计 算 机 研 究 与 发 展
Journal of Computer Research and Development
https://crad.ict.ac.cn
63(10): 2752−2766, 2026

基于 GCN-Stacking 协同增强的多模态金融舆情负向情感强度识别

李婷婷[1,2] 孙守强[3] 李青青[4] 万 玲[5]
[1](重庆理工大学经济金融学院 重庆 400054)
[2](金融风险智能预警及现代治理实验室(重庆理工大学) 重庆 400054)
[3](西安电子科技大学经济与管理学院 西安 710126)
[4](武汉大学信息管理学院 武汉 430072)
[5](湖南农业大学信息与智能科学技术学院 长沙 410128)
(litingting@cqut.edu.cn)

GCN-Stacking Collaborative Enhancement-Based Multimodal Financial Public Opinion Negative Sentiment Intensity Recognition

Li Tingting[1,2], Sun Shouqiang[3], Li Qingqing[4], and Wan Ling[5]
[1](*School of Economics and Finance*, *Chongqing University of Technology*, *Chongqing* 400054)
[2](*Lab of Financial Risk Intelligent Early Warning and Modern Governance* (*Chongqing University of Technology*), *Chongqing* 400054)
[3](*School of Economics and Management*, *Xidian University*, *Xi'an* 710126)
[4](*School of Information Management*, *Wuhan University*, *Wuhan* 430072)
[5](*School of Information and Intelligence Science and Technology*, *Hunan Agricultural University*, *Changsha* 410128)

Abstract The recognition of negative sentiment intensity in financial public opinion on social media is of great significance for financial risk prevention and market stability. Given the prevalence of multimodal information combining text and images in financial sentiment expression, and the need for further exploration of structural correlations between modalities, we propose a multimodal negative sentiment intensity recognition model based on the collaborative enhancement of graph convolutional networks and Stacking ensemble learning. The model first employs the CLIP model to extract semantic features from text and images. Subsequently, a multimodal graph structure is constructed and graph convolutional networks are introduced to achieve structural association embedding of features, thereby enhancing the ability to capture potential inter-modal correlations. Finally, a two-layer Stacking ensemble learning framework is built to optimize the decision performance of sentiment intensity recognition by integrating the prediction results of multiple diverse base classifiers. Experimental results show that under the optimal configuration, the model achieves an $F1$-score of 83.57% and an AUC of 95.05%, outperforming the optimal single-text model, the optimal single-image model, and the original CLIP multimodal model by 2.48, 5.27, and 1.87 percentage points in $F1$-score, respectively. These results verify the effectiveness of the proposed method in mining modal correlations and improving recognition performance, providing technical support for intelligent financial public opinion analysis and financial risk monitoring.

Key words financial public opinion; multimodal information; negative sentiment recognition; graph convolutional network; Stacking ensemble

收稿日期: 2025-12-24; **修回日期**: 2026-05-15

基金项目: 重庆市教委科学技术研究项目(KJQN202501134); 重庆市社会科学规划项目(2025BS031); 重庆理工大学科研基金项目(2024ZDR032)
This work was supported by the Science and Technology Research Program of Chongqing Municipal Education Commission (KJQN202501134), the Chongqing Social Science Planning Project (2025BS031), and the Scientific Research Foundation of Chongqing University of Technology (2024ZDR032).

摘　要　社交媒体中金融舆情的负向情感强度识别对金融风险防控与市场稳定具有重要意义。针对图文多模态信息在金融情感表达中的普遍性，以及模态间结构关联有待深入挖掘的问题，提出一种基于图卷积网络与 Stacking 集成学习协同增强的多模态负向情感强度识别模型。该模型首先利用 CLIP 模型提取文本与图像的语义特征；随后，通过构建多模态图结构并引入图卷积网络，实现特征的结构化关联嵌入，以增强对模态间潜在关联的捕捉能力；最后，构建双层 Stacking 集成学习框架，通过融合多个差异化基分类器的预测结果，优化情感强度识别的决策性能。实验结果表明，在最优配置下，模型的 $F1$ 分数达 83.57%，AUC 达 95.05%，相较于单一文本最优模型、单一图像最优模型及原始 CLIP 多模态模型，$F1$ 分数分别提升 2.48 个百分点，5.27 个百分点，1.87 个百分点。实验结果验证了所提方法在挖掘模态关联与提升识别性能方面的有效性，可为金融舆情智能化分析与金融风险监测提供技术支撑。

关键词　金融舆情；多模态信息；负向情感识别；图卷积网络；Stacking 集成

中图法分类号　TP39

DOI: 10.7544/issn1000-1239.202550930　　**CSTR:** 32373.14.issn1000-1239.202550930

在数字经济与信息社会背景下，金融舆情已成为影响投资者信心和市场波动的重要因素。尤其是负面舆情，其传播速度快、覆盖范围广，极易引发市场恐慌、加剧金融风险乃至形成系统性危机[1]。在金融舆情分析中，负向情感不仅需要被识别，更应关注其强度，因为情感强度直接决定了舆情可能带来的市场冲击。例如，轻度负面舆情可能仅反映个体不满，但强烈负面舆情则可能迅速引发群体共鸣并放大风险[2]。因此，识别金融舆情负向情感强度，不仅对金融监管部门实现前瞻性风险治理具有理论意义，也为金融机构优化决策、投资者防范风险提供了实践价值。

近年来，社交媒体已成为公众表达情绪和观点的主要平台，其中微博因其开放性、实时性和交互性，成为金融舆情的重要载体[3]。随着用户表达形式的多样化，微博等平台上的信息传播呈现图文并茂的特征。用户不仅通过文本表达观点，还会配合使用图片、表情包甚至短视频等多模态信息，以增强表达效果和情绪感染力。已有研究表明，单一模态的情感识别容易忽视隐含在图片或图文交互中的重要线索，从而降低模型的精度[4]。多模态情感识别因能够融合不同模态间的关联性和互补性，被广泛应用于社交舆情、产品评价及金融文本分析等任务中[5]。在金融舆情场景下，若能有效融合图文信息，不仅有助于更全面地刻画舆情内容，还能有效提升负向情感强度识别的准确性。

然而，现有研究在融合图文信息时，对模态间潜在的、深层的结构化关联挖掘不足；同时，单一模型的决策能力有限，难以在复杂多变的金融舆情场景下保持稳定和泛化性能。基于此，本文提出一种基于 GCN-Stacking 协同增强的多模态金融舆情负向情感强度识别模型。首先，在数据层面，本文通过爬取微博金融热门话题数据，并采用多模型一致性标注策略，在缺乏人工标注的条件下获得高置信度的情感强度标签。其次，在特征建模层面，本文利用对比语言-图像预训练（contrastive language-image pretraining, CLIP）模型提取文本与图像的多模态特征，并引入图卷积网络（graph convolutional network, GCN）实现特征转换，以捕捉模态间的结构化关系。随后，将原始与转换后的多模态特征进行融合，形成增强特征输入堆叠集成（stacked generalization, Stacking）学习框架，进一步提升模型的决策能力。最后，通过对比实验验证了所提模型的有效性，结果显示本文方法在各项指标上均优于单一模态和传统融合模型。研究结论不仅为金融舆情负向情感强度识别提供新的技术路径，也为金融风险监测与舆情治理提供有益参考。

1　相关研究

金融舆情通过分析金融事件、投资者评论等，揭示市场情绪倾向与行为预期，对资产价格、市场稳定有直接或间接影响[6]。随着社交媒体的发展，微博作为主要的金融舆情传播平台，其情感倾向对市场影响显著[7]，且用户常以文字、图片、走势截图等多模态形式表达观点，使舆情情绪更直观。这种传播模式下，负面舆情规模性扩散易引发投资者恐慌，甚至诱发系统性风险。研究表明，负面舆情集中易引发投资者恐慌，且股指期货市场对负向情感反应更强，低落情绪会削弱其避险能力[8-9]。因此，研究金融负向情感可为风险预警、市场稳定提供支撑，而准确识

别负向情感及强度，能助金融机构在市场波动初期施策，降低系统性风险概率。

1.1 金融情感分析

金融情感分析是指通过自然语言处理、机器学习以及深度学习等技术，对金融文本中隐含的情绪和态度进行分析和分类。其核心目标是准确提取投资者、媒体或公众在金融事件中的情绪倾向，同时关注情绪的强度和变化趋势[10]。金融情感分析不仅可以为市场情绪监测提供数据支持，还能够为投资决策、风险评估以及政策制定提供重要依据。

金融情感分析方法主要包括词典与规则驱动方法、传统机器学习方法、深度学习方法等。姚加权等人[11]构建了中文金融情绪词典，并提出年报语调和社交媒体情绪指标，在预测股票相关指标上表现优越。姜富伟等人[12]基于人工筛选和 Word2Vec 扩充词典，构建财经媒体情绪指标，能够有效反映投资者情绪并预测股票回报。Agarwal[13]提出了一种基于知识库和领域特定表示的金融情感分析模型。该方法通过将通用词向量转换为金融领域的专用词向量，并利用知识库扩充训练词汇，从而增强模型对稀有词和未登录词的处理能力。Popoola 等人[14]提出了一种基于机器学习的金融舆情情感分析方法，使用词频-逆文本频率(term frequency-inverse document frequency, TF-IDF)进行特征提取，通过机器学习算法进行预测，得到了显著的结果。Ahmad 等人[15]采用机器学习和深度学习对金融文本情绪进行极性分类。结果显示，多项式朴素贝叶斯和逻辑回归分别达到良好和较高的准确率，而深度学习模型的准确率则非常优秀，表明预处理步骤对提升模型准确性有积极作用。刘壮等人[16]使用基于 Transformer 架构的预训练语言模型 BERT(bidirectional encoder representations from Transformers)开发了金融预训练模型 F-BERT。该模型通过对大规模金融文本进行无监督训练，自动学习并存储金融领域知识，无需针对具体任务调整模型结构，即可直接应用于金融情感分析等下游任务。Huang 等人[17]开发了一个适用于金融领域的先进大型语言模型 FinBERT(a finance domain-adapted large language models based on Google's BERT algorithm)。他们证明了 FinBERT 能够整合金融知识，并更好地概括金融文本中的上下文信息。与此同时，Kim 等人[18]基于预训练语言模型对金融舆情进行建模，实现了在有限标注数据下的高精度预测。在 FinBERT 的基础上，Xu 等人[19]等提出了一个专用的双向编码器 FinBERT2，该模型已在目前规模最大中文金融预训练语料库进行训练，成功弥补大语言模型在金融部署方面的差距。Blythe[20]采用基于 BERT 的深度学习模型，有效捕捉复杂上下文语义信息，在金融情感分析任务中表现出突出优势。

