

基于MD&A文本隐式情感的上市公司财务困境预测

闫志华¹, 黄晓辉^{2,3}, 唐锡晋^{2,3*}, 闫绪娴¹

(1. 山西财经大学管理科学与工程学院, 山西太原 030006; 2. 中国科学院数学与系统科学研究院, 北京 100190; 3. 中国科学院大学经济与管理学院, 北京 100049)

摘要: 管理层讨论与分析(Management discussion and analysis, MD&A)文本中包含大量的隐式情感信息, 无法使用情感词直接判断情感类别. 提出融合句子结构特征、上下文语义特征和外部金融知识与情感信息的金融文本隐式情感识别模型: 使用依存句法分析将 MD&A 文本中的语句表示为依存句法树, 基于图注意力机制获得关键语句的结构特征; 使用 Bi-LSTM 和注意力机制获取关键上下文语义信息; 使用 FinBERT 金融预训练模型引入金融领域知识和情感信息. 在此基础上, 将 MD&A 文本的隐式情感特征应用于企业财务困境预测任务, 提升机器学习算法的预测能力. 实验表明, 该模型可以提升 MD&A 文本的隐式情感特征识别能力, 融合隐式情感特征的机器学习模型可以大幅度提升上市公司财务困境预测效果.

关键词: 财务困境预测; 隐式情感分析; 深度学习; 图神经网络

中图分类号: F272

文献标识码: A

文章编号: 1000-5781(2026)01-0180-14

doi: 10.13383/j.cnki.jse.2026.01.012

Financial distress prediction of listed companies based on implicit sentiment of MD&A

Yan Zhihua¹, Huang Xiaohui^{2,3}, Tang Xijin^{2,3*}, Yan Xuxian¹

(1. School of Management Science and Engineering, Shanxi University of Finance and Economics, Taiyuan 030006, China; 2. Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China; 3. School of Economics and Management, University of Chinese Academy of Sciences, Beijing, 100049, China)

Abstract: Management discussion and analysis (MD&A) text contains a large amount of implicit sentiment information, which makes it difficult to use sentiment words to directly determine sentiment polarity. This paper proposes a financial text implicit sentiment detection model (SC-ISA) that fuses sentence structural features, contextual semantic features, and external financial sentiment information. Specifically, the model: (1) uses dependent syntactic analysis to represent the sentence as a tree structure to obtain key sentence structural features; (2) employs Bi-LSTM with an attention mechanism to obtain key contextual semantic information; and (3) utilizes FinBERT to incorporate financial domain knowledge information. Experiments show that the SC-ISA model can improve the performance of implicit sentiment detection based on MD&A, and that the implicit sentiment features can substantially improve the accuracy of the prediction of financial distress in listed companies.

Key words: financial distress prediction; implicit sentiment analysis; deep learning; graph neural network

收稿日期: 2024-11-18; 修订日期: 2025-08-09.

基金项目: 国家社会科学基金资助项目(23&ZD331); 山西省哲学社会科学规划资助项目(2024YB090); 山西省研究生教育创新计划支持资助项目(2024TD25).

*通信作者

1 引言

财务困境是指由于企业内部决策失误或者外部市场竞争而出现盈利能力下降, 甚至出现破产倒闭的现象。企业陷入财务困境会降低企业的信用评级, 对股价和债券价格等产生负面影响, 给投资人带来经济损失^[1,2]。例如, 康得新连续四年财务造假, 虚增利润 115.3 亿元, 2014 年至 2017 年的财务报表均显示存在经营现金流波动大、高存高贷、关联股东占用资金和商誉高企等问题。因此, 有效的财务困境预测, 不仅可以帮助管理层更加清晰的掌握企业的运营状态, 及时调整经营策略, 提高抵御市场风险的能力, 而且能够保护股东和债权人的权益。国内外大量的金融机构提供企业财务困境预测的服务, 如国泰君安、Eikon&Datastream 等。

通常, 财务困境预测基于反映企业经营状态的关键财务指标构建预测指标体系, 使用统计学方法和机器学习算法对财务状况进行预测。常见的预测模型包括 Z-score 模型^[3]、LightGBM^[4]、深度神经网络^[5] 等。合理的预测指标体系是企业财务困境预测的关键, 最初的财务困境预测模型以企业的财务指标为主, 随着自然语言处理技术的发展, 从企业披露的财务文本中提取反映企业运营状态的指标可以提升模型的预测效果。MD&A 是企业财务报表的重要组成部分, 包含管理层对政策环境、市场风险、企业经营状况和企业发展战略等的讨论与预测, 反映了企业未来的经营状况, 在企业财务困境预测研究中应用广泛。现有的研究从 MD&A 文本中提取情感值、风险值、可读性和专业性等特征指标, 可以丰富企业财务困境预测指标体系, 提高企业财务困境模型的预测能力^[6,7]。

MD&A 文本情感特征包括显式情感和隐式情感, 是企业管理层对未来市场环境、经营风险和业务发展的预期。据统计, MD&A 文本中约 20% 的观点表达采用隐式情感。现有的研究主要使用情感词典法对显式情感进行识别与测度, 而隐式情感由于情感关键词缺乏、情感语义特征提取困难等原因关注较少。企业发展存在巨大的不确定性, 管理层为了提振投资者信心, 改善企业信用评级, 对企业的负面信息往往采用隐式情感进行表达。这些表达所采用的金融词汇具有较高的专业性, 句子的长度大, 包含的主体众多, 关键信息捕获与理解具有一定的难度, 给投资者和信用评级机构掌握企业的真实运营状况造成障碍。与当前隐式情感研究使用较多的社交媒体数据不同, MD&A 文本的句子较长, 语句结构复杂, 语义信息理解需要结合上下文语境。典型的 MD&A 隐式情感表述如表 1 所示。“具有跨行业应用的发展潜力”和“应用领域广泛”是对企业的技术实力和产品竞争力的积极评价, 表明语句 1 为正面隐式情感; 语句 2 为企业供应链的客观事实的描述, 为中性隐式情感; “较高的市场地位”、“垄断地位”和“客户集中度较高”的描述对象为“客户”, 客户的强势地位意味着企业的市场地位较低, 对企业的未来发展为负面评价, 结合上下文语义信息和金融专业知识才能判断语句 3 为负面隐式情感。因此, 采用机器学习方法和自然语言处理技术, 通过提取文本数据的上下文语义信息, 从海量的金融文本数据中识别隐式情感, 无论是对投资者的投资决策和评级机构的企业信用评价, 还是对金融市场政策的制定都非常有意义。

表 1 包含隐式情感的 MD&A 文本语句
Table 1 MD&A text statements containing implicit sentiments

序号	隐式情感语句	情感类别
1	由于高分子复合材料技术延展性好, 应用领域广泛, 公司相关产品具有跨行业应用的发展潜力。	正面
2	绝缘材料生产企业进入下游客户的供应链体系通常具有较高的门槛。	中性
3	客户在风力发电装备领域具有较高的市场地位, 并在国内轨道交通车辆制造领域占据垄断地位, 下游行业的市场格局导致公司客户集中度较高。	负面

综上所述, 目前的企业财务困境预测研究仍存在以下问题:

