财务指标本身包含的行业语义。

2.2 大语言模型在金融风险识别的应用

大语言模型（large language model，LLM）能够从金融信息中提取与违约、财务困境和偿债能力相关的语义信号，已被用于企业披露分析、信用违约预测及可解释信用建模[18]。

现有研究主要以既有金融文本为输入。Che等融合财务指标、临时公告和企业网络开展多模态预测[19]，Fan等利用生成式大语言模型从调研纪要中识别供应链风险[20]。在供应链金融场景中，DeepRRP融合最新披露文本与财务指标，并利用FinBERT和BiLSTM建模多期语义信息[21]；Wu等利用ChatGPT优化贷款评估文本并与结构化变量融合[22]。

上述研究表明，金融文本语义能够为风险预测提供增量信息，但其输入仍以公告、调研纪要、披露文件和贷款描述等既有文本为主。如何将财务表格等结构化数据转化为语言模型可利用的语义对象，仍缺乏充分研究。基于此，本文利用GLM-4对结构化财务数据进行行业语义增强。

2.3 时序企业信用风险

企业信用风险通常经历跨期累积和演化过程。Wang等指出，单期预测难以反映风险随时间的变化[23]；Sigrist和Leuenberger发现，随着预测期限延长，变量间交互效应更加明显[15]。因此，多年度信息是刻画信用风险变化的重要依据。

相关研究逐步从传统模型转向深度时序建模。Wang和Chi构建成本敏感Stacking模型，验证了财务与宏观变量对长期财务困境预测的作用[11]。在深度时序建模方面，Korangi等利用Transformer联合建模基本面、市场和定价数据中的跨期信息[24]；4C-FinNet将财务数据转化为多通道特征图并融合时空特征[25]；TemFRC通过时间折叠和风险对比学习捕捉趋势、周期及企业间风险相似性[26]。

现有模型虽然使用多年度数据，但通常将不同年份和指标整体输入网络，未充分区分同一年度内不同指标的组合关系与企业状态在年度之间的连续演化。本文据此采用双阶段Transformer，依次进行年度内指标编码与跨年度状态编码。

3 供应链金融信用风险评估模型

本文提出行业语义增强信用风险评估模型ISemTAR。模型首先利用大语言模型生成行业感知的财务风险文本，并通过FinBERT2提取文本语义表示；随后按照“年度内指标编码—跨年度状态编码”的顺序处理财务面板数据；最后通过跨模态对齐，将行业风险语义迁移至结构化财务表示，并完成信用风险预测。训练阶段同时使用文本分支与表格分支，推理阶段仅保留对齐增强后的表格分支。整体架构如图1所示。

3.1 文本特征提取

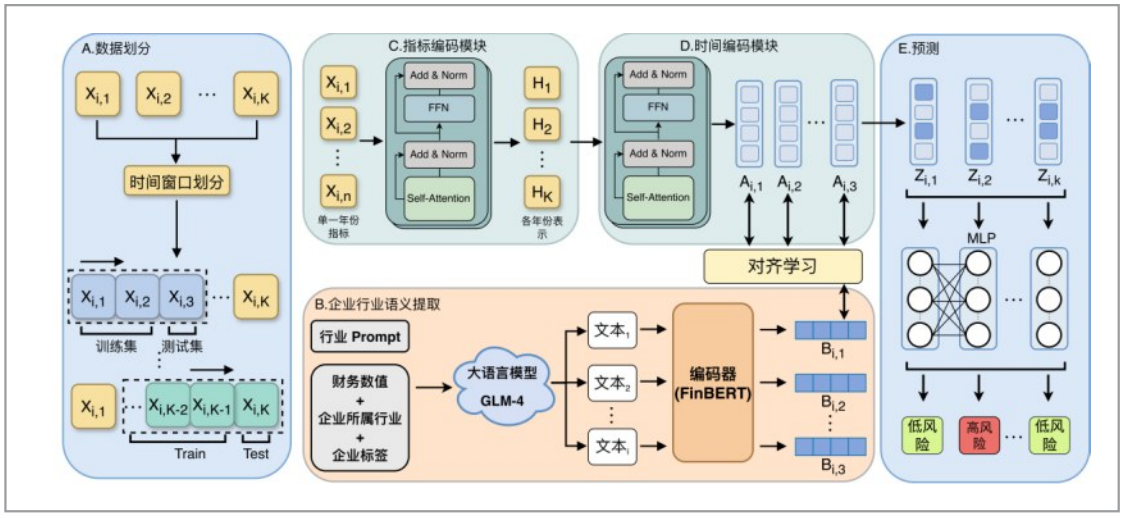

图1 整体架构图

Fig. 1 Comprehensive Architectural Diagram

3.1.1 行业文本生成

财务指标的风险含义受行业经营模式及指标组合影响。为引入行业背景，本文将企业所属行业、核心财务指标及其年度变化组织为提示词，引导 GLM-4 生成行业感知的财务风险文本。具体而言如图 2 所示，对于企业样本 i 设其在观察窗口内的财务面板数据表示为 x_i，行业类别表示为 c_i。文本生成过程表示为：