现有研究在情感倾向识别上取得了一定成果，但对负向情感的强度分析仍显不足。负向情感对市场波动的影响更大，并且在舆情传播中更容易形成恐慌效应。因此，对负向情感的细粒度量化，能够帮助监管机构及金融机构在市场情绪异常时提前采取干预措施，从而降低系统性金融风险的可能性。

1.2 多模态情感分析

随着微博、投资论坛等社交媒体的普及，金融舆情信息日益呈现出多模态融合的特征。用户在发表市场观点或情绪表达时，不仅依托文本描述，还频繁借助相关图片、走势截图等视觉信息作为补充。这类图文结合的表达方式，不仅提供了更为立体的情感线索和语境信息，也有助于更全面捕捉投资者的真实情绪与心理倾向，从而为精准识别金融舆情中的负面情感强度提供信息支撑。

为有效挖掘图文数据中的情感语义，学者们提出了多种多模态融合方法。Zhu 等人[21]提出了图像-文本交互网络，用于研究情感图像区域与文本之间的关系，并用于多模态情感分析。在公共数据集上的比较表明，所提出的模型优于单模态情感分析方法与基线算法。Li 等人[22]提出了一种基于多模态交叉注意力机制的情感分析模型，用于融合图像和文本信息以进行情绪极性分类。实验结果表明，多模态融合策略显著优于单模态情感分析方法，验证了模型的有效性。王顺杰等人[23]提出了对抗软对比调制动态图学习方法，以提升多模态会话情绪识别的性能。该方法通过构建动态话语图建模会话依赖，并采用对抗性软对比训练机制来增强特征的判别性与鲁棒性。Zhang 等人[24]提出了一种多模态语义融合网络，用于深入探索图像和文本之间的语义关系，并用于多模态情感分析。消融实验证实了多模态语义融合网络的情感分析优于单模态情感分析。Ding 等人[25]提出了一种高效的多模态转换器，该转换器配备了模态感知的自适应训练策略，以促进对多模态序列输入进行最佳的联合训练。该框架在充分利用跨模态交互的同时，确保对单模态特征进行充分学习，进而更全面地保证单模态特征和多模态特征的有效性。Huan 等人[26]提出了包含翻译模块和预测模块的多模态框架。翻译模块利用现有模态的信息翻译缺失模态。预测模块利用注意力机制融合多模

态信息并生成预测。实验结果表明，所提框架能够有效地处理缺失模态的生成任务和未对齐多模态序列的理解任务。陈海丰等人[27]基于混合专家模型，通过自适应路由模块动态评估各模态贡献度并指导融合专家选择，同时引入专家引导损失优化选择过程。多模态融合专家针对不同贡献度组合实现互补融合，并设置共享专家以保留全局信息和减少参数冗余，从而实现动态高效的多模态融合。Liu 等人[28]提出多模态注意力融合网络，用于社交媒体评论情感分析。模型包含交叉注意力与残差 2 个单元：前者选择 1 个核心模态并与其余模态进行交互，生成多组二元注意力特征；后者融入全局信息以增强表示。该方法实现了有效的跨模态融合与全局信息互补，提升了情感分析性能。随着多模态情感分析技术的进步，已有研究者考虑将该技术应用于金融市场预测。Farimani 等人[29]提出一种自适应价格回归模型，用于学习多模态数据与价格变动的关系。模型融合新闻概念向量、市场情绪和技术指标，采用循环卷积网络提取时序特征，并通过自适应加权融合多模态信息，有效捕捉数据分布与市场动态。Song 等人[30]基于股票评论视频构建的多模态情感分析数据集，设计了一种多模态多任务情感分析方法。该方法能够挖掘多模态短语级情感线索，并自适应地抑制各模态短语级特征中的噪声信息。Todd 等人[31]提出了一种融合音频与文本的多模态金融情绪分析方法。该方法整合 FinBERT 文本特征与副语言音频特征，并采用深度学习分类器进行预测。相比传统词典方法与机器学习模型，该多模态分类器在样本外测试中表现出更高的预测准确率。

尽管现有多模态情感分析方法已取得显著进展，但其在金融舆情分析中的应用仍较为有限，尤其针对负面情感强度识别的研究极为稀缺。现有方法难以充分适应金融领域特有的模态表达形式（如 K 线图、股评截图、符号化表情包等），同时也缺乏对负面情感强度连续变化（从轻微负面到极度负面）的系统性建模能力。针对上述问题，本文聚焦于金融舆情中的负面情感强度识别研究，采用 CLIP 模型构建文本与图像的统一语义表征，充分利用其跨模态对齐能力以提取高质量的多模态特征。在此基础上，引入 GCN 深入挖掘模态内和模态间的结构化语义关联，实现特征增强与交互建模。进一步地，通过 Stacking 集成学习框架融合多个基学习器的预测结果，以提升最终情感强度判别的准确性。本文在社交媒体收集的真实金融多模态数据集上进行实验，全面验证所提模型的有效性，以期为金融风险感知与智能化舆情监管提供技术支撑。

2 多模态金融舆情负向情感强度识别模型

在金融舆情分析场景中，本文构建了一种基于 GCN-Stacking 协同增强的负向情感强度识别模型。首先，在数据层面，本文爬取微博金融热门话题数据，并采用多模型一致性标注策略，在缺乏人工标注的条件下生成高置信度的情感强度标签，以保证训练数据的可靠性。其次，在特征建模层面，本文利用 CLIP 提取文本与图像的多模态特征，并引入 GCN 进行特征转换，以捕捉模态间的结构化关系。随后，将原始与转换后的多模态特征进行融合，并输入至 Stacking 集成学习框架，从而进一步提升模型的判别能力。基于 GCN-Stacking 协同增强的多模态金融舆情负向情感强度识别框架如图 1 所示。

2.1 多模态特征构建

为充分挖掘金融舆情文本-图像双模态信息的关联价值，本文设计通过 CLIP 实现双模态语义对齐，结合 GCN 捕捉模态间结构化关系，最终生成兼具全局语义与局部关联的增强型多模态特征。

1）CLIP 多模态特征提取

CLIP 模型通过大规模文本-图像对预训练，使文本与图像特征映射至同一语义空间，具备天然的跨模态对齐能力[32]。本文选用 CLIP 作为基础模型，分别对微博文本与配图进行特征提取，确保双模态特征的语义一致性与维度统一性。

针对微博文本的口语化、碎片化特点，执行文本预处理（去特殊符号、停用词等），随后通过 CLIP 文本编码器完成特征转换。CLIP 先通过内置分词器将文本转换为 Token 序列并添加特殊标记；再通过嵌入层将其映射为包含词汇语义、词序信息的向量矩阵；随后经多层 Transformer 编码器捕捉文本内部语义关联，提取句首标记对应的输出向量并归一化；最终得到维度为 1×512 的文本特征向量，实现文本语义的数值化表征。

$$\boldsymbol{f}_{\text{text}} = Norm\left(CLIP_{\text{text-enc}}(Tokenize(\boldsymbol{x}_{\text{text}}))_{[\text{CLS}]}\right), \quad (1)$$

其中 $\boldsymbol{x}_{\text{text}}$ 为单条微博预处理后文本，$Tokenize(\cdot)$ 为分词函数，$CLIP_{\text{text-enc}}(\cdot)$ 为 CLIP 文本编码器，$Norm(\cdot)$ 为 L2 归一化操作，$\boldsymbol{f}_{\text{text}} \in \mathbb{R}^{512\times 1}$ 为最终文本特征向量。

针对微博配图的尺寸多样性与内容复杂性，采用 CLIP 图像预处理与视觉编码器完成特征提取。

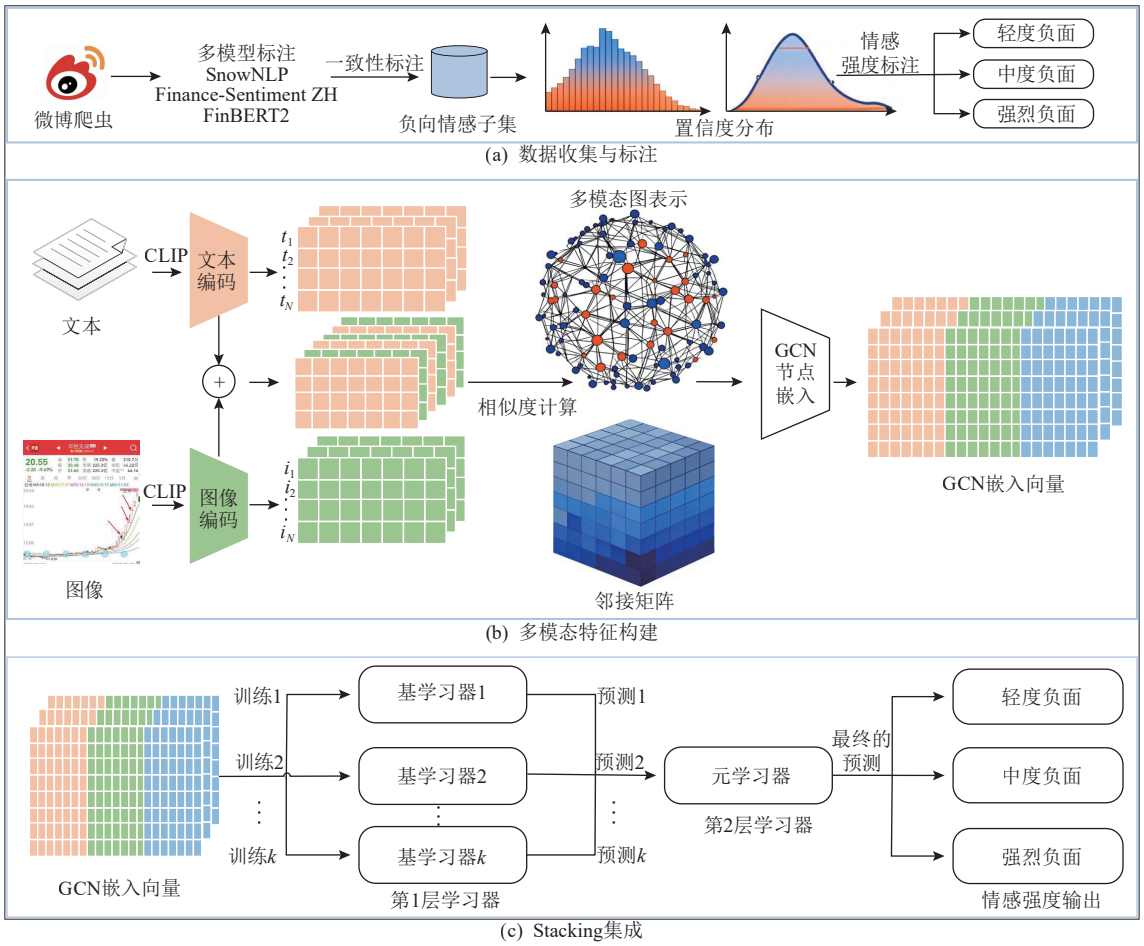


Fig. 1 GCN-Stacking collaborative enhancement-based multimodal financial public opinion negative sentiment intensity recognition