1) 隐式情感识别的研究对象主要是社交媒体文本, 针对金融文本的隐式情感的研究较少。与社交媒体的隐式情感识别不同, 以 MD&A 为代表的金融文本隐式情感识别具有词语序列敏感性、上下文语义相关性、情感特征长距离依赖和金融用语专业性等问题。

2) 现有的研究通过引入常识知识谱图^[8]、大语言模型^[9,10]解决领域专业用语的问题,使用 Bi-LSTM 和图注意力机制对上下文语义进行表征^[11],缺乏针对金融文本的隐式情感识别算法,没有解决金融文本中隐式情感特征识别困难的问题。

综合使用图神经网络、依存句法分析和大型语言模型等构建金融隐式情感识别算法,并应用于企业财务困境预测,主要贡献为:提出金融隐式情感的识别算法,使用 FinBERT 模型引入金融领域情感信息,使用依存句法分析构建图神经网络模型识别句子要素之间的层次结构信息,使用 Bi-LSTM 和注意力机制获得句子的上下文语义信息,实现隐式情感的高效识别;将隐式情感应用于企业财务困境预测,从 MD&A 文本中提取金融隐式情感作为特征变量,提升决策树、SVM 和 Catboost 等机器学习算法在财务困境预测任务的效果。

2 相关文献回顾

2.1 隐式情感分析

隐式情感分析是在没有显式情感词语的情况下,识别文本的情感类别,较显式情感分析任务难度更大。词语语义表示和隐式情感表示是隐式情感分析研究的基础问题,从文本中提取情感特征和引入外部情感信息是隐式情感分析的主要思路^[12]。

隐式情感分析的关键是情感特征信息的提取,如上下文语义信息和文本结构信息等。赵荣梅等^[13]提出基于 LSTM 模型和注意力机制提出单词级和句子级的语义和结构特征,进行中文隐式情感分析。张玉莹等^[14]使用 FinBERT 预训练模型生成词向量,使用 Bi-GRU 获得上下文语义信息,构建嵌入情感词汇的两极注意力机制实现情感增强,通过图卷积网络获得文本的语义依存特征,从判定句子的隐式情感类别。杨善良等^[15]构建文本和词语的异构图谱,使用图注意力机制计算词语对文本情感的贡献,通过注意力正交约束和分值约束提升模型效果。

知识图谱和大型语言模型包含了大量的领域情感信息,为隐式情感分析提供了新的思路。张军等^[16]提出基于 RoBERTa 融合双向长短期记忆网络及注意力机制的隐式情感识别模型,解决隐式情感分析中表达隐晦且缺少情感词的问题。Yin 等^[17]基于预训练模型对句子和上下文语义信息进行表示学习,提出隐式情感查询注意力机制进行隐式情感识别。Chen 等^[18]使用 BERT 预训练模型表示语义信息,使用 HowNet 表示句法结构,基于句子中情感与修辞之间的层次性和相关性进行隐式情感的识别。Fei 等^[19]提出三跳推理框架模拟思考过程,基于大语言模型和提示学习进行隐式情感的识别。Liao 等^[11]引入 ConceptNet 知识库中的情感信息增强语句的情感表达,提出情感常识图嵌入的多极正交注意力模型,解决隐式情感识别中情感词缺乏的问题。

综上所述,文本语义特征建模是当前隐式情感识别研究的主要思路,对文本的结构特征和大语言模型的研究较少。文本的结构特征反应了文本元素之间的层次结构特征,对于识别长句子的关键语义信息是非常重要的。大语言模型可以增加专业领域知识,更好的表征专业文本的语义信息,为隐式情感的识别提供了新的思路。

2.2 企业财务困境预测

企业财务困境预测是根据财务指标和非财务指标,使用统计学方法和机器学习方法判断企业未来是否出现财务问题。企业财务困境预测的研究包括财务困境的定义、预测指标的确定和预测方法的选择。

财务困境是企业所处的不良经营状态。衡量企业财务困境的指标有多种,如财务亏损、企业破产和债权人清算等。企业出现经营困难,可能是由于行业周期的影响,虽然不意味着会破产,但会影响企业的偿债能力。因此,选择恰当的财务指标确定企业财务困境是非常重要的。Zhao 等^[20]认为如果连续两年税息折旧及摊销前利润低于财务费用,则企业处于财务困境。Sun 等^[21]的研究将连续两年累计盈利为负或每股净资产

低于票面价值作为企业财务困境的判断标准. Che 等^[22]使用中国 A 股上市企业作为研究对象, 以“ST”标记判断企业是否陷入财务困境.

企业陷入财务困境首先表现为财务指标异常, 合适的财务指标选择是企业财务困境预测的关键. Kamaluddin 等^[23]从现金流量表出发, 使用 Z-score 模型检验偿付能力比率、盈利能力比率和效率比率与企业财务困境的关系, 为企业财务困境预测提供支持. Fernández-Gómez^[24]的实证分析表明国家的宏观经济和规制的异质性影响企业财务困境预测的效果. Liang 等^[25]的研究表明企业治理指标可以提升企业破产预测的效果. Citterio 和 King^[26]认为 ESG 指标可以降低将处于困境的企业预测为正常状态的企业, 从而提升模型的机器学习模型的预测能力. 此外, 使用自然语言处理技术从财务报告和社交媒体等数据中提取情感特征和语义特征都可以提升企业财务困境预测的效果, 如 Wang 等^[27]提出融合情感信息和文本信息的企业财务困境预测模型 IST-RS.

企业财务困境预测的方法可以分为统计学方法和机器学习方法. 常见的统计学方法包括判别式法、Logistic 回归和因子分析法等. Pereira 等^[28]使用 lasso 回归和岭回归模型对企业的财务困境进行预测. Almamy 等^[29]提基于 Z-score 模型的企业财务困境预测方法. 统计学方法具有易操作和解释性强的优点, 但是对数据满足分布的要求较高. 随着机器学习理论和方法的日益完善, 逻辑回归、SVM、随机森林、深度神经网络和图神经网络等方法在企业财务困境预测领域应用广泛. Wu 等^[30]提出融合多层感知神经网络 MLP-ANN 和 Z-score 模型混合预测方法, 对企业早期财务困境进行预测. 然而, 基于机器学习的企业财务困境预测面临着训练样本不足和数据样本不均衡的难题. Chen 等^[31]基于集成学习和标签支持向量机构造 Bagged-pSVM 和 Boosted pSVM 模型, 解决企业财务困境预测中训练样本不足的问题. Shen 等^[32]提出自适应递归 SMOTE 方法解决财务企业财务困境预测中数据样本不均衡和数据漂移的问题.

综上所述, 现有的企业财务困境预测研究以预测指标和预测方法的选择、样本不均衡问题等为主, 显式情感的识别方法、预测效果已经得到广泛的研究, 但是对金融文本中常见的隐式情感关注较少. 与显式情感相比, 隐式情感在企业财务报表中表达负面观点时更为常见, 可以丰富企业财务困境预测的指标体系, 提升机器学习算法预测的效果.