$$x_{text} = G(x_i, c_i) \tag{1}$$

其中，G 大语言模型，x_{text} 表示生成的行业感知财务语义文本。需要说明的是，标签仅用于训练集的文本生成，所有测试样本在预测过程中均不输入真实标签信息。

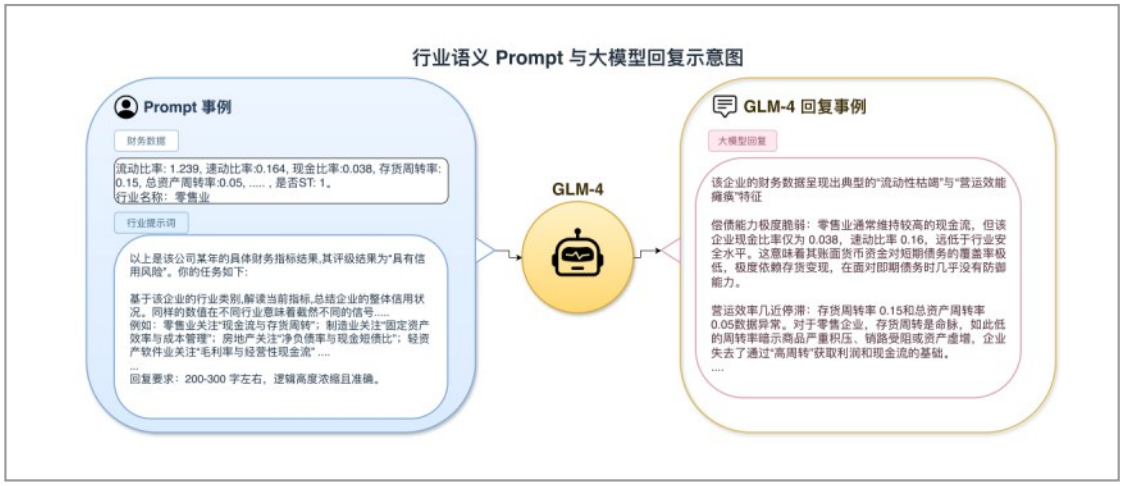

图2 基于 GLM-4 的行业感知企业财务表格数据语义增强示例

Fig. 2 Example of semantic enhancement of industry-aware corporate financial spreadsheet data using GLM-4

3.1.2 文本特征提取

考虑到生成的文本是具有强烈情感倾向的金融文本，因此本文采用金融领域专用预训练语言模型 FinBERT2 作为文本编码器。FinBERT2 能够更好地捕捉金融文本中的专业术语、语义倾向和上下文关系。因此，FinBERT2 提取的文本特征可以简洁地表示为：

$$z_i^{text} = FinBERT2(x_{text}) \tag{2}$$

其中，z_i^{text} 是提取的文本特征，x_{text} 是生成的文本。

3.2 双阶段 Transformer

设企业样本的财务面板表示为 $X \in R^{T \times N}$，其中 T 表示观察窗口内的年份数，N 表示财务指标数，第 t 行向量 $x_t \in \mathrm{R}^N$ 表示企业在第 t 年的全部财务指标。第一阶段通过指标编码学习同一年度内财务指标之间的组合关系，并聚合形成年度表示；第二阶段将各年度表示按时间顺序进行编码，学习企业财务状态的跨年度依赖。

在第一阶段中，模型首先将第 t 年第 j 个财务指标的数值映射到 d 维隐空间，并结合指标身份嵌入保留变量语义。其数值嵌入可表示为：

$$e_{t,j}^{val} = W_p x_{t,j} + b_p \tag{3}$$

其中，$W_p \in \mathrm{R}^d$ 和 $b_p \in \mathrm{R}^d$ 为可学习参数。随后，模型将数值嵌入与对应的指标身份嵌入相加，并在序列首部加入可学习的 [CLS] 标记，用于汇聚该年份的整体财务状态。为避免公式冗余，本文将指标身份编码、分类标记拼接以及第一阶段指标编码统一表示为：

$$H_t^{ind} = Transformer_{ind}(E_t) \tag{4}$$

其中，$Transformer_{ind}$ 表示指标维编码器，H_t^{ind} 为第 t 年的指标交互表示，取 [CLS] 位置输出作为该年份的年度表示 h_t。该过程使模型能够在同一年度内识别盈利能力、偿债能力、现金流与运营效率等指标之间的组合风险信号。

在第二阶段中，将各年份表示 $[\mathrm{h}_1, \mathrm{h}_2, \ldots, \mathrm{h}_t]$ 按时间顺序组成序列，并叠加位置编码后输入时间维 Transformer，以学习企业财务状态的跨期依赖关系：

$$Z = Transformer_{temp}([\tilde{h}_1, \tilde{h}_2, \ldots, \tilde{h}_t]) \tag{5}$$

其中，$Transformer_{temp}$ 表示时间维编码器，$\mathrm{Z} = [\mathrm{z}_1, \mathrm{z}_2, \ldots, \mathrm{z}_t]$ 为第二阶段的输出序列。通通过第二阶段编码，模型能够刻画风险从早期经营压力到后续偿债能力恶化的连续演化过程。