图 1 基于 GCN-Stacking 协同增强的多模态金融舆情负向情感强度识别

CLIP 先对图像进行标准化处理，再通过 ViT(vision Transformer)将图像分割为固定大小的图像块，经线性投影转换为嵌入向量并添加 [CLS] 标记。随后通过多层 Transformer 编码器捕捉图像块间的空间关联与语义依赖，最终提取 [CLS] 标记对应的输出向量并归一化，得到维度为 1×512 的图像特征向量，实现图像语义的数值化表征。

$$\boldsymbol{f}_{\text{image}} = Norm\left(CLIP_{\text{image-enc}}\left(Resize(\boldsymbol{x}_{\text{image}})\right)_{[\text{CLS}]}\right), \quad (2)$$

其中 $\boldsymbol{x}_{\text{image}}$ 为单条微博配图，$Resize(\cdot)$ 为图像尺寸标准化函数，$CLIP_{\text{image-enc}}(\cdot)$ 为 CLIP 图像编码器，$\boldsymbol{f}_{\text{image}} \in \mathbb{R}^{512\times 1}$ 为最终图像特征向量。

2) 多模态图表示

为捕捉样本间的模态关联与语义相似性，本文基于 CLIP 提取的双模态特征构建图结构数据(记为 $\mathcal{G}=(\mathcal{V},\mathcal{E},\boldsymbol{X},\boldsymbol{W})$)，其中 $\mathcal{V}$ 为节点集，$\mathcal{E}$ 为边集，$\boldsymbol{X}$ 为节点特征矩阵，$\boldsymbol{W}$ 为边权重矩阵。

将每条多模态样本视为 1 个节点，即节点集 $\mathcal{V}=\{v_1,v_2,\cdots,v_N\}$，节点 v_i 对应第 i 条样本。设文本特征权重 $\alpha=0.5$，图像特征权重 $\beta=0.5$(经交叉验证确定)，通过标准化消除特征尺度差异，再计算加权融合特征

$$\boldsymbol{x}_i = \alpha\cdot Scaler(\boldsymbol{f}_{\text{text},i}) + \beta\cdot Scaler(\boldsymbol{f}_{\text{image},i}), \quad (3)$$

其中 $Scaler(\cdot)$ 为 Z-score 标准化函数，$\boldsymbol{f}_{\text{text},i}$ 为第 i 个节点的文本特征，$\boldsymbol{f}_{\text{image},i}$ 为第 i 个节点的图像特征，$\boldsymbol{x}_i \in \mathbb{R}^{512}$ 为第 i 个节点的融合特征。最终节点特征矩阵

$$\boldsymbol{X} = (\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_N)^{\mathrm{T}} \in \mathbb{R}^{N\times 512}。 \quad (4)$$

采用 K 近邻来衡量与中心节点相似的候选节点，避免检查所有不相关的节点对，节省计算复杂度和时间。相似度阈值用来确保连接质量，去除弱连接，只保留真正有意义的连接。本文基于余弦相似度衡量节点间融合特征的语义相似性，第 i 个节点与第 j 个节点的相似度为

$$s_{i,j} = \frac{\boldsymbol{x}_i\cdot\boldsymbol{x}_j^{\mathrm{T}}}{\|\boldsymbol{x}_i\|\cdot\|\boldsymbol{x}_j\|}, \quad (5)$$

其中 $s_{i,j}\in[0,1]$，值越大表示 2 节点语义越相似。对每个节点 v_i，筛选相似度最高的 K 个邻居，排除节点自

身($s_{i,i}=1$)；设置相似度阈值θ，仅保留邻居中相似度$s_{i,j}>\theta$的节点对，构建边$(v_i,v_j)\in\mathcal{E}$，并将$s_{i,j}$作为边权重$w_{i,j}$。因此，基于上述策略构建了多模态图表示G。

3）GCN 特征转换与跨模态融合

GCN 通过聚合节点自身特征与邻居特征，实现对图结构中结构化信息的挖掘。本文设计l层 GCN 对多模态图G进行节点嵌入，捕捉模态内依赖与模态间关联，再与原始 CLIP 特征融合，生成增强型多模态特征。

GCN 通过图卷积层实现特征聚合，第l层 GCN 的特征转换公式为

$$\boldsymbol{H}^{(l)}=\sigma\left(\tilde{\boldsymbol{A}}\boldsymbol{H}^{(l-1)}\boldsymbol{W}^{(l)}+\boldsymbol{b}^{(l)}\right), \tag{6}$$

其中$\boldsymbol{H}^{(l)}\in\mathbb{R}^{N\times d_l}$，为第$l$层的节点特征矩阵，$d_l$为第$l$层特征维度。$\tilde{\boldsymbol{A}}=\boldsymbol{D}^{-1/2}(\boldsymbol{A}+\boldsymbol{I})\boldsymbol{D}^{-1/2}$为归一化邻接矩阵，$\boldsymbol{A}$为原始邻接矩阵，$\boldsymbol{I}$为单位矩阵，$\boldsymbol{D}$为度矩阵。$\boldsymbol{W}^{(l)}\in\mathbb{R}^{d_{l-1}\times d_l}$为第$l$层可学习权重矩阵，$\boldsymbol{b}^{(l)}\in\mathbb{R}^{d_l}$为偏置向量。$\sigma(\cdot)$为 ReLU 激活函数。

融合原始 CLIP 双模态特征与 GCN 结构特征，生成最终的增强型多模态特征。第i个节点的融合特征公式为

$$\boldsymbol{F}_i=(\boldsymbol{f}_{\text{text},i},\boldsymbol{f}_{\text{image},i},\boldsymbol{f}_{\text{g},i}), \tag{7}$$

其中$\boldsymbol{F}_i\in\mathbb{R}^{512+512+d=1\,024+d}$为第$i$个样本的多模态特征。最终所有样本的特征矩阵为$\boldsymbol{F}=(\boldsymbol{F}_1,\boldsymbol{F}_2,\cdots,\boldsymbol{F}_N)^{\mathrm{T}}\in\mathbb{R}^{N\times(1\,024+d)}$。

2.2 Stacking 集成情感强度识别

针对微博金融舆情负向情感强度识别任务，本文基于多模态特征矩阵$\boldsymbol{F}\in\mathbb{R}^{N\times(1\,024+d)}$，设计双层 Stacking 集成学习框架。该框架通过第 1 层基础分类器提取多模态特征的差异化判别信息，第 2 层元学习器融合多源预测结果以优化最终分类性能，有效提升情感强度识别的精度与鲁棒性。本文的 Stacking 框架分为基础分类器筛选与双层集成建模 2 个阶段，先通过性能对比确定最优基础分类器集合，再构建双层结构完成最终预测。

1）基础分类器筛选

为确保基础分类器的多样性与互补性，本文选取了覆盖不同学习范式的分类模型作为候选，涵盖线性模型、树模型、集成树模型、神经网络及注意力机制模型，可充分适配多模态特征中文本语义、图像视觉以及图结构关联的复杂信息。研究以多模态特征$\boldsymbol{F}$为输入，通过多分类任务核心指标对候选模型进行性能排序，筛选出识别性能较高的模型作为 Stacking 第 1 层的基础分类器。

2）双层集成建模

筛选出第 1 层的基础分类器的集合为$\mathcal{M}=\{M_1,M_2,\cdots,M_k\}$，$k$为第 1 层的基础分类器的数量。多模态特征矩阵划分为训练集$\boldsymbol{F}_{\text{tr}}\in\mathbb{R}^{N_{\text{tr}}\times(1\,024+d)}$与测试集$\boldsymbol{F}_{\text{te}}\in\mathbb{R}^{N_{\text{te}}\times(1\,024+d)}$($N_{\text{tr}}+N_{\text{te}}=N$)，情感强度标签向量$\boldsymbol{y}\in\{0,1,2\}^N$（0=中度负面、1=强烈负面、2=轻度负面）。对每个基础分类器$M_i\in\mathcal{M},i\in\{1,2,\cdots,k\}$，基于训练集$F_{\text{tr}}$与$y_{\text{tr}}$（训练集标签）训练模型，学习特征到情感强度的映射关系M_i：$\mathbb{R}^{(1\,024+d)}\to\{0,1,2\}$。利用训练好的$M_i$对训练集$F_{\text{tr}}$与测试集$F_{\text{te}}$分别输出类别概率分布。

训练集元特征：

$$\boldsymbol{h}_{i,\text{tr}}=(p_{i,\text{tr}}(0),p_{i,\text{tr}}(1),p_{i,\text{tr}}(2))^{\mathrm{T}}, \tag{8}$$

其中$p_{i,\text{tr}}(c)$表示M_i预测训练样本为类别c的概率，$\boldsymbol{h}_{i,\text{tr}}\in\mathbb{R}^{N_{\text{tr}}\times 3}$。