3 研究方法

给定 MD&A 文本句子集合 $M = \{S, C\}$, 其中 S 为包含隐式情感的语句, C 为同一段落中语句 S 的上下文语句. 提出基于句子 S 的结构特征、上下文语义特征和外部金融情感信息的金融文本隐式情感识别算法 SC-ISA, 实现给定语句 S 隐式情感的识别. 该算法包括语句特征识别模块、上下文语义识别模块和隐式情感判断模块, 如图 1 所示. 语句特征识别模块使用依存句法分析识别句子中短语结构和短语之间的层次关系, 通过图注意力机制 GAT 将隐式情感语句表示为向量. 上下文语义识别模块使用 Bi-LSTM 和注意力机制识别隐式情感语句及上下文语句中的关键信息, 获得语义信息的向量表示. 隐式情感判断模块使用 BiAffine 操作将句子结构特征和上下文语义特征融合, 使用 softmax 函数对 MD&A 文本句子进行分类, 确定句子的隐式情感类别.

3.1 MD&A 隐式情感识别

3.1.1 语句结构特征识别

使用依存句法分析识别 MD&A 文本中句子要素之间的层次结构关系, 基于图注意机制将文本表示为向量的形式.

1) 词语嵌入层. 为了将 MD&A 文本表示为向量, 使用 LTP 工具进行切词和词性标注. 切词后可以每个句子包含 m 个词语, 其中 w_i 表示第 i 个词语, 使用 word2vec 将词语表示为向量. 为了考虑词性、词语位置对句子隐式情感的影响, 分别对词性、词语位置采用随机初始化的方式表示为向量的形式. 词语向量、词性向量和词语位置向量拼接后得到句子的向量表示 $S_s = [w_{s1}, w_{s2}, \dots, w_{sm}]$.

2) 句子结构特征表示层. 与社交媒体文本不同, MD&A 文本以长句子为主, 依存句法分析可以识别句子中短语结构和短语之间的层次语法关系, 为隐式情感识别提供重要的信息. 使用 CoreNLP 工具对句子进行依存句法分析得到依存句法树, 使用无向图 $G = (V, E)$ 表示, 其中 V 表示节点的集合, E 表示边的集合, $(v_i, v_j) \in E$ 表示词语 w_{s_i} 和 w_{s_j} 之间的依存关系. 为了解决长句子中包含较多的噪声信息的问题, 引入图注意力机制对临近节点进行加权求和. 在第 l 层图注意力网络中, 词语 w_{s_i} 与邻居词语 w_{s_j} 的相似系数为 $e_{ij}^{(l)} = f([W^{(l)}S_{s_i}^{(l)}, W^{(l)}S_{s_j}^{(l)}])$, 则 w_{s_i} 的图注意力系数为

$$\alpha_{ij}^{(l)} = \frac{\exp(L(e_{ij}^{(l)}))}{\sum_{k \in N_i} L(e_{ik}^{(l)})}, \tag{1}$$

其中 L 为带泄露线性整流激活函数.

经过运算, 词语 w_{s_i} 向量表示为

$$h_{s_i}^{(l+1)} = \delta \left(\sum_{j \in N_{s_i}} \alpha_{ij}^{(l)} S_s^{(l)} h_{s_j}^{(l)} \right), \tag{2}$$

其中 N_{s_i} 表示词语 w_{s_i} 的邻居节点, δ 为激活函数. 最后, 句子 S 的结构特征表示为矩阵 $H_s = [h_{s_1}^{(l+1)}, \dots, h_{s_i}^{(l+1)}, \dots, h_{s_m}^{(l+1)}]$.

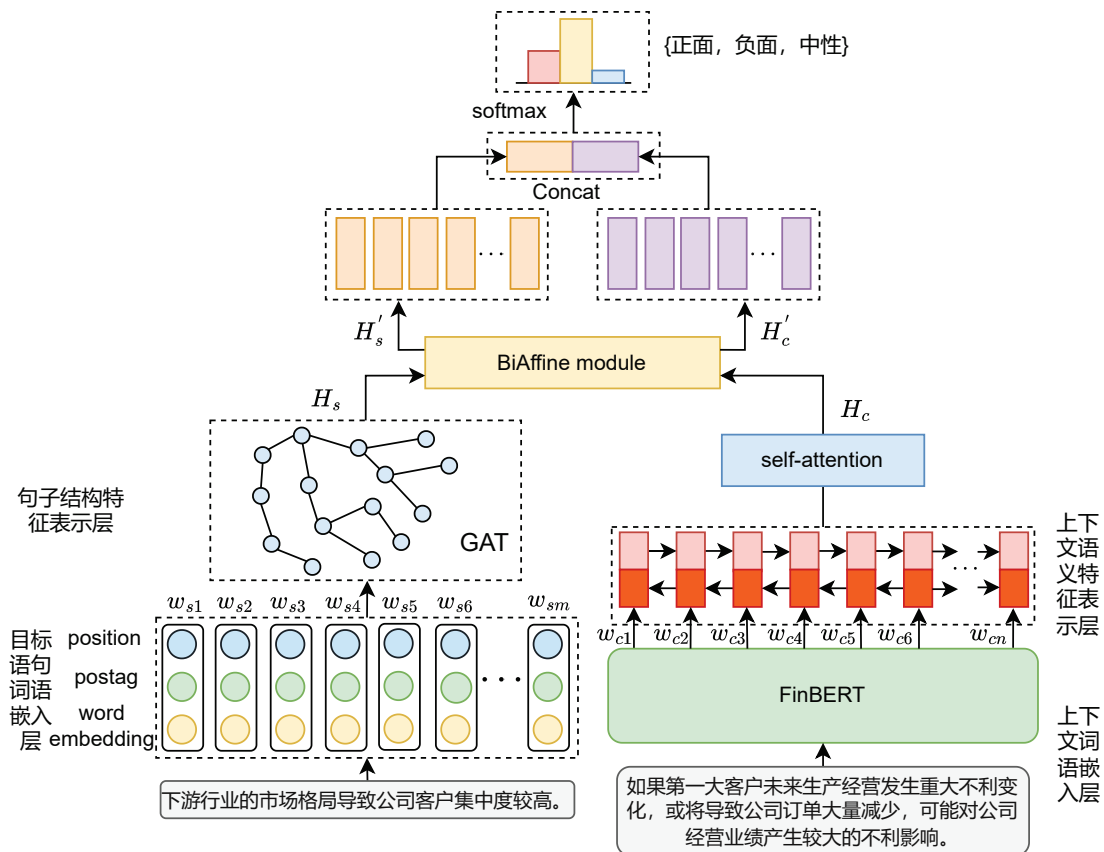


图 1 MD&A 文本隐式情感识别框架

Fig. 1 Framework of implicit sentiment analysis based on MD&A

3.1.2 上下文语义特征识别

MD&A 文本包含大量的金融专业术语和惯用表达方式, 可以隐晦的表达句子的情感信息. 为了识别这些句子所表达的隐式情感, 使用 FinBERT 预训练模型引入金融领域知识^[33], 将上下文语句 C 表示为词向量. 在此基础上, 使用 Bi-LSTM 模型获取上下文语义信息^[34], 使用 self-attention 注意力机制获取关键语义

信息^[35].

1) 词语嵌入层. FinBERT 预训练模型是在海量的金融财经新闻、上市公司公告和金融类百科词条的基础上训练得到的, 包含大量的金融领域专业知识, 可以有效解决部分 MD&A 文本语句因缺乏情感关键词而难以判断情感的问题. 将该预训练模型引入 SC-ISA 模型, 将语句 S 和上下文语句 C 表示为向量 $\mathbf{S}_c = [w_{c1}, w_{c2}, \dots, w_{cm}]$.