考虑到不同企业的历史年份长度可能不一致，本文采用 padding 与 attention mask 对序列进行补齐，并仅保留有效年份对应的编码结果。最终取最后一个有效年份的输出作为企业级表格表示：

$$z = z_t \tag{6}$$

3.3 跨模态对齐学习

在获得表格分支的结构化表示后，本文通过对齐学习将表格表示与文本表示映射到统一语义空间，拉近同一样本的跨模态表示、拉远异样本表示，从而将文本语义注入表格特征。

以同一样本的表格表示与文本表示为正样本、其余组合作为负样本，构造表格与文本对比损失，定义为：

$$\mathrm{L} = -\frac{1}{B}\sum_{i=1}^{B} \log \frac{\exp\left(sim\left(\mathrm{z}_i^{\sim num}, \mathrm{z}_i^{\sim text}\right)\right)}{\sum_{j=1}^{B} \exp\left(sim\left(\mathrm{z}_i^{\sim num}, \mathrm{z}_j^{\sim text}\right)\right)} \tag{7}$$

其中，$\tilde{z}_i^{num}$ 和 $\tilde{z}_i^{text}$ 分别表示第 i 个样本的表格表示与文本表示，sim 是余弦相似度函数。该损失鼓励第 i 个表格样本与其对应的第 i 个文本样本具有更高相似度，同时与 batch 内其他文本样本保持区分。

训练完成后，保留表格分支投影网络生成对齐后的表格特征：

$$h_i^{align} = \frac{f_{num}(x_i^{num})}{|f_{num}(x_i^{num})|_2} \tag{8}$$

其中，h_i^{align} 为对齐增强后的企业表示，作为后续分类器的输入。

3.4 风险预测

在获得经过对齐学习后的融合语义表示后，本文采用多层感知机（MLP）[27]对企业的最终表示进行分类预测。具体而言，对于第 i 个企业样本，其融合后的语义表示记为 h_i，则模型的预测输出为：

$$\hat{y}_i = Softmax(MLP(\boldsymbol{h}_i)) \tag{9}$$

考虑到高风险与低风险样本分布不均衡，模型采用加权交叉熵损失：

$$\mathrm{L} = -\sum_{i=1}^{N} \omega_{y_i} y_i \log(\hat{y}_i) \tag{10}$$

其中，y_i 为企业 i 的真实标签，$\hat{y}_i$ 为通过 Softmax 函数得到的企业 i 的预测概率；ω_{y_i} 为类别 y_i 的权重，类别权重依据训练集中各类别样本数量的反比设置，使样本数量较少的高风险类别获得更高的损失权重，以缓解类别不均衡对模型训练的影响。权重计算仅使用训练集的类别分布信息。

4 实验结果与分析

4.1 数据集

本文数据来自中国经济金融研究数据库（CSMAR）。首先，以注册资本不低于 2 000 万元且经营历史较长的企业为候选核心企业，再依据供应商关系识别其上游一级企业和关键供应商。考虑到 2016 年以来供应链数字化转型和金融科技应用持续推进，本文选取 2016 至 2025 年的企业年度数据，以刻画链上企业信用风险随时间变化的动态演化特征。最终，从 A 股中筛选约 700 家上市企业作为研究样本，涵盖化学原料及化学品制造业、非金属矿物制品业以及计算机应用服务业等相关细分行业。按企业与年度整理后，共形成覆盖 2016 至 2025 年的 6,895 条观测记录。样本中低风险与高风险企业的比例约为 20∶1。在信用风险刻画方面，本文采用“特别处理”（ST 或*ST）标识作为企业信用风险的客观代理变量。其中，ST 是中国证券交易所对财务状况异常企业施加的监管标签，而 *ST 则表示企业已面临较高的退市风险。这两类标识在中国资本市场中被广泛视为企业信用恶化的重要信号。因此，本文将被标记为 ST 或 *ST 的企业视为高风险企业。在数据预处理方面，本文统一变量类型并完成特征匹配。对于连续型财务指标的缺失值，采用训练集中相应指标的中位数进行填补，测试集沿用训练集确定的填补值。随后，采用 Z-score 方法对连续型财务指标进行标准化处理。各指标的均值和标准差仅由训练集估计，测试集沿用训练集确定的变换参数，以保持训练与测试数据转换的一致性。观察期内有 128 家企业发生过状态变化，共观察到 269 次状态切换，包括 140 次进入风险状态和 129 次退出风险状态。本文以企业—目标年份为样本单位构建风险预测任务，对于目标年份 T，选取该企业此前连续 m 个年

度，即 T−m 至 T−1 年的财务数据作为观察窗口，并以企业在 T 年的ST/*ST状态构建风险标签。其中，目标年份处于ST或 *ST 状态的企业标记为高风险样本，其余企业标记为低风险样本。因此，本文实验考察的是利用历史财务信息预测下一年度企业风险状态的能力。为确保数据集能够完整刻画企业信用风险的演化过程，本文并未剔除最终退市或破产的企业，而是将其保留在样本中。对于此类企业，其退市年份被视为时间序列的终止点，从而使模型能够学习企业在退出市场前的动态风险演化特征。模型输入包括标的企业自身的财务指标、基本信息、行业属性、客户集中度、供应商集中度和供应链集中度。具体指标见表1和表2。