测试集元特征

$$\boldsymbol{h}_{i,\text{te}}=(p_{i,\text{te}}(0),p_{i,\text{te}}(1),p_{i,\text{te}}(2))^{\mathrm{T}}, \tag{9}$$

同理，$\boldsymbol{h}_{i,\text{te}}\in\mathbb{R}^{N_{\text{te}}\times 3}$。

将k个基础分类器的元特征按列拼接，形成元学习器的输入“元特征矩阵”。

元学习器训练集

$$\boldsymbol{H}_{\text{tr}}=(\boldsymbol{h}_{1,\text{tr}},\boldsymbol{h}_{2,\text{tr}},\cdots,\boldsymbol{h}_{k,\text{tr}})\in\mathbb{R}^{N_{\text{tr}}\times k\times 3}。 \tag{10}$$

元学习器测试集

$$\boldsymbol{H}_{\text{te}}=(\boldsymbol{h}_{1,\text{te}},\boldsymbol{h}_{2,\text{te}},\cdots,\boldsymbol{h}_{k,\text{te}})\in\mathbb{R}^{N_{\text{te}}\times k\times 3}。 \tag{11}$$

上述过程的数学表达为

$$\boldsymbol{H}_{\text{tr}}=(M_1(F_{\text{tr}};\theta_1)_{\text{prob}},M_2(F_{\text{tr}};\theta_2)_{\text{prob}},\cdots,M_k(F_{\text{tr}};\theta_k)_{\text{prob}}), \tag{12}$$

$$\boldsymbol{H}_{\text{te}}=(M_1(F_{\text{te}};\theta_1)_{\text{prob}},M_2(F_{\text{te}};\theta_2)_{\text{prob}},\cdots,M_k(F_{\text{te}};\theta_k)_{\text{prob}}), \tag{13}$$

其中θ_i为基础分类器M_i的模型参数，$(;)_{\text{prob}}$表示模型输出的类别概率。

第 2 层元学习器旨在如何最优地融合基础分类器的概率输出。本文选用逻辑回归作为元学习器，其在 Stacking 集成框架中已被广泛应用并验证有效。以元学习器训练集$\boldsymbol{H}_{\text{tr}}$为输入，训练集标签$y_{\text{tr}}$为监督信号，训练逻辑回归模型$L:\mathbb{R}^{k\times 3}\to\{0,1,2\}$，学习元特征到情感强度的线性映射关系：

$$L(\boldsymbol{H};\boldsymbol{\omega},\boldsymbol{b})=softmax(\boldsymbol{H}\boldsymbol{\omega}+\boldsymbol{b}), \tag{14}$$

其中$\boldsymbol{\omega}\in\mathbb{R}^{k\times 3}$为权重矩阵（每行对应 1 个基础分类器的 3 类概率权重，每列对应 1 个情感强度类别的权重），$\boldsymbol{b}\in\mathbb{R}^3$为偏置向量，函数$softmax(\cdot)$将线性输出转换为类别概率分布。

利用训练好的元学习器L，对元学习器测试集$\boldsymbol{H}_{\text{te}}$输出最终类别概率分布$\hat{p}(\boldsymbol{H}_{\text{te}};\boldsymbol{\omega},\boldsymbol{b})$，确定最终情感

强度标签：

$$\hat{\boldsymbol{y}}_{\text{te}} = \arg\max_{c\in\{0,1,2\}} L(\boldsymbol{H}_{\text{te}};\boldsymbol{\omega},\boldsymbol{b})_c \tag{15}$$

其中$\hat{\boldsymbol{y}}_{\text{te}} \in \{0,1,2\}^{N_{\text{te}}}$为测试集的最终情感强度预测结果，$L(\boldsymbol{H}_{\text{te}};\boldsymbol{\omega},\boldsymbol{b})_c$表示元学习器预测测试样本为类别$c$的概率。

3 实证研究

3.1 数据收集与标注

本文以“寒武纪”“华胜天成”“领益智造”等39个金融领域热门讨论对象为关键词，从微博平台获取用户生成内容，去重后共获得36 019条舆情数据。为了保证文本质量，利用Python的正则化模块剔除了URL、特殊符号(@, $, &, *, #等)、括号、表情符号、英文字符及数字等无关信息。针对金融舆情标注普遍存在的人工成本高、领域适配性差的问题，本文采用弱监督学习框架下的多模型一致性标注方法。通过多个金融领域适配的预训练模型对同一文本进行独立标注，以投票一致原则筛选高置信度负向样本，从而有效规避单一模型的标注偏差[33]。本文筛选适配金融领域的预训练模型进行数据标注，详情如表1所示。

设3个标注模型的输出为$f_1, f_2, f_3 \in \{-1,0,1\}$，最终标签$y_i$公式如下：

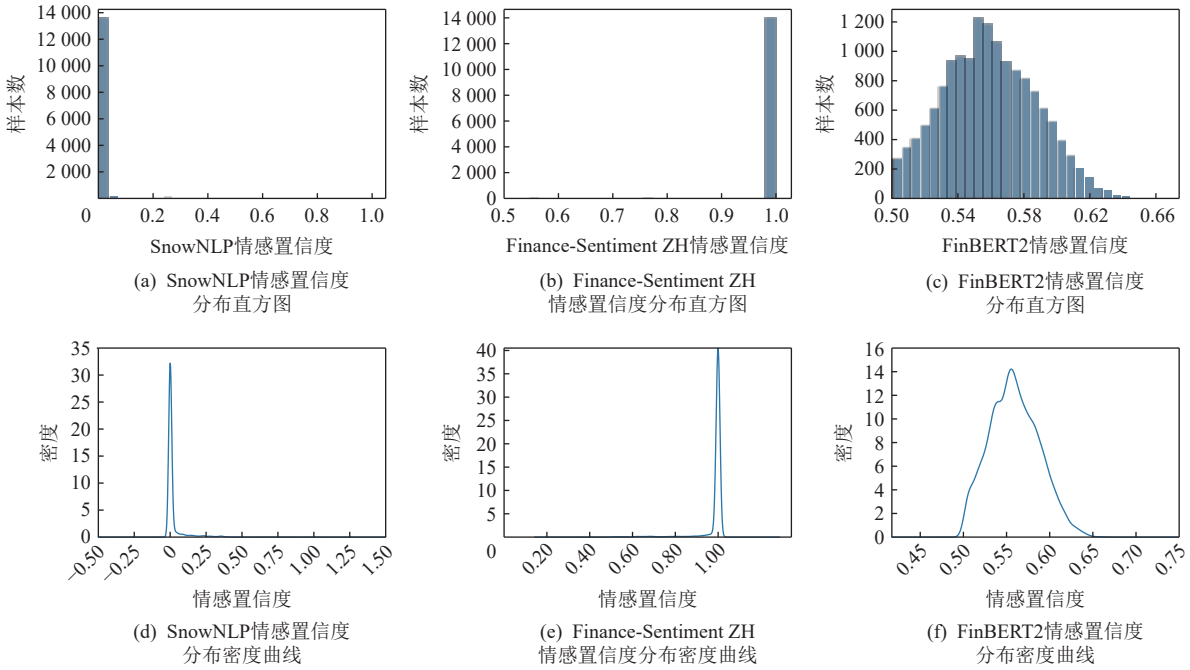
$$y_i = \begin{cases} c, \text{ 若存在} c \in \{-1,0,1\}, \\ \quad \text{使得至少2个分类器输出为} c; \\ \text{丢弃, 其他情况,} \\ \quad \text{即} f_1, f_2, f_3 \text{各取不同值。} \end{cases} \tag{16}$$

当2个或2个以上分类器输出同一值c时，直接取c，当三者分别为 −1, 0, 1时，直接丢弃。因此，本文对36 019条微博文本完成标注后，筛选出$c=-1$的所有文本，即只选用高置信度负向情感样本进行金融情感强度识别。随后，从负向文本中进一步选取包含图片的子集，共得到15 133条图文样本。在负样本基础上，本文进一步对3种模型的负向情感置信度分布进行了对比分析如图2所示。结果显示SnowNLP的置信度分布过度偏斜，缺乏对负向情感

Table 1 Multi-Model Consistency Labeling

表1 多模型一致性标注

标注模型	模型类型	领域适配优化措施	标注输出形式
SnowNLP	传统机器学习模型	替换原始产品评论词典为金融领域专用词典（整合姚加权等人[11]与姜富伟等人[12]的金融文本词典，更适用于社交媒体等金融领域文本的情感分析）	情感极性（正向/中性/负向）+情感得分（0~1）
Finance-Sentiment ZH	轻量级Transformer模型	直接采用BardsAI公司开源的中文金融情感模型（基于Hugging Face Transformers框架调用，预训练数据含50万条中文金融文本），该模型已被广泛用于各种情感分析任务[34]	情感极性（正向/中性/负向）+置信度（0~1）
FinBERT2	金融专用BERT模型	采用在32B金融领域Token上预训练的FinBERT2-Base版本，适配金融术语的语义理解[19]	情感极（正向/中性/负向）+置信度（0~1）



(a) SnowNLP情感置信度分布直方图 (b) Finance-Sentiment ZH情感置信度分布直方图 (c) FinBERT2情感置信度分布直方图

(d) SnowNLP情感置信度分布密度曲线 (e) Finance-Sentiment ZH情感置信度分布密度曲线 (f) FinBERT2情感置信度分布密度曲线

Fig. 2 The confidence distribution of negative emotion

图2 负向情感置信度分布

强度的有效区分；Finance-Sentiment ZH 的置信度高度集中在接近 1 的区间，导致所有负向文本被统一视为极强负面，难以体现细粒度差异；FinBERT2 的置信度分布相对均衡，能够自然刻画不同层次的负向情绪梯度，因此更适合用于后续的情绪强度划分。

基于 FinBERT2 的置信度输出，本文采用分位数方法作为无监督分箱标准，对负向情绪进行 3 档划分。FinBERT2 作为在 32B 金融领域 Token 上预训练的专用模型，其置信度输出已在金融情感分析任务中展现出优异的判别能力[19]；分位数划分则是情感强度分级研究中成熟的无监督方法[35]。通过中位数与 90% 分位点的划分策略，本文在无需人工标注的条件下构建了具有统计学区分度的情感强度数据集，为后续多模态负向情感强度识别提供了可靠的数据基础。具体划分如下：

$$Strength_{score}=\begin{cases}\text{轻度}, & score \leqslant Q_{0.5},\\ \text{中度}, & Q_{0.5}<score \leqslant Q_{0.9},\\ \text{强烈}, & score > Q_{0.9},\end{cases} \tag{17}$$