2) 上下文语义特征表示. 使用 Bi-LSTM 获取文本的上下文语义信息, 使用 self-attention 机制获取关键语义信息. 给定输入词向量 $\mathbf{S}_c = [w_{c1}, w_{c2}, \dots, w_{cm}]$, 可以获得输出

$$\vec{p}_{ct} = \overrightarrow{\text{LSTM}}(\vec{p}_{t-1}, w_{ct}), \quad (3)$$

$$\overleftarrow{p}_{ct} = \overleftarrow{\text{LSTM}}(\overleftarrow{p}_{t-1}, w_{ct}). \quad (4)$$

其中 $\overrightarrow{\text{LSTM}}$ 和 $\overleftarrow{\text{LSTM}}$ 分别为 Bi-LSTM 的正向模块和逆向模块.

词语 w 经过 Bi-LSTM 可以获得上下文语义信息的表示 $\mathbf{h}_{ct} = [\vec{p}_{ct}, \overleftarrow{p}_{ct}]$, $[\cdot, \cdot]$ 表示向量之间的拼接. 通过 self-attention 机制可以得到上下文语义向量 $\text{Att}(\mathbf{Q}, \mathbf{K}, \mathbf{V})$ 为

$$\text{Att}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}, \quad (5)$$

其中 $\mathbf{Q} = w_{ct}\mathbf{W}_q$, $\mathbf{K} = w_{ct}\mathbf{W}_k$, $\mathbf{V} = w_{ct}\mathbf{W}_v$, \mathbf{W}_q 、 \mathbf{W}_k 和 \mathbf{W}_v 均为权重矩阵, d_k 表示词语向量维度, softmax 为激活函数. 最终, 句子 S 的上下文语义特征表示为矩阵 \mathbf{H}_c .

3.1.3 隐式情感判断

使用 BiAffine 操作充分融合 MD&A 语句的结构特征和上下文语义特征, 分别得到语句的结构特征向量 \mathbf{H}'_s 和上下文语义特征向量 \mathbf{H}'_c

$$\mathbf{H}'_s = \text{softmax}(\mathbf{H}_s\mathbf{W}_1\mathbf{H}_c^T)\mathbf{H}_c, \quad (6)$$

$$\mathbf{H}'_c = \text{softmax}(\mathbf{H}_c\mathbf{W}_2\mathbf{H}_s^T)\mathbf{H}_s, \quad (7)$$

其中 \mathbf{W}_1 和 \mathbf{W}_2 为系数矩阵. 将 \mathbf{H}'_s 和 \mathbf{H}'_c 进行拼接, 使用 softmax 函数识别语句 T 的隐式情感

$$\hat{y}_T = \text{softmax}([\mathbf{H}'_s, \mathbf{H}'_c]). \quad (8)$$

交叉熵(cross entropy)常用于文本分类任务中评估分类模型的输出值概率分布与真实值分布的差异. 采用交叉熵损失函数作为优化方向, 分类模型输出值和实际值交叉熵值越小, 两个数值的概率分布就越接近, 分类模型预测正确情感类别的概率越高, 交叉熵的计算公式为

$$L = - \sum_{y \in D} y \ln \hat{y}, \quad (9)$$

其中 D 为训练集, y 是真实值的独热编码, \hat{y} 是预测值的情感概率分布.

3.2 基于隐式情感的企业财务困境预测

3.2.1 财务指标选择

企业的财务状况可以揭示企业的盈利能力, 成长能力、资本结构和产品竞争力等, 进而预测企业的财务困境状态. 现有研究所使用的财务指标体系侧重于企业当前经营状况的描述, 对企业未来的经营状况和企业经营规范化程度重视不足^[23]. 在现有企业财务困境预测指标的基础上, 使用发展能力指标、现金流量指标和股权结构指标刻画企业的成长能力、经营健康程度和法人治理情况, 最后共得到 8 个财务一级指标和 33 个财务二级指标, 如表 2 所示. 一级财务指标包括偿债能力、发展能力、运营能力、盈利能力和现金流量等, 二级财务指标包括资产负债率、流动比率、负债权益比、存货周转率、债务保障率、销售净利润率和 Herfindah 指数等.

表2 企业财务指标
Table 2 Financial indicators of enterprises

类别	财务指标	指标代码	计算方法
偿债能力	资产负债率	Asset-liability ratio	总负债/总资产
	流动比率	Liquidity ratio	总负债/总资产
	负债权益比	Debt-equity ratio	总负债/股东权益总额
	速动比率	Liquidity ratio	(流动资产-存货)/流动负债
	已获利息倍数	Times interest earned	息税前利润/利息费用
发展能力	有形资产负债率	Tangible assets to liabilities ratio	负债总额/(资产总额-无形资产总额-商誉净额)
	营业收入增长率	Revenue growth rate	(本年营业收入-去年营业收入)/去年营业收入
	净利润增长率	Net profit growth rate	本年净利润/去年净利润
	资本积累率	Capital accumulation rate	本年所有者权益增长额/年初所有者权益
	总资产增长率	Total assets growth rate	本年总资产/去年总资产
运营能力	营业总成本增长率	Growth rate of operating costs	(本年营业成本-去年营业成本)/去年营业成本
	存货周转率	Inventory turnover rate	营业成本/存货期末余额
	流动资产周转率	Current assets turnover ratio	营业收入/流动资产期末余额
	非流动资产周转率	Non-current asset turnover ratio	营业收入/(非流动资产合计)期末余额
	固定资产周转率	Fixed asset turnover	销售收入/固定资产
盈利能力	应收账款周转率	Accounts receivable turnover	主营业务收入/应收账款
	资本密集度	Capital intensity	总资产/营业收入
	销售净利润率	Net profit margin	净利润/销售收入
	资产息税前利润率	Earnings before interest and tax	息税前利润/总资产
	资产收益率	Return on total assets	净利润/总资产
现金流量	净资产收益率	Return on equity	净利润/净资产
	营业利润率	Return on equity	营业利润/营业收入
	营业净利率	Operating net profit margin	净利润/营业收入
	销售净利润	Net profit from sales	净利润/销售收入
	债务保障率	Debt coverage ratio	经营活动现金流量/负债总额
股权结构	营业收入现金比率	Sales cash ratio	经营活动净流量/营业收入
	全部资产现金回收率	All cash flow return on investment	经营活动现金净流量/期末资产总额
结构稳健	Z指数	Z-index	第一大股东持股数/第二大股东持股数
	Herfindah指数	Herfindah-index	公司前三位大股东持股比例平方之和
单股指标	固定资产比率	Fixed asset ratio	固定资产/总资产
	留存收益比率	Fixed asset ratio	留存收益/总资产
单股指标	每股收益	Earnings per share	净利润/总股数
	每股净资产	Debt coverage ratio	所有者权益总额/实收资本

3.2.2 情感指标计算

MD&A 包含管理层对企业未来发展前景的预期, 具有一定的情感倾向, 可以用于企业财务困境预测^[36]. 从 MD&A 文本提取显式情感信息和隐式情感信息, 作为机器学习算法的特征变量.