4.2 实验设置

实验在Windows系统下完成，硬件配置为NVIDIA RTX 4080 SUPER 16 GB GPU、14核Intel CPU和64 GB内存，软件环境为CUDA 12.1、Python 3.10和PyTorch 2.1.0。模型采用Adam优化器，学习率为0.05，正则化参数设置为 5e−2，Dropout为0.5。为保证对比实验的公平性，各基线模型均依据原始文献及公开实现中的推荐配置确定初始参数。在样本构建过程中，财务指标及其年度变化特征均限定于对应的历史观察窗口，目标年份的财务指标不作为该样本的预测输入。文本分支与表格分支采用一致的财务观察窗口，以保证两种表示在时间范围上的对应关系；测试阶段不向模型提供目标年份的真实风险标签。在数据划分方面，本文采用基于时间顺序的数据集划分方法，以模拟供应链金融风险预测中利用历史信息预测未来风险的实际场景。具体而言，本文基于连续年度数据构建滑动时间窗口，并按照窗口时间先后将历史样本划分为训练集和测试集：前期窗口样本用于模型训练，后期窗口样本仅用于最终性能评估。本文采用长度为3年的滑动观察窗口，以企业连续3年的历史财务数据作为输入，预测下一年度的风险状态。例如，以2016—2018年的财务数据为观察窗口，以2019年的ST/*ST状态作为对应风险标签；随后将窗口逐年向后移动，依次构建2017—2019年对应2020年、2018—2020年对应2021年等预测样本，直至以2022—2024年的财务数据预测2025年的风险状态。在分类决策阶段，以模型输出的高风险类别预测概率作为判别依据，采用0.5作为固定分类阈值，将预测概率转换为高风险或低风险类别。

4.3 指标体系

本文评价指标选取准确率、精确率、召回率、F1值、马修斯相关系数(MCC)、平衡准确率和精确率−召回率曲线下面积（AUPRC)。其中，TP表示被正确预测为正类的样本数，TN表示被正确预测为负类的样本数，FP表示被错误预测为正类的样本数，FN表示被错误预测为负类的样本数。

准确率（Accuracy)：

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

精确率 (Precision):

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

召回率（Recall)：

$$Recall = \frac{TP}{TP + FN} \tag{13}$$

F1值（F1-Score）：

$$F1=\frac{2\times(Precision\times Recall)}{Precision+Recall} \quad (14)$$

马修斯相关系数（MCC）：

$$MCC=\frac{TP\times TN-FP\times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (15)$$

平衡准确率（Balanced Accuracy，BA）：

$$BA=\frac{1}{2}\left(\frac{TP}{TP+FN}+\frac{TN}{TN+FP}\right) \quad (16)$$

精确率-召回率曲线下面积（area under the precision-recall curve，AUPRC）：

$$AUPRC=\int_0^1 P(R)\,dR \quad (17)$$

表1 财务指标

Table 1 Enterprise financial indicators

指标	变量	指标解释
	资产报酬率	衡量每单位资产带来的综合回报能力
	总资产净利润率	反映资产创造净利润的效率
盈利能力	营业毛利率	体现主营业务的盈利空间
	营业利润率	衡量主营业务获利能力
	成本费用利润率	反映投入成本费用后的盈利水平
	净利润增长率	企业净利润的年度增长速度
	利润总额增长率	整体利润规模的增长趋势
成长能力	营业利润增长率	主营业务利润的增长幅度
	总资产增长率	企业资产扩张的速度
	主营收入增长率	主营业务收入的增长能力
	收益留存率	反映企业内部利润再投资比例
	应收帐款周转率	显示应收账款的回收效率
	存货周转率	衡量存货变现速度与管理效率
运营能力	流动资产周转率	流动资产使用效率的体现
	固定资产周转率	固定资产利用创造收入的能力
	总资产周转率	总体资产运用效率的指标
	流动比率	衡量短期偿债能力的总体指标
	速动比率	衡量剔除存货后的流动性水平
偿债能力	现金比率	反映以现金立即偿还短期债务的能力
	长期负债权益比率	衡量资本结构中长期债务压力
	流动负债比率	衡量短期负债在总负债中的占比
	长期债务与营运资金比率	反映长期负债对营运资金的占用程度

4.4 实验结果分析

4.4.1 对比实验

对比方法包括传统机器学习模型LR、

表2 非财务指标

Table 2 Non-financial indicators

指标	变量	指标解释
企业基本数据	注册资本	企业的法定资本金
	成立年份	企业正式成立并开始运营的年份
	所属行业	企业主营业务所归属的行业类别
供应链运营程度	客户集中度	主要客户对企业收入的依赖程度
	供应商集中度	企业采购来源的集中程度
	供应链集中度	整体供应链关系的集中性与依赖度