其中 $Q_{0.5}$ 和 $Q_{0.9}$ 分别表示 FinBERT2 负向置信度分布的中位数和 90% 分位点。通过该方法，负向情感被划分为轻度负面、中度负面、强烈负面 3 个等级，为后续的情绪强度建模提供了基础数据支持。最终获得的实验数据详情如表 2 所示。

3.2 实验设置

针对金融舆情负向情感强度分类任务，为全面衡量模型性能并规避过拟合风险，实验采用 5 折交叉验证将数据集随机划分为训练集(80%)与测试集(20%)，重复 5 次训练与测试后取指标均值及标准差。考虑到样本类别不平衡特性(轻度负面 7 567 条、中度负面 6 053 条、强烈负面 1 513 条)，所有分类指标均采用加权平均计算，以平衡各类别样本占比的影响，核心评估指标包括 *F*1-分数(*F*1)、召回率(R)、精确率(P)、准确率(A)和 AUC。

Table 2 Data Details and Examples

表 2 数据详情及示例

数据信息	轻度负面	中度负面	强烈负面
文本	华胜天成 sh600410 这波已经 3 次反包了，今天深水，明天再次反包的可能性比较小。	领益智造大跌那天感觉天都塌了，然后打开通讯录找富婆姐姐要了 2 万零花钱加了一笔，现在可以开心的笑了	上周比较大的一个失误就是没有上长/城/军/工，这 2 天想上又觉得不想去冒险，剑管都喊话。A 股超话#股民交流##今日看盘##a 股#长城军工 sh601606
图像			
标签数量	7 567	6 053	1 513

3.3 数据平衡

针对数据类别不平衡问题(强烈负面样本占比仅 10%)，本文采用合成少数类过采样技术(synthetic minority oversampling technique, SMOTE)对训练集进行平衡处理。SMOTE 作为处理不平衡分类问题的经典方法，通过在少数类样本间进行线性插值生成新样本，能够有效缓解类别分布偏斜带来的模型偏向问题。由表 3 可知，SMOTE 过采样在 *F*1, AUC 等关键指标上均取得最优结果，显著优于无重采样及其他对比方法。尽管强烈负面样本原始基数较小，但 SMOTE 通过在高维特征空间中对少数类样本进行插值，能够在保持类别边界特征的同时生成多样化的合成样本，有效缓解了小样本带来的过拟合风险。

Table 3 Comparative Results and Standard Deviation of Different Resampling Methods

表 3 不同重采样方法对比结果及标准差 %

采样方法	*F*1	R	P	A	AUC
无重采样	82.11±0.27	82.32±0.26	82.99±0.28	82.32±0.26	94.38±0.25
SMOTE（本文）	**82.34±0.52**	**82.43±0.47**	82.69±0.53	**82.43±0.47**	**94.66±0.33**
随机过采样	82.19±0.63	82.36±0.59	**82.93±0.56**	82.36±0.59	94.61±0.33
ADASYN	82.22±0.50	82.30±0.52	82.48±0.49	82.30±0.52	94.59±0.24
随机欠采样	65.25±0.41	64.92±0.42	66.30±0.43	64.92±0.42	84.64±0.64
Tomek Links	81.34±0.72	81.54±0.72	82.20±0.64	81.54±0.72	93.96±0.35
SMOTE Tomek	81.56±0.45	81.65±0.43	81.89±0.44	81.65±0.43	94.24±0.25
SMOTE ENN	60.48±0.42	58.53±0.53	69.14±0.52	58.53±0.53	84.67±0.48

注：固定 $K=500$，$\theta=0.6$，$d=512$，以随机森林为基准分类器进行实验对比；黑体数值表示指标中的最佳值。“±”前数值表示平均值，“±”后数值表示标准差。

3.4 参数优化与确定

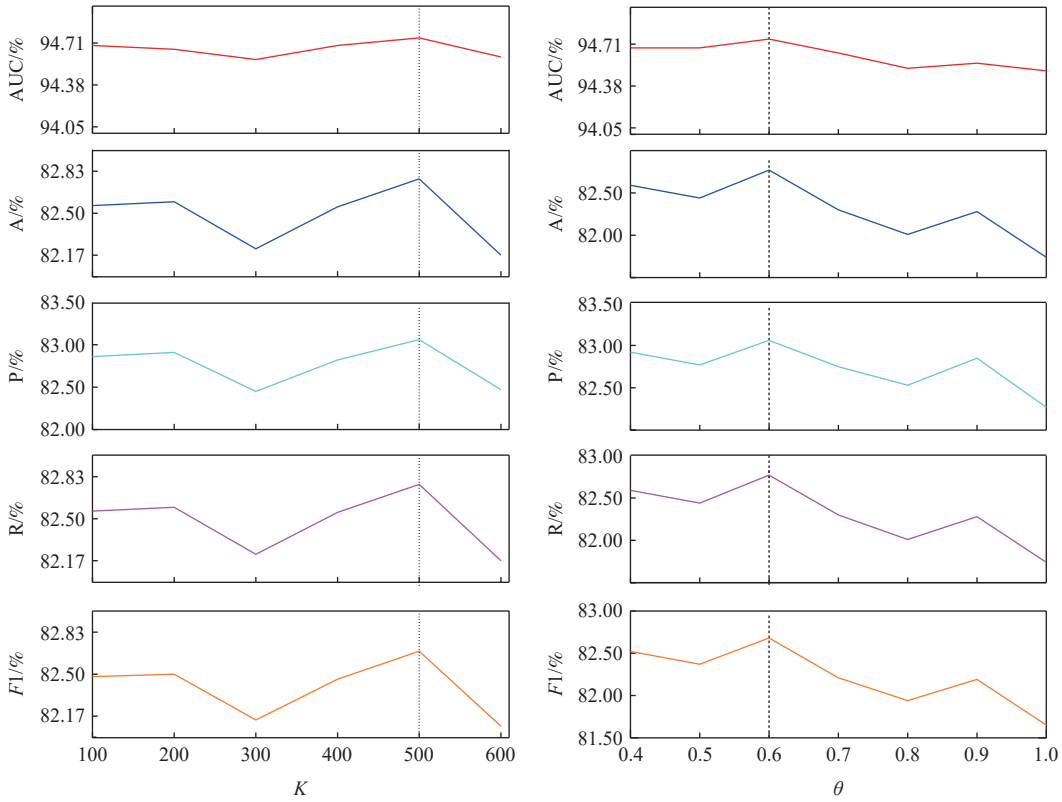
在本文所提模型的图构建与GCN节点嵌入过程中,邻居数K与相似度阈值θ是影响图结构合理性与模型性能的核心参数,二者分别承担候选节点筛选与连接质量控制的关键作用。

为确定最优邻居数K,实验固定参数组合:相似度阈值$\theta=0.6$、GCN嵌入维度$d=512$,以随机森林为基准分类器,在$K\in\{100,200,\cdots,600\}$(步长100)的范围内开展对比实验。如表4所示,选择该范围是由于预实验发现当$K\geqslant300$时,图中边数量与图密度趋于饱和,继续增加K值仅会增加计算成本而无法提升关联节点覆盖度,因此无需进行更高值的验证。实验结果如图3(a)所示,当$K=500$时,模型在各项评估指标上均取得最优结果,表明该值能在候选节点覆盖度与计算效率间取得最佳平衡。因此,后续实验固定$K=500$。

Table 4 Edge Numbers and Graph Density of Graph Network Corresponding to Different Values of K

表4 不同K值对应的图网络边数和图密度

K	边数量	图密度
100	211 931	0.000 925
200	255 611	0.001 116
300	268 626	0.001 173
400	268 626	0.001 173
500	268 626	0.001 173
600	268 626	0.001 173



(a) K值与对应的情感强度识别效果 (b) θ值与对应的情感强度识别效果

Fig. 3 The identification effect of emotional intensity corresponding to different values of K and θ

图3 不同K和θ的取值对应的情感强度识别效果

在确定K后,进一步优化相似度阈值θ,实验固定GCN嵌入维度$d=512$、基准分类器RF,在$\theta\in\{0.4,0.5,\cdots,1\}$(步长0.1)的范围内验证。选择该范围是由于预实验显示当$\theta<0.4$时,如表5所示,过低的阈值会导致图网络边数量激增,图结构过度复杂,引发内存溢出与计算崩溃;而当$\theta>1$时,无任何节点对满足相似度条件,图结构退化至孤立节点状态。图3(b)显示$\theta=0.6$时模型性能最优,低于该值时,弱关联边引入的噪声降低特征判别性;高于该值时,过强的过滤导致有效关联丢失,图结构信息不足。因此,最终确定最优参数组合为$K=500$,$\theta=0.6$,基于最优参数构建的多模态图表示及邻接矩阵如图4所示。