1) 显式情感指标计算. 显式情感值的计算方法包括机器学习法和情感词典法, 情感词典操作简单和适用性强的优点^[37]. 以姜富伟等^[38]和Li等^[39]的金融情感词典为基础, 构建包括 4 445 个正向情感词和 7 378 个负向情感词的金融情感词典. MD&A 显式情感值 S_e 使用文本正向情感词与负向情感词的相对差异表示, 计算方法为

$$S_e = \frac{P_w - N_w}{P_w + N_w}, \quad (10)$$

其中 P_w 表示 MD&A 中正向情感词的数量, N_w 表示 MD&A 中负向情感词的数量. 如果 S_e 为正值, 表示管理层对企业的发展持有积极的态度; 否则, 管理层对企业的发展持有消极的态度.

2) 隐式情感指标计算. MD&A 隐式情感值 S_i 使用文本中正向隐式情感的句子与负向隐式情感的相对差异表示, 计算方法为

$$S_i = \frac{P_t - N_t}{P_t + N_t}, \quad (11)$$

其中 P_t 表示 MD&A 中正向隐式情感句子的数量, N_t 表示 MD&A 中负向隐式情感的数量. 如果 S_i 为正值, 表示管理层对企业的发展持有积极的态度; 否则, 管理层对企业的发展持有消极的态度.

4 实证分析

4.1 样本选择与数据来源

以 2006 年至 2022 年非金融行业 A 股上市公司为实证研究对象, 财务数据来源于 Wind 数据库和 CSMAR 数据库, 财务报告来源于巨潮资讯网、上交所和深交所的公告. 上市公司根据经营状况分为 ST 公司(包括 ST、*ST 和 SST 等)和非 ST 公司. 如果企业符合最近一个会计年度经审计的净利润为负值、年度报告被出具无法表示意见或者否定意见等条件, 则企业被证券交易所标记 ST. 以是否被 ST 标记作为企业财务困境的标准: ST 企业为财务困境企业, 作为预测的负样本, 记为 0; 非 ST 公司为正常企业, 作为预测的正样本, 记为 1.

为了构建隐式情感识别的数据集, 首先使用 python 爬虫爬取上市公司年报 pdf 文件, 使用 pdfplumber 从年报中提取 MD&A 文本. 然后, 随机选择 127 家上市公司年报 MD&A 文本数据, 使用情感关键词进行逆向匹配获得隐式情感语句. 通过双人多轮标注的方法, 共标注不包含情感词的句子 1 500 条, 其中正向情感、负向情感和中性情感均为 500 条.

4.2 数据预处理与评价指标选择

1) 财务指标预处理

财务指标预处理包括缺失值的填补、异常值的处理和数据的标准化. 企业财务数据中存在缺失值, 如果样本缺失的比例高于 5%, 则舍弃该指标; 否则, 使用样本中该指标的均值进行填补. 如果样本与均值的差异超过三倍标准差, 则该指标为异常值, 使用样本均值正(负)三倍标准差进行填补. 由于财务指标数据的量纲和取值范围差异较大, 对机器学习算法的预测结果影响较大, 使用 Z-score 方法进行数据标准化, 计算方法为

$$z = \frac{x - \mu}{s}, \quad (12)$$

其中 x 为样本指标值, μ 为样本指标均值, s 为样本指标标准差.

2) 文本数据预处理

使用 pdfplumber 从企业年报中提取 MD&A 文本数据, 对数据的预处理操作包括: 去掉长度小于 20 的语句; 构建包含会计术语、金融术语和法律术语等的保留词典; 构建停用词典; 使用 LTP 工具对 MD&A 文本进行切词, 去掉长度小于 10 的语句.

3) SMOTE 过采样

由于处于财务困境的上市企业数量较少, 收集的正常公司与 ST 公司的比例为 2 007: 478, 样本数据存在一定的不均衡. 为了规避样本不均衡对模型预测效果的影响, 使用 SMOTE(synthetic minority over-sampling technique)过采样技术合成 ST 公司样本来增加其在数据集中的比重. 最终, 样本中正常公司与 ST 公司的数量均为 2 007 家.

4) 评估指标

使用分类任务常用的评价指标精准率、召回率和 F_1 值对隐式情感识别和企业财务困境预测结果进行评估. 模型的预测类别分为正类(Positive)和负类(Negative), 正类是指模型预测为正常公司的样本, 负类是指

模型预测为 ST 公司的样本. 精准率 P 、召回率 R 和 F_1 值定义为

$$P = \frac{TP}{TP + FP}, \quad (13)$$

$$R = \frac{TP}{TP + FN}, \quad (14)$$

$$F_1 = \frac{2PR}{P + R}, \quad (15)$$

其中 TP 为正类预测为正类的数量, FN 为正类预测为负类的数量, FP 为负类预测为正类的数量, TN 为负类被预测为负类的数量.

4.3 隐式情感识别结果分析

使用标注隐式情感的 MD&A 文本数据对 SC-ISA 模型的性能进行评估, 评价指标包括精准率 P 、召回率 R 和 F_1 值. 对比模型包括 RoBERTa 预训练模型、Bi-LSTM+attention 模型、GACNN 模型、Llama3.1 和 GPT-4o 等.

1) RoBERTa. 该方法使用 RoBERTa 预训练模型将包含隐式情感的语句进行向量表示, 使用微调的方法实现隐式情感文本分类任务.

2) Bi-LSTM+attention^[40]. 该方法使用 Bi-LSTM 获取上下文语义信息, 使用 attention 机制获取关键语义信息, 从而实现隐式情感分类.

3) GACNN^[41]. 该方法构建包含词语和文档的异构文本图, 使用图注意力机制提取上下文关键语义信息, 使用正交约束的多头注意力机制进行隐式情感识别.

4) KG-MPOA^[11]. 该方法提出嵌入多极正交注意力机制的知识图谱模型, 通过将外部情感知识与上下文语义信息融合实现隐式情感分类.

5) Llama3.1^[9]. 该方法是 Meta 公司在 2024 年 7 月 23 日推出的大型生成式预训练语言模型, 包含 80 亿参数和 4 050 亿个参数两个版本.

6) GPT-4o^[10]. 该方法是 OpenAI 推出的基于 Transformer 的大型生成式预训练语言模型, 可以对音频、视觉和文本进行推理.

MD&A 文本隐式情感识别比对试验结果如表 3 所示. 预训练模型 RoBERTa 的泛化能力和鲁棒性优异, 在文本分类任务中表现良好, 是本研究的基线模型, 精准率、召回率和 F_1 指标分别为 0.721 5、0.708 3 和 0.714 8. Bi-LSTM 可以获得文本的上下文语义信息, 通过与注意力机制的结合提取文本中的关键语义信息, 但是无法解决情感特征长距离依赖和金融用语专业性问题. GACNN 模型通过构建包含词语和文档的图模型, 可以解决情感特征长距离依赖的问题. KG-MPOA 通过引入外部知识图谱提高金融专业术语的表示能力, 隐式情感识别效果得到提升. 然而, 该方法受到外部知识图谱的规模的限制, 对金融专业术语的表示能力和识别能力有限. 大语言模型是基于海量文本数据训练而成的深度学习模型, 不仅包含大量的领域专业知识, 而且具有强大的逻辑推理能力. Llama3.1 和 GPT-4o 的隐式情感识别能力均接近或者优于 GACNN 和 KG-MPOA 等模型. 与 RoBERTa 等预训练模型相比, 大语言模型的训练语料的规模更大, 算法更加先进, 在本研究的 MD&A 文本隐式情感识别实验中表现更好. 提出的隐式情感识别模型 SC-ISA 使用 FinBERT 引入金融领域专业知识和情感信息, 通过依存句法分析和图注意力机制识别关键结构信息, 通过 Bi-LSTM 和注意力机制获取关键上下文语义信息, 精准率、召回率和 F_1 指标分别为 0.786 4、0.793 3 和 0.789 8, 均优于对比模型.