MLP 和 SVM，集成学习模型 RF、AdaBoost、XGBoost 和 LightGBM，以及深度表格模型 TabNet、AutoInt 和 FT-Transformer。此外，本文选取融合文本与财务信息的 DeepRRP[21]、采用多通道财务特征图的 4C-FinNet[25]和结合时间折叠与风险对比学习的 TemFRC[26]作为对比方法，以比较不同建模方式的风险识别表现。对比实验结果见表 3。本文实验采用多次独立运行并取均值的方式汇总结果，表 3 所列数值均为相应评价指标在多次运行中的平均值。

表3 基线模型比较结果

Table 3 Baseline model comparison results

模型	Accuracy	Balanced Acc	Precision	Recall	F1-Score	MCC	AUPRC
LR	0.6954	0.7008	0.5980	0.7277	0.6565	0.3936	0.5796
MLP	0.7265	0.7378	0.6243	0.7943	0.6991	0.4661	0.6382
XGBoost	0.7643	0.7760	0.6632	0.8342	0.7390	0.5407	0.6984
AdaBoost	0.7870	0.7978	0.6892	0.8516	0.7618	0.5836	0.7315
LightGBM	0.8028	0.8184	0.6972	0.8961	0.7842	0.6241	0.7718
TabNet	0.8342	0.8482	0.7340	0.9183	0.8159	0.6824	0.8126
SVM	0.8565	0.8704	0.7589	0.9398	0.8397	0.7258	0.8439
RF	0.8719	0.8850	0.7782	0.9505	0.8558	0.7546	0.8619
4C-FinNet	0.8754	0.8849	0.7926	0.9324	0.8568	0.7555	0.8617
AutoInt	0.8981	0.9079	0.8187	0.9570	0.8825	0.8010	0.8898
TemFRC	0.9009	0.9094	0.8268	0.9517	0.8849	0.8048	0.8915
FT-Transformer	0.9064	0.9161	0.8293	0.9644	0.8918	0.8172	0.9024
DeepRRP	0.9131	0.9188	0.8519	0.9476	0.8972	0.8259	0.9121
ISemTAR	0.9274	0.9296	0.8849	0.9408	0.9120	0.8514	0.9217

如表 3 所示，ISemTAR 的 Accuracy、Precision、F1-Score 和 MCC 分别达到 0.9274、0.8849、0.9120 和 0.8514，均高于各基线。FT-Transformer 取得最高 Recall（0.9644），但其 Precision 低于 ISemTAR，表明较高的风险识别敏感性伴

随更多误报。与FT-Transformer相比，ISemTAR的Accuracy、Precision、F1-Score和MCC分别提高2.10、5.56、2.02和3.42个百分点，说明行业语义增强和双阶段编码有助于改善综合分类表现。

4.4.2 消融实验

本文设置四个消融变体：ISemTAR-a去除行业语义文本，仅使用结构化指标；ISemTAR-b去除跨模态对齐；ISemTAR-c去除年度内指标编码层；ISemTAR-d去除跨年度年份编码层。

图3显示，去除任一模块后模型性能均有所下降。其中，ISemTAR-a在Precision和MCC上的降幅最明显，说明行业语义能够补充数值指标难以直接表达的风险含义；ISemTAR-b的结果表明，跨模态对齐是将文本语义迁移至表格表示的重要环节；ISemTAR-c和ISemTAR-d则分别验证了年度内指标组合关系与跨年度状态演化信息的作用。

4.4.3 超参数实验

为进一步考察关键超参数设置对模型性能的影响，本文围绕学习率与Transformer表示维度开展了对比实验。

图4给出了学习率和Transformer表示维度的敏感性结果。随着学习率增大，Accuracy和MCC均呈先升后降趋势，并在学习率为0.05时达到最高值；学习率过大会降低训练稳定性。表示维度增加至512时模型性能最优，继续增加至1 024维后略有下降，表明过高维度可能引入冗余参数并增加过拟合风险。