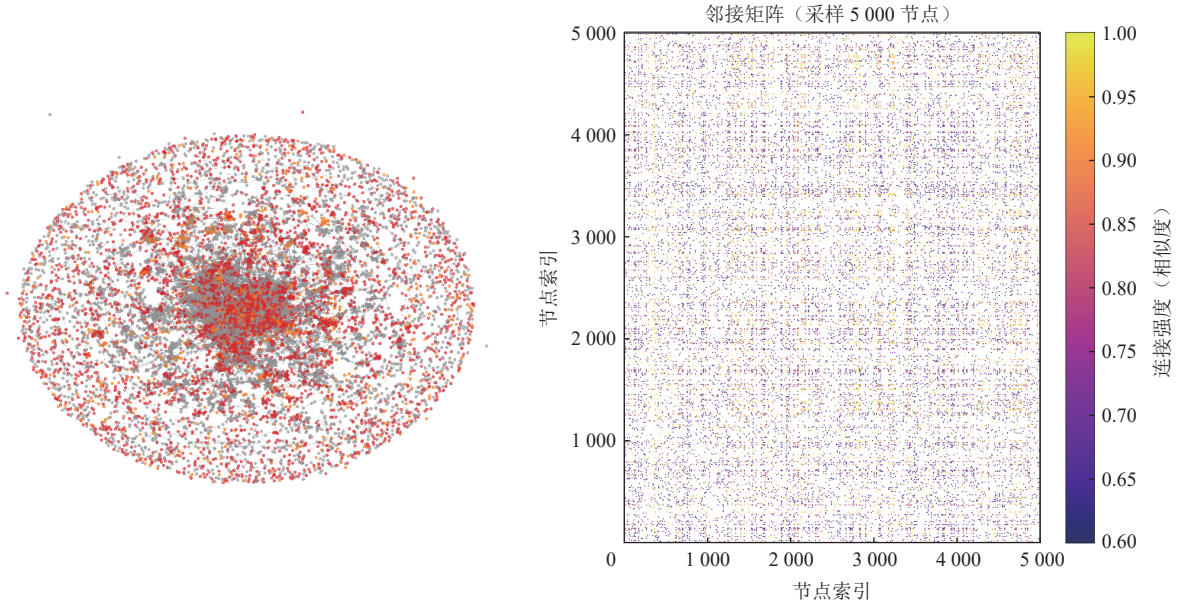
在确定图构建核心参数后,进一步对GCN的网络层数与各层嵌入维度进行优化,二者直接影响节点特征的抽象能力。层数过少易导致特征提取不充

Table 5 Edge Numbers and Graph Density of Graph Network Corresponding to Different Values of θ

表 5 不同θ值对应的图网络边数和图密度

θ	边数量	图密度
0.4	2 350 960	0.010 267
0.5	730 716	0.003 191
0.6	268 626	0.001 173
0.7	154 044	0.000 673
0.8	102 408	0.000 447
0.9	78 772	0.000 344
1	12 166	0.000 053

分，层数过多则可能引发梯度消失；嵌入维度过低会丢失细粒度语义信息，过高则增加计算复杂度且易过拟合。实验通过遍历不同层数（1~3 层）与维度组合（128~1 024 维），以模型性能为指标筛选最优 GCN 配置，前 9 组最佳配置结果如表 6 所示。实验结果显示 3 层 GCN 的最优配置（排名 1）性能优于 2 层与 1 层，3 层网络可通过多轮特征聚合，更充分地捕捉图中节点的多阶关联信息，尤其对多模态特征的跨模态关联融合更具优势。同等层数下，统一维度（如 [256, 256, 256]）的配置性能优于梯度维度（如 [256, 512, 1 024]），统一维度可避免特征维度骤升骤降导致的信息损耗，同时降低模型参数冗余，提升训练稳定性。实验最终确定 GCN 的最优配置为：网络层数为 3，各层嵌入维度依次为 [256, 256, 256]，并将该配置用于后续多模态特征增强与情感强度识别实验。



(a) 多模态图表示 (b) 邻接矩阵

Fig. 4 Multi-modal graph representation and its corresponding adjacency matrix

图 4 多模态图表示及其对应的邻接矩阵

Table 6 The Top 9 Best GCN Configurations and Their Corresponding Recognition Effects

表 6 前 9 个最佳 GCN 配置极其对应的识别效果

排名	层数	各层维度	*F*1/%	R/%	P/%	A/%	AUC/%
1	3	[256, 256, 256]	82.68	82.77	83.06	82.77	94.75
2	2	[1 024, 512]	82.29	82.38	82.71	82.38	94.63
3	1	[1 024]	82.29	82.38	82.68	82.38	94.48
4	1	[512]	82.20	82.30	82.59	82.30	94.56
5	2	[256, 256]	82.17	82.24	82.59	82.24	94.62
6	1	[256]	82.16	82.25	82.60	82.25	94.50
7	3	[256, 512, 1 024]	82.15	82.24	82.48	82.24	94.45
8	2	[512, 512]	82.06	82.16	82.40	82.16	94.51
9	3	[512, 256, 128]	82.03	82.12	82.41	82.12	94.61

3.5 图嵌入方法有效性验证与对比

为验证 GCN 在多模态金融舆情负向情感强度识别中节点嵌入的有效性，实验覆盖 8 种不同学习范式的分类模型：LR（logistic regression），DT（decision tree），RF，GBDT（gradient boosting decision tree），XGB（extreme gradient boosting），LGBM（light gradient boosting machine），MLP（multi-layer perception），TabNet（attentive interpretable tabular learning）。同时引入 GAT（graph attention network），GraphSAGE（graph sample and aggregated），GIN（graph isomorphism network），DGI（deep graph infomax），GCN+GAT 混合模型共 5 种主流图嵌入方法进行对比，将各类图嵌入模型输出的

节点嵌入特征与原始多模态特征拼接后输入基准分类器，通过性能差异验证 GCN 的特征增强能力，结果如图 5 所示。从实验结果可见，基于 GCN 的图嵌入方法在所有评估指标上均展现出更优表现力，尤

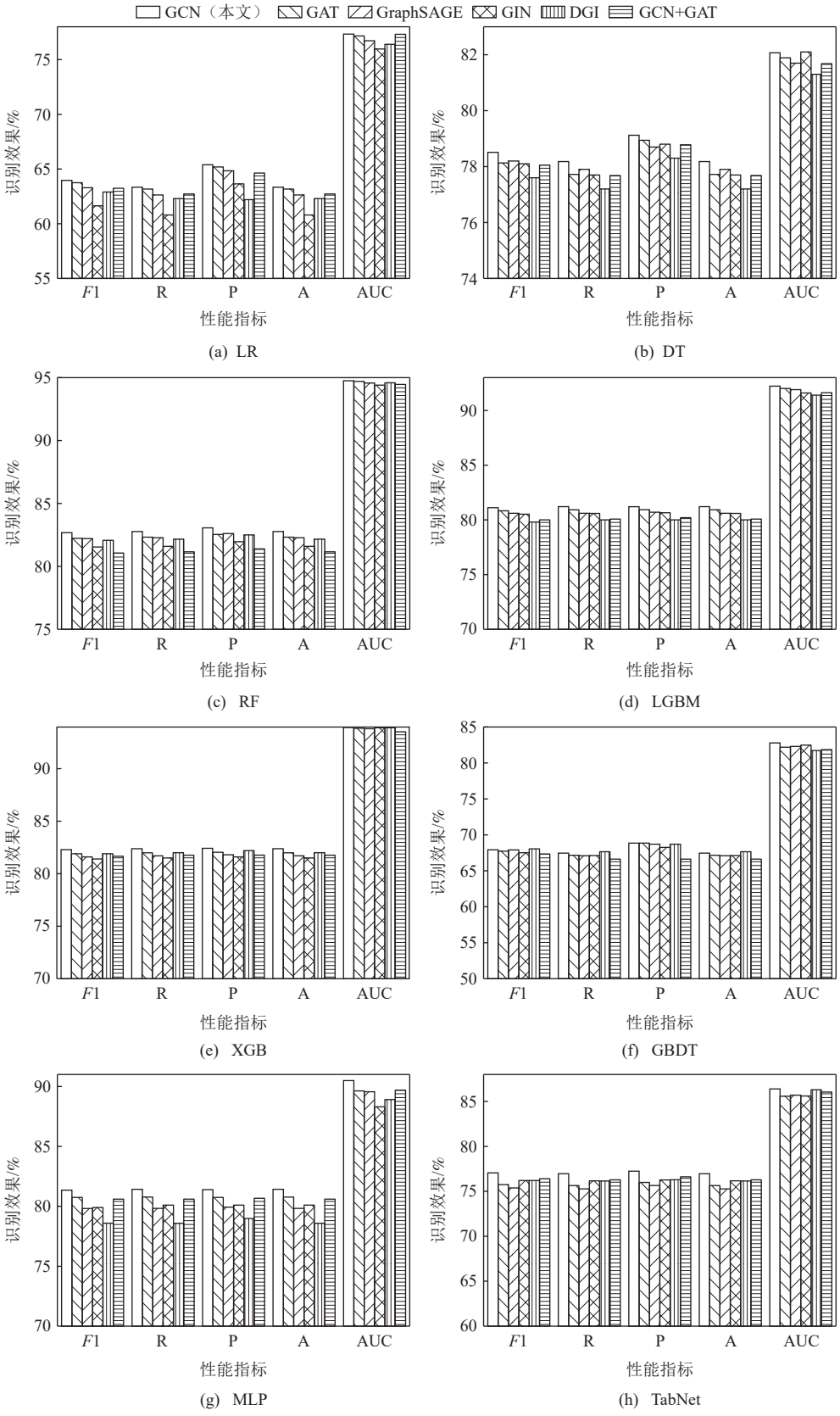


Fig. 5 The effectiveness comparison of graph embedding methods

图 5 图嵌入方法有效性对比

其在 DT, RF, LGBM, MLP, TabNet 上优势更为显著。GCN 之所以能取得更优效果，主要在于其邻域聚合机制。相比于其他图嵌入模型，GCN 通过拉普拉斯平滑聚合邻居特征，在保留节点共性信息的同时有效过滤噪声，更适用于构建在相似度之上的、具有同质性的图结构数据。这充分证明 GCN 能更有效捕捉多模态图中节点的语义关联与结构信息，验证了本文选用 GCN 进行图网络嵌入的合理性。

3.6 基于 GCN 特征增强的有效性验证

为验证 GCN 节点嵌入对多模态特征的增强价值，通过捕捉图结构关联补充隐藏语义信息，进而提升金融舆情负向情感强度识别性能，实验设计原始多模态特征与 GCN 增强多模态特征的对照实验，以 8 种候选分类器为统一基准，分别输入 2 类特征并对比其分类性能，见表 7。实验结果表明，引入 GCN 特征增强后，所有候选分类器的各项评估指标均实现显著提升，验证了 GCN 对多模态特征的结构信息补充作用。从提升幅度来看，LR 的性能增益最为突出：*F*1 分数较原始特征提升 2.26 个百分点，准确率提升 2.49 个百分点。这是因为 GCN 嵌入的结构信息有效弥补了 LR 对非线性关联捕捉能力的不足，丰富了特征判别维度。其次为 TabNet，其 *F*1 分数提升 1.60 个百分点、准确率提升 1.70 个百分点，表明 GCN 捕捉的全局节点关联与 TabNet 的局部特征注意力可形成互补，进一步优化关键信息聚焦效果。即使对 RF, XGB 等本身鲁棒性较强的集成模型，GCN 增强后仍能实现 0.71~1.35 个百分点的指标提升，证明结构信息对多模态特征的增益具有普适性。综上，本文提出的“GCN 多模态特征增强”策略可有效挖掘隐藏的节点关联信息，为后续情感强度识别任务提供更优的特征基础。