使用消融实验验证 FinBERT 金融预训练模型、语句结构特征和上下文语义特征的在 SC-ISA 模型中的有效性, 实验结果如表 4 所示. 在基于 FinBERT 金融预训练模型的于隐式情感识别实验中, 精准率、召回率和 F_1 值分别为 0.742 1、0.729 5 和 0.735 7. 在 FinBERT 金融预训练模型的基础上分别增加语句结构

特征和上下文语义特征, F_1 值分别提升 0.022 3 和 0.032 4, 表明通过依存句法分析构建神经网络和通过 Bi-LSTM 提取上下文语义信息, 均可以从 MD&A 文本中提取与文本情感有关的信息, 从而提升隐式情感识别的效果。

表 3 隐式情感识别结果分析
Table 3 Comparisons of implicit sentiment analysis

模型名称	精准率 P	召回率 R	F_1 值
RoBERTa	0.721 5	0.708 3	0.714 8
BiLSTM+attention	0.682 9	0.654 2	0.668 2
GACNN	0.765 2	0.735 5	0.750 1
KG-MPOA	0.752 3	0.741 8	0.747 0
Llama3.1	0.746 5	0.752 8	0.749 6
GPT-4o	0.768 3	0.762 5	0.765 4
SC-ISA (本模型)	0.786 4	0.793 3	0.789 8

表 4 消融实验结果分析
Table 4 Results of ablation experiment

模型名称	精准率 P	召回率 R	F_1 值
FinBERT	0.742 1	0.729 5	0.735 7
FinBERT+语句结构特征	0.763 1	0.752 9	0.758 0
FinBERT+上下文语义特征	0.775 3	0.761 1	0.768 1
SC-ISA (本模型)	0.786 4	0.793 3	0.789 8

4.4 企业财务困境预测结果分析

选择 2 485 家国内 A 股上市公司作为实证研究的样本, 样本特征包括资产负债率、流动比率、存货周转率、营业利润率、显式情感和隐式情感等 35 个财务指标和文本特征指标。样本数据划分为训练集和测试集, 采用五折交叉检验进行模型训练。对比模型包括决策树、SVM、MLP 和 LightGBM 等模型, 参数设置如表 5 所示。

表 5 模型参数设置
Table 5 Parameter settings of machine learning models

模型名称	参数设置
决策树	最大深度 max_depth = 52, 最大叶子节点数 max_leaf_nodes = 68
SVM	惩罚系数 $C = 381$, 核函数宽度 $\gamma = 0.000 1$
MLP	神经网络层数 n_layers = 2, 隐藏层大小 hidden_layer_sizes = 130
Catboost	决策树最大深度 depth = 4, 学习率 learning_rate = 0.03, 叶子节点权重的正则化系数 l2_leaf_reg = 12

对比实验分别基于财务指标数据、财务指标+显式情感数据和财务指标+显式情感+隐式情感数据, 评估指标包括精准率、召回率和 F_1 值, 实验结果如表 6 所示。对于所有的企业财务困境预测模型, 增加显式情感和隐式情感指标均可以提升模型预测的效果。SVM 作为经典的机器学习算法, 通过将低维数据映射到高维空间, 财务困境预测的效果较决策树和 MLP 算法均有一定的提升。Catboost 作为集成模型的代表, 通过集成多个决策树模型, 使用投票机制确定样本分类, 较决策树和 SVM 等模型具有更好的预测效果。通过增加显式情感特征和隐式情感特征, 该模型的 F_1 值从 0.865 5 分别提升至提升 0.888 4 和 0.902 5。

为了对研究结果的稳健性进行检验, 选择机械设备行业、医药生物行业、电子行业和基础化工等行业的企业共 1 592 家进行企业财务困境预测, 实验结果如表 7 所示。在决策树、SVM、MLP 和 Catboost 模型中, 增加显式情感特征和隐式情感特征均可以提升模型预测的效果, 表明所提出的企业财务困境预测模型的稳定性较好。

表 6 企业财务困境预测结果分析
Table 6 Results of corporate financial distress prediction

预测指标	模型名称	精准率 P	召回率 R	F_1 值
财务指标	决策树	0.738 9	0.756 6	0.747 6
	SVM	0.831 3	0.832 5	0.831 9
	MLP	0.802 7	0.819 6	0.811 1
	Catboost	0.849 2	0.882 5	0.865 5
财务指标+显式情感	决策树	0.751 5	0.794 5	0.772 4
	SVM	0.862 7	0.851 6	0.857 1
	MLP	0.820 5	0.863 1	0.841 3
	Catboost	0.870 4	0.907 2	0.888 4
财务指标+显式情感 +隐式情感	决策树	0.759 2	0.821 5	0.789 1
	SVM	0.876 8	0.862 5	0.869 6
	MLP	0.834 1	0.885 9	0.859 2
	Catboost	0.883 1	0.922 7	0.902 5

表 7 企业财务困境预测稳健性分析
Table 7 Robustness of corporate financial distress prediction

预测指标	模型名称	精准率 P	召回率 R	F_1 值
财务指标	决策树	0.748 5	0.743 9	0.746 2
	SVM	0.827 7	0.834 2	0.830 9
	MLP	0.810 4	0.805 5	0.807 9
	Catboost	0.877 3	0.862 5	0.869 8
财务指标+显式情感	决策树	0.782 9	0.791 1	0.787 0
	SVM	0.859 3	0.843 6	0.851 4
	MLP	0.847 2	0.839 5	0.843 3
	Catboost	0.891 3	0.883 7	0.887 5
财务指标+显式情感 +隐式情感	决策树	0.792 2	0.814 5	0.803 2
	SVM	0.874 1	0.860 8	0.867 4
	MLP	0.861 9	0.852 5	0.857 2
	Catboost	0.912 7	0.897 3	0.904 9

4.5 企业财务困境预测模型可解释性分析

使用 SHAP 方法对企业财务困境预测模型的可解释性进行分析, 通过计算模型中特征的 Shapley 值得到特征对模型预测结果的贡献度, 从而判断模型中特征的重要程度, 如图 2 所示. 营业利润率、留存收益比率、HerFindah 指数和资本密集度等是预测模型中最重要的特征, Shapley 值分别为 0.78、0.73、0.65 和 0.64. 显式情感值和隐式情感值反映了管理层对企业经营现状和前景的态度, 对企业财务困境预测结果的影响也较大, Shapley 值分别为 0.28 和 0.20.