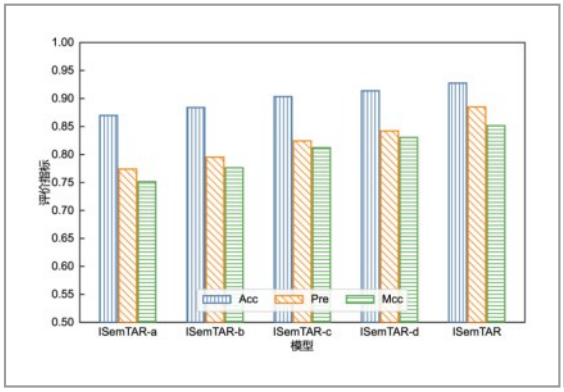

图3 消融实验

Fig. 3 Ablation experiment

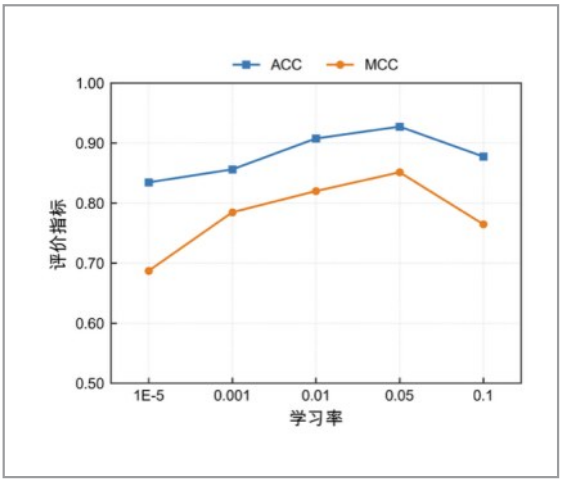

图4(a) 学习率实验

Fig. 4(a) Learning rate experiment

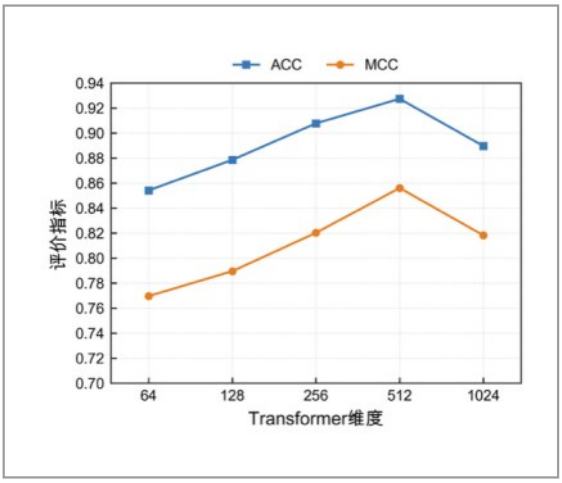

图4(b) Transformer维度实验

Fig. 4(b) Transformer Dimension Experiments

4.4.4 描述文本一致性评估

为检验生成文本能否准确反映原始财务数据，本文采用FactCC进行事实一致性评估[28]。由于本研究不存在标准参考文本，依赖参考文本表面相似度的BLEU和ROUGE并不适用。本文将“流动比率为2.87”“速动比率为1.46”等结构化指标转换为源陈述，再与生成文本共同输入FactCC。

实验结果如表4所示，本文生成文本的FactCC得分达到0.9648。该结果表明，GLM-4生成的行业语义增强文本与原始财务数据之间具有较高的事实一致性，能够较为准确地反映企业财务状况及其相关特征。同时，较高的一致性得分也说明，本文所采用的文本生成策略整体上具有较好的可靠性，生成文本中出现明显事实性错误的情况较少。

表4 描述文本事实一致性评估结果

Table 4 Results of the descriptive text fact consistency assessment

模型	得分
FactCC	0.9648

4.4.5 大模型评估实验

为分析提示信息的作用，本文设置三种对比方式：ISemTAR-e仅在训练集提示词中去除真实标签，保留财务指标和行业信息；ISemTAR-f去除行业信息，保留财务指标和训练标签；GLM-4则直接判断企业风险，不使用后续文本编码、表格编码和跨模态对齐模块。所有测试样本在预测过程中均不输入真实标签。

如图5所示，去除训练标签提示后，模型的Accuracy和MCC明显下降，说明标签约束能够减少过度中性的财务描述；去除行业信息后，模型性能同样下降，表明行业背景有助于解释相同财务指标在不同经营模式下的风险含义。仅使用GLM-4直接预测时，MCC为0.1228，说明大语言模型难以独立识别类别不平衡条件下的分类边界。

无标签提示变体的性能下降表明，训练阶段的标签语义监督有助于增强生成文本的风险指向性，并影响后续的跨模态对齐效果。其中，提示词中的标签信息仅用于训练样本的语义监督，测试阶段的风险判断仍基于历史观察窗口内的财务信息完成。

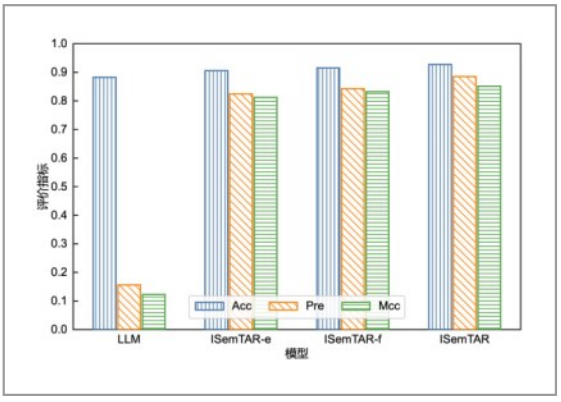

图5 大模型评估实验

Fig. 5 Large Model Evaluation Experiment

4.4.6 跨模态对齐学习可视化

为进一步直观分析跨模态对齐学习的实际作用，本文将对齐前后的表格表示与文本表示映射至二维空间，其中蓝色圆点和橙色三角形分别表示结构化财务数据和行业语义文本的表示。如图6(a)所示，对齐前两种模态分布在不同区域，匹配样本与非匹配样本的平均余弦相似度分别为0.094和0.096，差异较小，说明同一企业

的表格表示与文本表示尚未形成稳定关联。

如图 6(b)所示，对齐后匹配样本的平均余弦相似度提升至 0.392，非匹配样本降至−0.003，表明模型能够拉近同一企业的跨模态表示，并区分不同企业的表示。由此，文本中的行业背景、财务变化及风险含义能够迁移至结构化财务表示空间，使模型获得更具判别性的企业风险表示。