3.7 基于 GCN-Stacking 协同增强模型的有效性验证

为验证基于 GCN-Stacking 协同增强模型在金融舆情负向情感强度识别中的优势，实验先筛选 Stacking 第 1 层基础分类器，再通过多模态与单模态模型的横向对比，验证协同增强策略的有效性。

在基础分类器筛选阶段，以 GCN 增强后的多模态特征为输入，对比 8 种候选分类器的识别性能，如图 6 所示。分析发现，RF 与 GBDT 综合性能最优，二者单模型 *F*1 分数均超 82%，且学习机制存在显著互补性（RF 基于 Bagging 降低过拟合，GBDT 基于 Boosting 优化偏差），既能保留不同学习范式的优势，又避免因模型数量过多引入冗余噪声，因此确定为 Stacking 第 1 层基础分类器，元学习器选用 LR 以量化基础模型贡献权重并保障泛化性。

为进一步验证协同增强效果，实验构建多组对比模型，涵盖单文本、单图像、传统多模态 3 类基准，如表 8 所示。从实验结果可归纳出 3 点结论：1）多模态模型整体性能优于单模态模型，且优于带交叉注意力交互的多模态模型（BERT-ViT-CA 与 ViLT），表明本文采用的 CLIP 统一语义表征结合 GCN 结构化建模的方法在金融多模态场景下更具优势；2）GCN 特征增强与 Stacking 集成存在协同效应，对比 CLIP-RF（*F*1=81.49%）与CLIP-GCN（*F*1=82.68%）可见GCN 的图结构信息嵌入丰富了特征判别维度，进一步对比 CLIP-GCN 与 GCN-Stacking（*F*1=83.57%）则表明 Stacking 有效优化了模型偏差与方差，这 2 项策略协同实现了性能的叠加提升；3）GCN-Stacking 协同增强模型性能最优，*F*1 分数为 83.57%、AUC 为 95.05%，相较原始 CLIP 多模态模型 *F*1 分数提升 2.08 个百分点，且各项指标标准差均小于 0.6，兼具高精度与良好稳定性。

Table 7 Performance Comparison Between Feature Enhancement and Original Multi-Modal Features Recognition Based on GCN

表 7 基于 GCN 的特征增强与原始多模态特征的识别性能对比 %

算法	CLIP					CLIP-GCN（本文）					提升百分点	
	*F*1	R	P	A	AUC	*F*1	R	P	A	AUC	*F*1	A
LR	61.70	60.85	63.68	60.85	75.96	63.96	63.34	65.39	63.34	77.33	2.26	2.49
DT	77.20	76.71	78.17	76.71	81.33	78.51	78.18	79.12	78.18	82.07	1.31	1.47
RF	81.49	81.59	81.88	81.59	94.53	82.68	82.77	83.06	82.77	94.75	1.19	1.18
LGBM	79.99	80.07	80.04	80.07	91.86	81.11	81.21	81.20	81.21	92.22	1.12	1.14
XGB	80.93	81.03	81.08	81.03	93.57	82.29	82.38	82.42	82.38	93.92	1.36	1.35
GBDT	67.31	66.75	68.41	66.75	82.26	67.92	67.46	68.85	67.46	82.78	0.61	0.71
MLP	81.08	81.13	81.08	81.13	90.24	81.35	81.42	81.39	81.42	90.49	0.27	0.29
TabNet	75.45	75.26	75.88	75.26	86.44	77.05	76.96	77.24	76.96	86.39	1.60	1.70

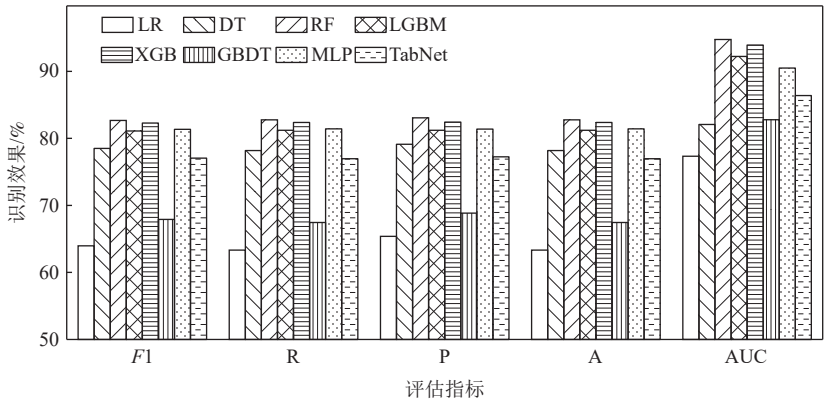


Fig. 6 Comparison of basic classifiers in the first layer of Stacking

图 6 Stacking 第 1 层的基础分类器对比

Table 8 Performance Comparison of Single Modal and Multi-Modal Recognitions

表 8 单模态、多模态识别性能对比 %

模态	算法	*F*1	R	P	A	AUC
文本	TF-IDF-LR	62.89±4.11	61.30±5.93	69.02±0.51	61.30±5.93	79.57±0.81
	Fasttext	53.40±1.00	57.34±0.71	76.50±0.89	57.34±0.71	64.24±0.30
	Bi-LSTM	67.00±3.40	65.84±4.58	71.49±1.65	65.84±4.58	80.49±2.80
	$(CLIP\text{-}RF)_{Text}$	81.09±0.59	81.15±0.58	81.29±0.57	81.15±0.58	94.12±0.24
图像	VGG16	42.93±0.62	43.44±0.51	42.47±0.71	43.44±0.51	49.67±0.45
	ResNet50	42.97±0.78	43.15±0.69	42.82±0.91	43.15±0.39	49.50±0.91
	$(CLIP\text{-}RF)_{Image}$	78.30±0.77	78.34±0.77	78.53±0.78	78.34±0.77	91.88±0.27
多模态	BERT-ViT	82.45±0.38	82.55±0.29	83.46±0.43	82.55±0.41	94.56±0.25
	BERT-ViT-CA	74.58±1.12	74.90±1.11	74.82±1.03	74.90±1.11	85.11±0.96
	ViLT	75.44±1.99	73.13±2.00	79.04±0.78	79.75±2.50	89.89±0.95
	OpenCLIP	79.85±0.64	79.88±0.61	80.04±0.63	79.88±0.61	92.43±0.49
	CLIP-RF	81.49±0.61	81.59±0.62	81.88±0.59	81.59±0.62	94.53±0.26
	CLIP-GCN	82.68±0.60	82.77±0.58	83.06±0.59	82.77±0.58	94.75±0.39
	GCN-Stacking（本文）	**83.57±0.57**	**83.78±0.54**	**84.22±0.54**	**83.78±0.54**	**95.05±0.19**

注：BERT-ViT-CA 表示基于 BERT 与 ViT 特征并引入交叉注意力交互；ViLT 表示 vision-and-language Transformer；黑体数值表示最佳值。“±”前数值表示平均值，“±”后数值表示标准差。

因此，本文提出基于 GCN-Stacking 协同增强模型可有效整合图结构信息与集成学习优势，显著提升多模态金融舆情负向情感强度识别效果，验证了模型设计的合理性。

4 结 语

本文针对图文多模态信息在金融情感表达中的普遍性，以及模态间结构关联有待深入挖掘的问题，提出一种基于图卷积网络与 Stacking 集成学习协同增强的多模态负向情感强度识别模型。研究表明，单一文本模型受限于语义理解的片面性，图像模型无法充分捕捉文本语义信息，而通过 CLIP 模型构建的多模态特征虽能实现模态间信息融合，但未能充分挖掘模态间的潜在结构关联。引入 GCN 对多模态特征进行图结构嵌入后，所有候选分类器性能均有所提升，证明 GCN 能有效挖掘节点间的隐藏语义关联，增强了情感强度识别的特征判别维度。此外，GCN-Stacking 协同策略通过特征增强、决策优化双重增益进一步提升了模型表现。GCN 捕捉复杂的多模态图结构，解决了传统多模态模型忽视模态间结构关联的问题，而 Stacking 集成决策则通过多模型融合降低了偏差，提高了决策精度。然而，本研究也为后续工作提供了进一步探索的方向。首先，本研究采用模型置信度划分情感强度的方法是在缺乏大规模人工标注下的有效折中方案，但未来若能引入更

精细的心理语言学标注体系，有望进一步提升标签的可靠性。其次，数据中强烈负面样本的相对稀缺对模型泛化能力提出了更高要求，探索更先进的生成式数据增强方法将是优化该问题的有益尝试。此外，当前的静态图构建方式在面向大规模实时数据时，其计算开销值得关注，未来结合增量学习或动态图更新机制，是提升模型在实时金融舆情监测中实用性的重要方向。综上所述，本文提出的 GCN-Stacking 协同增强模型有效解决了多模态金融舆情负向情感强度识别中的模态关联挖掘与决策优化问题，为金融舆情智能化分析、金融风险预警和市场稳定维护提供了切实可行的技术支持。

作者贡献声明：李婷婷负责研究设计、实验操作和论文撰写；孙守强、李青青负责论文优化调整；万玲负责文献调研、数据采集。

参 考 文 献

[1] Wu Jiang, Duan Yiqi. Trends and future prospects in sentiment analysis of financial reviews texts[J]. Journal of Information Resources Management, 2025, 15(1): 86−101 (in Chinese)
(吴江, 段一奇. 金融评论文本情感分析研究趋势与未来展望[J]. 信息资源管理学报, 2025, 15(1): 86−101)

[2] Ouyang Zisheng, Chen Shili, Yang Xite, et al. Economic policy uncertainty, network public opinion and systemic risk of financial institutions[J]. Journal of Management Sciences in China, 2023, 26(4): 62−86 (in Chinese)
(欧阳资生, 陈世丽, 杨希特, 等. 经济政策不确定性、网络舆情与金融机构系统性风险[J]. 管理科学学报, 2023, 26(4): 62−86)