5 结束语

为了探索 MD&A 文本识别隐式情感特征对企业财务困境预测的影响, 基于图神经网络、Bi-LSTM 和 FinBERT 金融大语言模型提出 MD&A 文本隐式情感识别模型 SC-ISA, 通过提取句子结构特征、上下文语义特征和引入外部金融情感信息解决隐式情感识别中情感特征缺乏的问题, 使用 2007 年至 2022 年国内上市公司的财务数据和 MD&A 文本数据评估 SC-ISA 模型识别 MD&A 文本识别隐式情感特征的效果. 在此基础上, 将隐式情感作为特征变量, 使用决策树、SVM、MLP 和 Catboost 等机器学习算法进行企业财务困境预测. 实验结果表明, MD&A 文本隐式情感特征均可以提升企业财务困境预测模型的精准率、召回率和 F_1 值. 隐式情感特征作为 MD&A 文本情感特征的重要组成部分, 可以更好的反映企业管理层对未来市场环境、经营风险和业务发展的预期, 为提升企业财务困境预测能力提供了新的思路. 后续的研究考虑引入

社交媒体数据和多模态业务数据, 使用大语言模型提升 MD&A 文本隐式情感的识别能力.

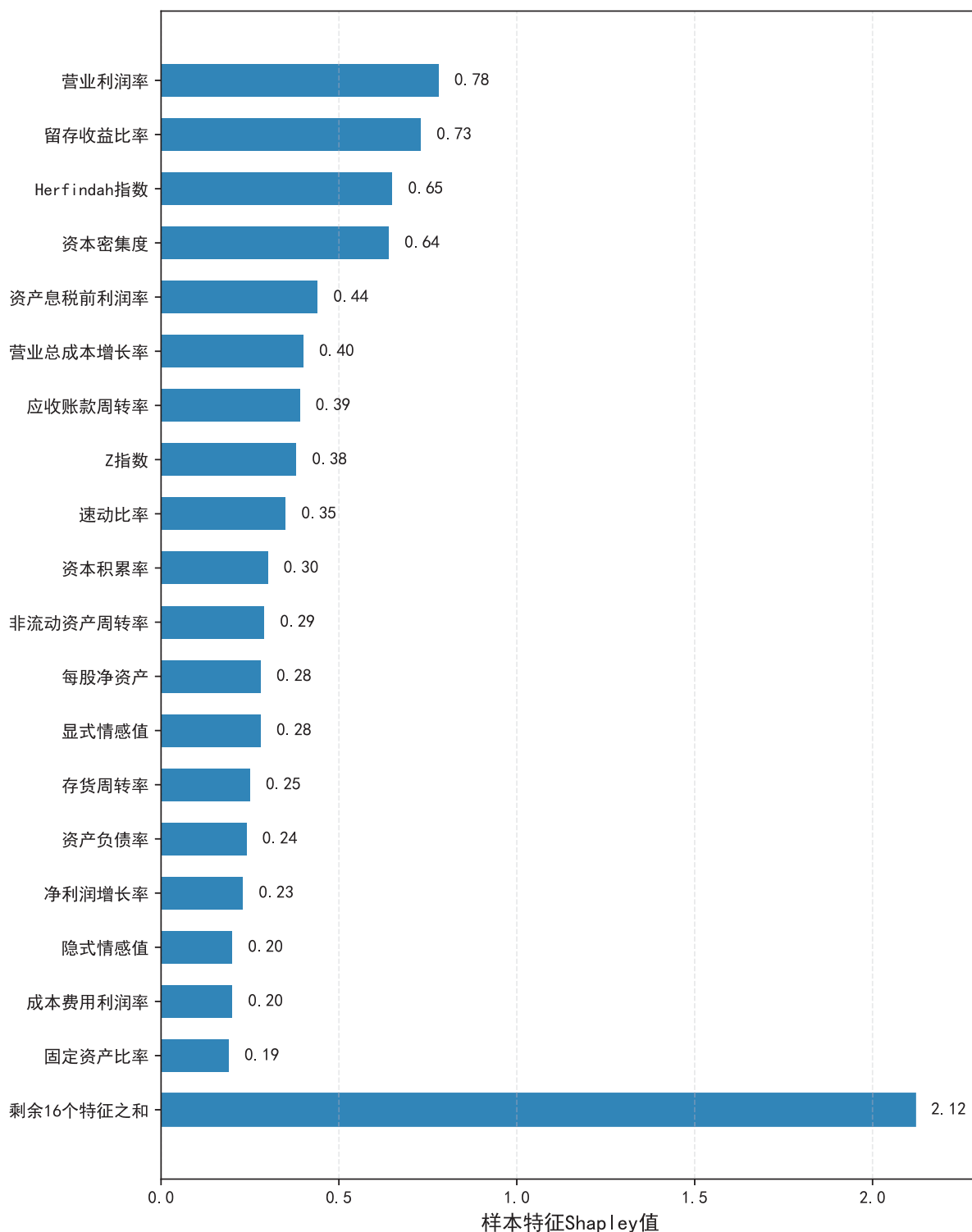


图 2 企业财务困境预测模型特征重要性

Fig. 2 Importance of features in financial distress prediction

参考文献:

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

- [2] 江俊毅, 蒋洪迅. 险境淘金: 基于并购重组事件挖掘的财务脱困预测. 系统工程学报, 2022, 37(2): 161–177.
Jiang J Y, Jiang H X. Gold panning: Financial relief forecasting of listed enterprises based on the mining of mergers and acquisitions. *Journal of Systems Engineering*, 2022, 37(2): 161–177. (in Chinese)
- [3] Zhao J, Ouenniche J, De S J. Survey, classification and critical analysis of the literature on corporate bankruptcy and financial distress prediction. *Machine Learning with Applications*, 2024, 15: 100527.
- [4] Liu J, Li C, Ouyang P, Liu J, Wu C. Interpreting the prediction results of the tree-based gradient boosting models for financial distress prediction with an explainable machine learning approach. *Journal of Forecasting*, 2023, 42(5): 1112–1137.
- [5] Hajek P, Munk M. Corporate financial distress prediction using the risk-related information content of annual reports. *Information Processing & Management*, 2024, 61(5): 103820.
- [6] 魏瑾瑞, 王若彤, 王 晗. 虚假评论网络的识别与特征发现. 系统工程学报, 2024, 39(5): 665–675.
Wei J R, Wang R T, Wang H. False comments network identification and feature discovery. *Journal of Systems Engineering*, 2024, 39(5): 665–675. (in Chinese)
- [7] Huang B, Yao X, Luo Y, Li J. Improving financial distress prediction using textual sentiment of annual reports. *Annals of Operations Research*, 2023, 330(1): 457–484.
- [8] Wei J, Liao J, Yang Z, Wang S, Zhao Q. BiLSTM with multi-polarity orthogonal attention for implicit sentiment analysis. *Neuro-computing*, 2020, 383: 165–173.
- [9] Agarwal P, Gupta A. Strategic business insights through enhanced financial sentiment analysis: A fine-tuned Llama 2 approach // *Proceedings of 2024 International Conference on Inventive Computation Technologies*, New York: IEEE Press, 2024: 1446–1453.
- [10] Leippold M. Sentiment spin: Attacking financial sentiment with GPT-3. *Finance Research Letters*, 2023, 55: 103957.
- [11] 武 杰, 那日萨, 郑雅璇. 基于在线评论的电商助农消费者抱怨分析. 系统工程学报, 2025, 40(1): 15–32.
Wu J, Narisa, Zheng Y X. Analysis of consumer complaints of the e-commerce to help farmers based on online reviews. *Journal of Systems Engineering*, 2025, 40(1): 15–32. (in Chinese)
- [12] Fang Z, Zhang Q, Tang X, Wang A, Baron C. An implicit opinion analysis model based on feature-based implicit opinion patterns. *Artificial Intelligence Review*, 2020, 53: 4547–4574.
- [13] 赵容梅, 熊 熙, 琚生根, 等. 基于混合神经网络的中文隐式情感分析. 四川大学学报(自然科学版), 2020, 57(2): 264–270.
Zhao R M, Xiong X, Ju S G, et al. Implicit sentiment analysis for Chinese texts based on a hybrid neural network. *Journal of Sichuan University (Natural Science Edition)*, 2020, 57(2): 264–270. (in Chinese)
- [14] 张玉莹, 朱广丽, 谈光璞. 基于情感增强和语义依存的金融隐式情感分析模型. 计算机工程与科学, 2024, 46(6): 1112–1120.
Zhang Y Y, Zhu G L, Tan G P. A financial implicit sentiment analysis model based on sentiment enhancement and semantic dependency. *Computer Engineering & Science*, 2024, 46(6): 1112–1120. (in Chinese)
- [15] 杨善良, 常 征. 基于图注意力神经网络的中文隐式情感分析. 计算机工程与应用, 2021, 57(24): 161–167.
Yang S L, Chang Z. Chinese implicit sentiment analysis based on graph attention neural network. *Computer Engineering and Applications*, 2021, 57(24): 161–167. (in Chinese)
- [16] 张 军, 张 丽, 沈凡凡, 等. RoBERTa 融合 BiLSTM 及注意力机制的隐式情感分析. 计算机工程与应用, 2022, 58(23): 142–150.
Zhang J, Zhang L, Shen F F, et al. Implicit sentiment analysis based on RoBERTa fused with BiLSTM and attention mechanism. *Computer Engineering and Applications*, 2022, 58(23): 142–150. (in Chinese)
- [17] Yin W, Shang L. Contextbert: Enhanced implicit sentiment analysis using implicit-sentiment-query attention // *Proceedings of 2022 International Joint Conference on Neural Networks*. New York: IEEE Press, 2022: 1–8.
- [18] Chen X, Wang S, Li X, Hai Z, Li Y, et al. Identifying implicit emotions via hierarchical structure and rhetorical correlation. *International Journal of Machine Learning and Cybernetics*, 2023, 14(11): 3753–3764.
- [19] Fei H, Li B, Liu Q, Bing L, Li F, et al. Reasoning implicit sentiment with chain-of-thought prompting // *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*. Massachusetts: MIT Press, 2023: 1171–1182.
- [20] Zhao Q, Xu W, Ji Y. Predicting financial distress of Chinese listed companies using machine learning: To what extent does textual disclosure matter. *International Review of Financial Analysis*, 2023, 89: 102770.
- [21] Sun J, Li J, Fujita H, Ai W. Multiclass financial distress prediction based on one-versus-one decomposition integrated with improved decision-directed acyclic graph. *Journal of Forecasting*, 2023, 42(5): 1167–1186.
- [22] Che W, Wang Z, Jiang C, Abedin M Z. Predicting financial distress using multimodal data: An attentive and regularized deep learning method. *Information Processing & Management*, 2024, 61(4): 103703.
- [23] Kamaluddin A, Ishak N, Mohammed N F. Financial distress prediction through cash flow ratios analysis. *International Journal of Financial Research*, 2019, 10(3): 63–76.

- [24] Fernández-Gámez M Á, Soria J A C, Santos J A C, et al. European country heterogeneity in financial distress prediction: An empirical analysis with macroeconomic and regulatory factors. *Economic Modelling*, 2020, 88: 398–407.
- [25] Liang D, Tsai C F, Wu H T. The effect of feature selection on financial distress prediction. *Knowledge-Based Systems*, 2015, 73: 289–297.
- [26] Citterio A, King T. The role of Environmental, Social, and Governance (ESG) in predicting bank financial distress. *Finance Research Letters*, 2023, 51: 103411.
- [27] Wang G, Chen G, Chu Y. A new random subspace method incorporating sentiment and textual information for financial distress prediction. *Electronic Commerce Research and Applications*, 2018, 29: 30–49.
- [28] Pereira J M, Basto M, Da Silva A F. The logistic lasso and ridge regression in predicting corporate failure. *Procedia Economics and Finance*, 2016, 39: 634–641.
- [29] Almamy J, Aston J, Ngwa L N. An evaluation of Altman’s Z-score using cash flow ratio to predict corporate failure amid the recent financial crisis: Evidence from the UK. *Journal of Corporate Finance*, 2016, 36: 278–285.
- [30] Wu D, Ma X, Olson D L. Financial distress prediction using integrated Z-score and multilayer perceptron neural networks. *Decision Support Systems*, 2022, 159: 113814.
- [31] Chen Z, Chen W, Shi Y. Ensemble learning with label proportions for bankruptcy prediction. *Expert Systems with Applications*, 2020, 146: 113155.
- [32] Shen F, Liu Y, Wang R, Zhou W. A dynamic financial distress forecast model with multiple forecast results under unbalanced data environment. *Knowledge-Based Systems*, 2020, 192: 105365.
- [33] Huang A H, Wang H, Yang Y. FinBERT: A large language model for extracting information from financial text. *Contemporary Accounting Research*, 2023, 40(2): 806–841.
- [34] Arbane M, Benlamri R, Brik Y, Alahmar A D. Social media-based COVID-19 sentiment classification model using Bi-LSTM. *Expert Systems with Applications*, 2023, 212: 118710.
- [35] Ouyang J, Xuan C, Wang B, Yang Z. Aspect-based sentiment classification with aspect-specific hypergraph attention networks. *Expert Systems with Applications*, 2024, 248: 123412.
- [36] Wang G, Chen G, Chu Y. A new random subspace method incorporating sentiment and textual information for financial distress prediction. *Electronic Commerce Research and Applications*, 2018, 29: 30–49.
- [37] Frankel R, Jennings J, Lee J. Disclosure sentiment: Machine learning vs. dictionary methods. *Management Science*, 2022, 68(7): 5514–5532.
- [38] 姜富伟, 孟令超, 唐国豪. 媒体文本情绪与股票回报预测. *经济学(季刊)*, 2021, 21(4): 1323–1344.
Jiang F W, Meng L C, Tang G H. Media textual sentiment and Chinese stock return predictability. *China Economic Quarterly*, 2021, 21(4): 1323–1344. (in Chinese)
- [39] Li S, Shi W, Wang J, Zhou H. A deep learning-based approach to constructing a domain sentiment lexicon: A case study in financial distress prediction. *Information Processing & Management*, 2021, 58(5): 102673.
- [40] Yang S, Xing L, Li Y, Chang Z. Implicit sentiment analysis based on graph attention neural network. *Engineering Reports*, 2022, 4(1): 1–16.

作者简介:

闫志华 (1987—), 男, 山西朔州人, 博士, 讲师, 研究方向: 数据挖掘与决策支持, Email: zhyan@amss.ac.cn;

黄晓辉 (1995—), 男, 广东珠海人, 博士, 博士后, 研究方向: E-mail: 机器学习, Email: huangxiaohui@amss.ac.cn;

唐锡晋 (1967—), 女, 江苏无锡人, 博士, 研究员, 研究方向: 综合集成、决策支持和知识科学, Email: xjtang@iss.ac.cn;

闫绪娴 (1978—), 女, 山西朔州人, 博士, 教授, 研究方向: 应急管理, Email: yanxux@163.com.