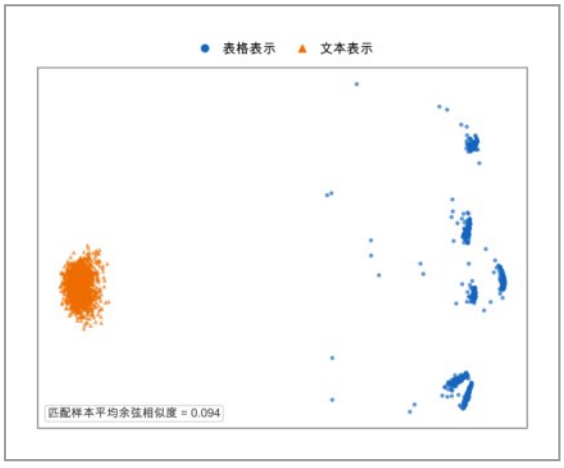

图 6(a)　跨模态对齐前的表格表示与文本表示分布

Fig. 6(a)　Distribution of tabular and textual represen - tations before cross–modal alignment

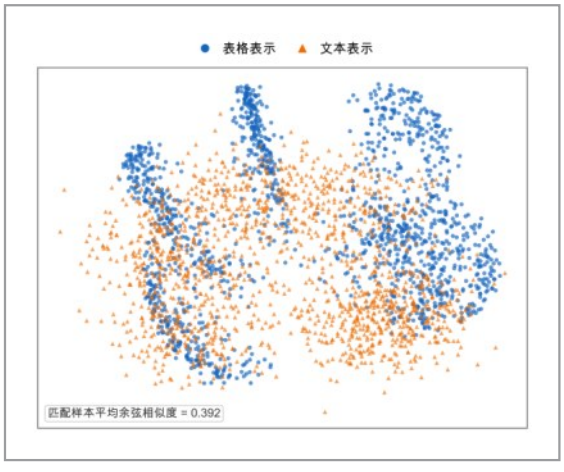

图 6(b)　跨模态对齐后的表格表示与文本表示分布

Fig. 6(b)　Distribution of tabular and textual represen - tations after cross–modal alignment

5　结论

本文提出 ISemTAR，一种基于大语言模型的行业语义增强供应链金融企业信用风险评估方法。该方法利用 GLM−4 将企业所属行业、财务指标及年度变化转化为行业感知描述文本，采用 FinBERT2 提取语义特征，并通过跨模态对齐将行业语义迁移至结构化财务表示；进一步通过双阶段 Transformer 依次学习年度内指标组合关系与跨年度状态变化。实验结果表明，ISemTAR 的 AUPRC、F1 值和平衡准确率分别达到 0.9217、0.9120 和 0.9296，MCC 达到 0.8514，综合分类表现优于所比较的基线模型。消融实验和 FactCC 评估进一步验证了行业语义、跨模态对齐及双阶段编码的有效性，以及生成文本与源财务信息的一致性。

本文实证分析主要基于供应链关联上市企业的财务面板数据，采用 ST/*ST 状态作为企业信用风险的代理标签。现有特征主要反映企业财务状况、行业属性及供应链集中程度，尚未充分覆盖应收账款确权、交易账期、核心企业增信及上下游风险传导等业务信息，模型在具体供应链融资业务中的适用性仍需进一步检验。后续可结合供应链交易关系，匹配上下游企业的流动性、现金流、应付账款和付款能力等指标，并在文本生成与风险评估中显式建模供应链风险传导关系。

参考文献：

[1] Lokanan M E, Maddhesia V. Supply chain fraud prediction with machine learning and artificial intelligence[J]. In -

ternational Journal of Production Research, 2025, 63(1): 286–313.

[2] Ni Z, Jiang M, Zhan W. Default prediction with industry-specific default heterogeneity indicators based on the forward intensity model[J]. Axioms, 2023, 12(4): 402.

[3] Wu S, Irsoy O, Lu S, et al. BloombergGPT: A large language model for finance [EB/OL]. (2023-03-30) [2026-11-09]. https://arxiv.org/abs/2303.17564.

[4] Yang H, Liu X Y, Wang C D. FinGPT: Open-source financial large language models[EB/OL]. (2023-06-09) [2026-11-09]. https://arxiv.org/abs/2306.06031.

[5] Xie Q, Han W, Zhang X, et al. PIXIU: A comprehensive benchmark, instruction dataset and large language model for finance[C]//Advances in Neural Information Processing Systems 36. Red Hook: Curran Associates, 2023: 33469-33484.

[6] 程大伟, 贾仁军, 李江彤, 等. 知识增强的中文金融大模型研究[J]. 大数据, 2025, 11(2): 5-18. DOI: 10.11959/j. issn. 2096-0271.2025021.
Cheng D W, Jia R J, Li J T, et al. Research on knowledge-augmented Chinese financial large language model[J]. Big Data Research, 2025, 11(2): 5-18. DOI: 10.11959/j.issn.2096-0271.2025021.