[3] Wang Hu, Wu Haowei, Jiang Changbin. Evolution of Weibo public opinion: Heat changes, topic dynamics and sentiment trends[J]. Journal of Intelligence, 2024, 43(11): 144−151, 128 (in Chinese)
(王虎, 吴浩伟, 江长斌. 聚焦热度变化、主题动态与情感趋势的微博舆情演化研究[J]. 情报杂志, 2024, 43(11): 144−151, 128)

[4] Li Tingting, Zeng Ziming, Li Qingqing, et al. Integrating GIN-based multimodal feature transformation and multi-feature combination voting for irony-aware cyberbullying Detection[J]. Information Processing & Management, 2024, 61(3): 103651

[5] Zeng Ziming, Sun Shouqiang, Li Qingqing. Multimodal negative sentiment recognition in online public opinion during public health emergencies based on fusion strategy[J]. Journal of the China Society for Scientific and Technical Information, 2023, 42(5): 611−622 (in Chinese)
(曾子明, 孙守强, 李青青. 基于融合策略的突发公共卫生事件网络舆情多模态负面情感识别[J]. 情报学报, 2023, 42(5): 611−622)

[6] Wang Xueqiu. Research on the characteristics and laws of financial public opinion events[J]. Information Science, 2021, 39(4): 54−61 (in Chinese)
(王雪秋. 突发金融舆情事件信息传播规律与对策研究[J]. 情报科学, 2021, 39(4): 54−61)

[7] Yang Lingling, Liu Xinze. Simulation and demonstration of financial network public opinion evolution mechanism[J]. Journal of Statistics and Information, 2024, 39(1): 74−86 (in Chinese)
(杨玲玲, 刘馨泽. 金融网络舆情演化机制仿真与实证[J]. 统计与信息论坛, 2024, 39(1): 74−86)

[8] Fang Yi, Wang Yanru. Bank-tech sentiment and systemic risk[J]. Journal of Sun Yat-sen University: Social Science Edition, 2023, 63(3): 190−206 (in Chinese)
(方意, 王晏如. 金融科技媒体情绪与银行系统性风险[J]. 中山大学学报: 社会科学版, 2023, 63(3): 190−206)

[9] Tang Yong, Zhan Yuanyi, Lin Juanjuan. The asymmetric effect of text sentiment on the hedging function of stock index futures[J]. Financial Theory & Practice, 2023(7): 98−107 (in Chinese)
(唐勇, 詹元毅, 林娟娟. 文本情绪对股指期货避险功能的非对称效应研究[J]. 金融理论与实践, 2023(7): 98−107)

[10] Li Shixuan, Shi Wenxuan, Wang Jiancheng, et al. A deep learning-based approach to constructing a domain sentiment lexicon: A case study in financial distress prediction[J]. Information Processing & Management, 2021, 58(5): 102673

[11] Yao Jiaquan, Feng Xu, Wang Zanjun. et al. Tone, sentiment and market impacts: The construction of Chinese sentiment dictionary in finance[J]. Journal of Management Sciences in China, 2021, 24(5): 26−46 (in Chinese)
(姚加权, 冯绪, 王赞钧, 等. 语调、情绪及市场影响: 基于金融情绪词典[J]. 管理科学学报, 2021, 24(5): 26−46)

[12] Jiang Fuwei, Meng Lingchao, Tang Guohao. Media textual sentiment and Chinese stock return predictability[J]. China Economic Quarterly, 2021, 21(4): 1323−1344 (in Chinese)
(姜富伟, 孟令超, 唐国豪. 媒体文本情绪与股票回报预测[J]. 经济学(季刊), 2021, 21(4): 1323−1344)

[13] Agarwal B. Financial sentiment analysis model utilizing knowledge-base and domain-specific representation[J]. Multimedia Tools and Applications, 2023, 82(6): 8899−8920

[14] Popoola G, Abdullah K K, Fuhnwi G S, et al. Sentiment analysis of financial news data using TF-IDF and machine learning algorithms[C]//Proc of the 3rd IEEE Int Conf on AI in Cybersecurity (ICAIC). Piscataway, NJ: IEEE, 2024: 1−6

[15] Ahmad H O, Umar S U. Sentiment analysis of financial textual data using machine learning and deep learning models[J]. Informatica, 2023, 47(5): 153−158

[16] Liu Zhuang, Liu Chang, Wayne Lin. et al. Pretraining financial language model with multi-task learning for financial text mining[J]. Journal of Computer Research and Development, 2021, 58(8): 1761−1772 (in Chinese)
(刘壮, 刘畅, Wayne Lin, 等. 用于金融文本挖掘的多任务学习预训练金融语言模型[J]. 计算机研究与发展, 2021, 58(8): 1761−1772)

[17] Huang A H, Wang H, Yang Y. FinBERT : A large language model for extracting information from financial text[J]. Contemporary Accounting Research, 2023, 40(2): 806−841

[18] Kim J, Kim H S, Choi S Y. Forecasting the S&P 500 index using

mathematical-based sentiment analysis and deep learning models: A FinBERT Transformer model and LSTM[J]. Axioms, 2023, 12(9): 835

[19] Xu Xuan, Wen Fufang, Chu Beilin, et al. FinBERT2: A specialized bidirectional encoder for bridging the gap in finance-specific deployment of large language models[C]//Proc of the 31st ACM SIGKDD Conf on Knowledge Discovery and Data Mining V2. New York: ACM, 2025: 5117−5128

[20] Blythe C. Enhancing financial sentiment analysis with BERT and data augmentation for market impact prediction[J]. Journal of Computer Technology and Software, 2025, 4(7): 1−6

[21] Zhu Tong, Li Leida, Yang Jufeng, et al. Multimodal sentiment analysis with image-text interaction network[J]. IEEE Transactions on Multimedia, 2022, 25: 3375−3385

[22] Li Hongchan, Lu Yantong, Zhu Haodong. Multi-modal sentiment analysis based on image and text fusion based on cross-attention Mechanism[J]. Electronics, 2024, 13(11): 2069

[23] Wang Shunjie, Cai Guoyong, Lu Guangrui. Adversarial soft-contrast modulated dynamic graph learning for multimodal emotion recognition in conversations[J]. Journal of Computer Research and Development, 2026, 63(2): 305−323 (in Chinese)
(王顺杰, 蔡国永, 吕光瑞. 对抗软对比调制动态图学习的多模态会话情绪识别方法[J]. 计算机研究与发展, 2026, 63(2): 305−323)

[24] Zhang Shunxiang, Liu Jiajia, Jiao Yixuan, et al. A multimodal semantic fusion network with cross-modal alignment for multimodal sentiment analysis[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2025, 21(10): 3744648

[25] Ding Chaoyue, Zong Daoming, Li Baoxiang, et al. Multimodal sentiment analysis via efficient multimodal Transformer and modality-aware adaptive training strategy[C]//Proc of the 4th on Multimodal Sentiment Analysis Challenge and Workshop: Mimicked Emotions, Humour and Personalisation. New York: ACM, 2023: 11−17

[26] Huan Ruohong, Zhong Guowei, Chen Peng, et al. Unimf: A unified multimodal framework for multimodal sentiment analysis in missing modalities and unaligned multimodal Sequences[J]. IEEE Transactions on Multimedia, 2023, 26: 5753−5768

[27] Chen Haifeng, Li Xi, Li Yan, et al. Multimodal emotion recognition method based on mixture of fusion experts[J]. Journal of Computer Research and Development, 2026, 63(6): 1478−1491 (in Chinese)
(陈海丰, 李溪, 李岩, 等. 基于混合融合专家的多模态情感识别方法[J]. 计算机研究与发展, 2026, 63(6): 1478−1491)

[28] Liu Ziyu, Yang Tao, Chen Wen, et al. Sentiment analysis of social media comments based on multimodal attention fusion Network[J]. Applied Soft Computing, 2024, 164: 112011

[29] Farimani S A, Jahan M V, Fard A M. An adaptive multimodal learning model for financial market price prediction[J]. IEEE Access, 2024, 12: 121846−121863

[30] Song Lingyun, Chen Siyu, Meng Ziyang, et al. FMSA-SC: A fine-grained multimodal sentiment analysis dataset based on stock comment Videos[J]. IEEE Transactions on Multimedia, 2024, 26: 7294−7306

[31] Todd A, Bowden J, Cummins M, et al. A multimodal sentiment classifier for financial decision making[J]. International Review of Financial Analysis, 2025, 105: 104322

[32] Radford A, Kim J W, Hallacy C, et al. Learning transferable visual models from natural language supervision[C]//Proc of the 38th Int Conf on Machine Learning. New York: PMLR, 2021: 8748−8763

[33] Malik U, Bernard S, Pauchet A, et al. Pseudo-labeling with large language models for multi-label emotion classification of french Tweets[J]. IEEE Access, 2024, 12: 15902−15916

[34] Song Piaopeng, Lu Hanglin, Zhang Yongjie. Unveiling tone manipulation in MD&A: Evidence from ChatGPT Experiments[J]. Finance Research Letters, 2024, 67: 105837

[35] Tellamekala M K, Amiriparian S, Schuller B W, et al. COLD Fusion: Calibrated and ordinal latent distribution fusion for uncertainty-aware multimodal emotion recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(2): 805−822



Li Tingting, born in 1994. PhD, master supervisor. Her main research interests include social media data mining and online public opinion analysis.
李婷婷, 1994 年生。博士, 硕士生导师。主要研究方向为社交媒体数据挖掘、网络舆情分析。



Sun Shouqiang, born in 1995. PhD, lecturer. His main research interest includes multimodal data mining and management.
孙守强, 1995 年生。博士, 讲师。主要研究方向为多模态数据挖掘与管理。



Li Qingqing, born in 1998. PhD. Her main research interests include social media data mining and online public opinion analysis.
李青青, 1998 年生。博士。主要研究方向为社交媒体数据挖掘、网络舆情分析。



Wan Ling, born in 2001. Master candidate. Her main research interests include multimodal data mining, and geographic information and remote sensing technology.
万　玲, 2001 年生。硕士研究生。主要研究方向为多模态数据挖掘、地理信息遥感技术。