[7] 王敏, 高晓影, 汪诗蕊, 等. 基于知识图谱和大语言模型增强的推荐系统研究[J]. 大数据, 2025, 11(2): 29-46. DOI: 10.11959/j.issn.2096-0271.2025026.
Wang M, Gao X Y, Wang S R, et al. Research on recommendation system based on knowledge graph and large language model enhancement[J]. Big Data Research, 2025, 11(2): 29-46. DOI: 10.11959/j.issn.2096-0271.2025026.

[8] 乔永博, 陈胜利. 科技型中小企业供应链金融风险测量研究[J]. 财务与金融, 2021(5): 6-14.
Qiao Y B, Chen S L. Research on supply chain finance risk measurement of technology-based SMEs[J]. Finance and Finance, 2021(5): 6-14.

[9] 王鑫, 王莹. 基于LSTM-CNN的中小企业信用风险预测[J]. 系统科学与数学, 2022, 42(10): 2698-2711.
Wang X, Wang Y. Credit risk prediction of small and medium-sized enterprises based on LSTM-CNN[J]. Journal of Systems Science and Mathematical Sciences, 2022, 42(10): 2698-2711.

[10] 黄苒, 冯小钰. 供应链信用风险传染、银行策略与风险控制[J]. 系统管理学报, 2024, 33(1): 137-149.
Huang R, Feng X Y. Supply chain credit risk contagion, bank strategy and risk control[J]. Journal of Systems & Management, 2024, 33(1): 137-149.

[11] Wang S, Chi G. Cost-sensitive stacking ensemble learning for company financial distress prediction[J]. Expert Systems with Applications, 2024, 255: 124525.

[12] Pellegrino M, Lombardo G, Adosoglou G, et al. A multi-head LSTM architecture for bankruptcy prediction with time series accounting data[J]. Future Internet, 2024, 16(3): 79.

[13] Huang M N, Lee H H. Inter-industry network and credit risk[J]. International Review of Economics & Finance, 2024, 92: 598-625.

[14] Spatareanu M, Manole V, Kabiri A, et al. Bank default risk propagation along supply chains: Evidence from the U.K. [J]. International Review of Economics & Finance, 2023, 84: 813-831.

[15] Sigrist F, Leuenberger N. Machine learning for corporate default risk: Multi-period prediction, frailty correlation, loan portfolios, and tail probabilities [J]. European Journal of Operational Research, 2023, 305(3): 1390-1406.

[16] Bitetto A, Filomeni S, Modina M. Machine learning for the unlisted: Enhancing MSME default prediction with public market signals[J]. Journal of Corporate Finance, 2025, 94: 102830.

[17] Zhang W, Wang J. Credit risk contagion in complex companies network: Empirical research based on listed agricultural companies[J]. Economic Analysis and Policy, 2024, 82: 938-953.

[18] Golec M, Alabduljalil M. Interpretable LLMs for credit risk: A systematic review and taxonomy[J]. Expert Systems with Applications, 2026, 306: 130941.

[19] Che W, Wang Z, Jiang C, et al. Predicting financial distress using multimodal data: An attentive and regularized deep learning method[J]. Information Processing & Management, 2024, 61(4): 103703.

[20] Fan S, Wu Y, Yang R. Measuring firm-level supply chain risk using a generative large language model[J]. Finance Research Letters, 2025, 77: 107111.

[21] Hou L, Du S, Bi G. From text to risk: Predicting repayment risk in supply chain finance with deep learning and large language models[J]. International Journal of Production Economics, 2026, 293: 109906.

[22] Wu Z, Dong Y, Li Y, et al. Unleashing the power of text for credit default prediction: Comparing human-written and generative AI-refined texts[J]. European Journal of Operational Research, 2025, 326(3): 691-706.

[23] Wang Z, Jiang C, Zhao H. Depicting risk profile over time: A novel multiperiod loan default prediction approach[J]. MIS Quarterly, 2023, 47(4): 1455-1486.

[24] Korangi K, Mues C, Bravo C. A transformer-based model for default prediction in mid-cap corporate markets [J]. European Journal of Operational Research, 2023, 308(1): 306-320.

[25] Yang S, Li C, Li Z, et al. 4C-FinNet: A multi-channel feature map and spatio-temporal feature fusion method for financial risk prediction[J]. Expert Systems with Applications, 2025, 284: 127789.

[26] Wang M, Huang W, Xie W, et al. Tem-FRC: Enterprise financial risk prediction with temporal folding and risk contrast [J]. Information Processing & Management, 2026, 63(7): 104780.

[27] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors[J]. Nature, 1986, 323 (6088): 533-536.

[28] Kryściński W, McCann B, Xiong C, et al. Evaluating the factual consistency of abstractive text summarization[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. Online: Association for Computational Linguistics, 2020: 9332-9346.

作者简介



刘颖（1979—），女，博士研究生，吉林财经大学人工智能与计算机学院，教授，主要研究方向为图神经网络、金融科技等。



陈嘉茂（2001—），男，硕士研究生，吉林财经大学人工智能与计算机学院，主要研究方向为大语言模型、金融科技等。



孙楠（2000—），女，硕士研究生，吉林财经大学人工智能与计算机学院，主要研究方向为图神经网络、金融科技等。

录用日期: XXXX-XX-XX
通信作者:
基金项目:
Foundation Items: