

多模态信息整合的互补性和边界 对管理学实证研究的启示

□ 刘 壮 畅 一鸣

领域编辑推荐语：

近年来，情绪对于股票市场的影响日益明显。这篇论文探讨人工智能在评估多模态情绪在股票预测中的影响。在研究方法上具有很强的创新性，特别是关于多模态情绪在股票预测中的效果取决于模态间互补性的观点非常有启发价值。

——程聪

摘 要：在人工智能尤其是大模型及其多模态技术快速发展的背景下，研究者能够更加轻松地从文本、图像、音频等非结构化数据中提取出潜在特征并构造研究指标，多模态融合也逐渐成为管理学研究中的重要信息表征策略。现有研究普遍认为多模态数据的应用和整合可以提升信息表征效果和预测精度，但是现实有时并非如此。本文从信息处理视角出发，提出“多模态并非必然带来更优的信息表征效果，其有效性取决于模态间的互补性结构，而非单纯的数量堆叠”。本文以比亚迪为实证对象，构建了涵盖股票市场、文本、图像与语音的多源数据集，并在相同模型框架下比较不同模态融合效果。研究发现，本案例中多模态未形成有效互补，核心特征权重被稀释，在更换模态组合、融合方式后，新增模态仍未提供任务相关信息增量。本文验证了边界效应的存在，并给出了互补性的形式化定义，为管理学研究中的多模态信息表征、指标构造与研究设计提供了方法启示。

关键词：多模态融合；边界效应；互补性结构；信息表征效果；管理学研究

一、引言

（一）研究背景与问题提出

随着人工智能技术的广泛应用，情绪信息已逐渐成为管理研究中刻画外部信息环境的重要来源。并且随着文本、音频、图像等数据在近年来的快速积累，基于文本等非结构化信息提取情绪特征并据此整合构造研究指标，也日益成为相关管理学

研究中的常见做法（部慧等，2018；石善冲等，2018；许雪晨和田侃，2021；于孝建等，2024；周颖刚等，2024；张飞鹏等，2024；Fazlija & Harder，2022；Ouf et al.，2024）。这一趋势显示人工智能正在对管理研究方法进行深刻重塑，尤其是以大模型和多模态融合技术为代表的生成式人工智能正在以前所未有的速度推动管理决策研究从传统统计范式走向智能化、数据驱动的新方法论上。但上述趋势隐含一个关键问题，即：信息模态的扩展与融合所构造的研究指标是否拥有有效的信息表征效果？

目前，一些研究已开始尝试使用深度学习、预训练模型，直接从财经文本中学习与市场反应相关的信号，而不仅仅停留在简单的情绪极性判断（Devlin et al.，2019；Du et al.，2024；Moradi - Kamali et al.，2024）。与此同时，在通用情感计算领域，融合文本、图像、音频等多模态信息被普遍认为有助于更全面地识别情绪（Das & Singh，2023；Han et al.，2021；Lai et al.，2023），这一趋势也开始进入具体应用研究，已有工作尝试将多种非结构化信息与任务相关数据进行融合分析（Todd et al.，2024a；Wang et al.，2025）。从技术视角看，多模态学习的优势在于不同模态之间可能提供互补信息，从而提升模型的表达能力与预测表现（Baltrušaitis et al.，2019；Zadeh et al.，2018）。因此，大量相关工作显示主流研究普遍认为多模态由于能够获得更丰富的维度信息，相较于单一模态在泛化能力、预测精度、解释性等多方面都要更加优秀。

但是从管理学中的信息处理视角来看，信息整合并非简单的数量累加过程，在实际应用

过程中多模态的表现并不稳定，很多时候它的表现力要低于单模态。例如，Simon 指出，信息丰富往往伴随着注意力稀缺（Simon，1971），Eppler 和 Mengis 也指出，过量信息可能导致信息过载，进而降低决策质量（Eppler & Mengis，2004）。换言之，更多的信息并不必然带来更高质量的判断。特别是在各个模态之间来源不同、同质化严重、信息维度混杂等多方面下，模型最后很可能由于冗余的信息导致有效特征被稀释，导致模型的性能下降（Kiela et al.，2019；Liang et al.，2018；Ngiam et al.，2011）。

因此，目前“多模态整合后的特征具有更有效的信息表征效果”这一论题值得被继续谈论探究。本文通过收集到的文本、图像、音频等数据结合大模型来获得多模态情感信息，再与市场信息对齐融合后通过双通道 LSTM 来预测单一股票的走势、投资策略的制定、可解释性分析等，并与单一模态文本下得到的结果进行对比，发现多模态不仅没有形成有效的信息表征，反而造成了预测精度的降低、关键特征的稀释以及模拟投资下的亏损。这一现象表明：信息整合的有效性并非单纯由模态数量决定，模态互补性比模态数量更为关键。这一结论揭示了多模态信息整合的边界效应，同时为研究者在管理学研究上进行信息筛选与变量测量提供了有意义的实证反思。与此同时，可解释人工智能研究也不断强调：在高风险管理决策场景中，模型不应只追求准确率，还应关注其可解释性与适用边界（Yeo et al.，2023）。

本文选择股票市场作为研究情境，主要基于以下考虑。一方面，金融市场是一个信息高度密集且反应迅速的决策环境，市场参与者需

要在短时间内整合不同来源的信息（如市场数据、媒体信息等）并作出决策，因此为研究多模态信息整合机制提供了典型场景。另一方面，媒体情绪、新闻信息等非结构化信息对市场行为的影响已在既有研究中得到广泛关注，使得该情境能够为检验多模态信息整合的有效性提供较为丰富的数据基础。因此，以股票预测任务作为研究对象，有助于从信息整合视角探讨多模态信息融合的边界效应。

（二）研究目标与核心贡献

针对上面提出的“信息和模态的融合扩展具有有效的信息表征效果，以及其有效性所依赖的结构条件”的问题，本研究用比亚迪 2019—2024 年的股票市场作为实证对象，用人民日报中关于比亚迪的相关新闻作为文本信息，用比亚迪发布会、研讨会的视频作为图像信息和音频信息的来源，采用双通道 LSTM，系统对比了多模态和单模态在这些条件下模型的预测能力、特征贡献性分析等，基于此，本研究重点探讨以下核心问题：（1）信息来源（模态）的扩展是否真的能够形成更有效的信息表征效果？（2）在何种条件下多模态会出现性能下降？（3）模态数量与互补性如何影响模型预测能力？（4）如何形式化界定信息互补性，并据此判定多模态整合的适用边界？

为解决上述问题，本研究的主要贡献在于：

（1）从信息处理的角度出发，修正了“信息模态的扩展能够改进信息表征效果”这一隐含假设，通过实证验证了一味地增加模态数量并不会使模型效果提高，反而可能会下降，强调了多模态融合的有效性取决于模态间是否形成了有效互补，而非简单的在数量维度的信息

堆叠。

（2）给出了模态互补性的形式化界定，综合“信息非冗余性、任务相关增量与融合可利用性”三点作为互补性的判定标准（互补性三条件），明确了多源信息整合存在的边界条件。

（3）构建了一个多模态的多维结构数据，通过新闻、语音、图像数据情感分析，与比亚迪股票信息进行对齐，得到了一个全方位的多模态股票市场数据。通过系统性的对照实验、可解释性分析、稳健性检验、消融实验等从实证角度验证了信息整合的边界效应的存在，以及互补性对信息表征效果的影响路径。

（4）将多模态融合问题从单纯的技术性能比较，提升为信息表征与指标构造的适用边界问题，提出基于互补性判定的多源信息筛选框架，为研究者在管理学研究中应用多模态方法进行信息表征、指标构造与研究设计提供了方法启示。

二、相关工作

近年来，情绪信息逐渐被纳入金融与管理研究之中，成为刻画外部信息环境的重要来源。大量研究显示无论数据是来自新闻还是社交媒体都能够反映出投资者情绪和股市行情存在较强的正相关关系（部慧等，2018；石善冲等，2018；许雪晨和田侃，2021；于孝建等，2024；朱莉等，2022）。例如，张宸瑞（2022）通过深度学习技术挖掘出文本情绪和股价之间的关联，张飞鹏等（2024）的研究也验证了将新闻文本情感加入预测模型中能够提供有价值的增量信息，同时还有研究通过情感挖掘来实现交易策

略的制定和资产配置（林建浩等，2022；孟勇等，2022），尹海员和吴兴颖（2023）则是在情绪基础上又引入了订单流不均衡；同样，在英文文献也有这种类似的研究（Muhammad et al.，2024）：Fazlija 和 Harder（2022）基于情感分析将新闻中的情绪因子进行量化，对股价涨跌方向进行了有效预测，并取得了优于其他基准模型的精确度；Ouf 等（2024）利用 Twitter 社交媒体文本中的情绪结合深度学习实现了基于社交情绪的时间序列建模并显著降低了预测误差，这些文章显示文本情绪是股票市场的一个重要信息源，其所包含的情绪特征能够与其他金融数据相结合，被进一步转化为研究指标（Kielbaso et al.，2019；Ngiam et al.，2011）。周颖刚等（2024）将市场情绪应用到资产定价中，这些研究表明情绪信息作为一种外部信息能够提供传统市场信息所不具备的额外信息，为管理学研究以及组织决策等提供增量信息。

早期金融情感分析以词典规则或机器学习为主（徐鹏和尚维，2022；姚加权等，2021；张宗新和吴钊颖，2021）。而近年来由于深度学习和预训练模型的出现，情感分类的精准度显著提高（Devlin et al.，2019；Jiang et al.，2020；Muhammad et al.，2024），研究者从非结构化文本中提取情绪与语义特征的技术门槛显著降低。Du 等（2024）对金融情感分析技术进行了综述，指出如 LSTM、Transformer 这样的神经网络已被大量地应用在新闻情绪分析中，并利用得到的情绪做行情预测等（Devlin et al.，2019；Sang & Li，2024；Wang，2023），尤其像 BERT 框架下的预训练模型大大提升了大模型对各种金融文本的情绪识别的能力并降低了使用

门槛（Devlin et al.，2019），越来越多的研究人员设计了一些针对金融市场的特定情感分类器来提取乐观/悲观的情绪信号，从而进行金融方面的预测，并取得较好的效果（张飞鹏等，2024）。例如，Liagkouras 和 Metaxiotis（2025）采用情绪指标混合 LSTM 预测框架的方式证明增加新闻情绪指标的混合模型，可以提升整体预测精度，而 Li 等则采用 Transformer 架构进行预测（Li et al.，2023）。不仅如此，还有一些学者利用股票论坛、Twitter 中的评论和推文等数据，探索在主流社交媒体中的情感因子，再用来进一步建模（Liagkouras & Metaxiotis，2025；Usmani & Shamsi，2021）。总体而言，将社会舆论或者个人观点中的情感融入量化决策之中已成为一种趋势：不管是财经新闻文本（张飞鹏等，2024；Du et al.，2024），还是社交媒体平台如 Twitter 的消息都可以通过情感量化作为模型的输入并提高对市场未来预测的效果（Liagkouras & Metaxiotis，2025；Usmani & Shamsi，2021），其所包含的情绪信息都越来越多地被转化为可用于研究的指标变量。随着信息处理能力的提升，管理学研究者能够处理更多非结构化数据，这为进一步拓展信息表征方式提供了技术基础，但同时也提高了信息整合与指标构造的复杂性。

在信息处理能力提升的基础上，现阶段，相关研究开始将目光转向更加丰富的数据来源和更复杂的模型架构，以期从非结构化信息中提取更深层次的情绪与语义特征。何勇等提出了用大型预训练模型加上一些另类数据如新闻、公告等与传统金融数据来赋能量化投资，这显示学界试图用当下最领先的人工智能工具去捕

获金融市场情绪（何勇等，2024）。一些最新的研究尝试利用大语言模型对财经新闻做出深度解读并从中得到更深层次的情感信息以及对更复杂性的事件进行分析。例如，Moradi - Kamali 等（2025）重新审视了金融情感分析的各种方法，采用领域语言模型直接基于文本和市场反应训练预测模型，显示出相较传统情绪指标方法更强的表征能力。而 Zhang（2023）也利用一系列大语言模型深入解读相关报道中蕴含的情绪信息，并将其用于构建投资组合，说明大模型能够支持研究者从文本中提取更丰富的语义信号。这些研究表明，随着深度学习和大模型的发展，情感分析在金融领域的应用已从简单的情绪极性判断拓展到更深层的语义理解与情境解释。Yeo 等（2023）的综合评述也强调了金融领域可解释人工智能的重要性，提示相关研究不应仅关注模型表现，还应关注其可解释性与适用边界。这一转变说明研究者对非结构化信息的利用正在由“情绪识别”走向“语义理解与情境解释”，信息表征的层次与复杂度显著提升。情感分析已经从简单的极性分析升级为了更加复杂的信息表征工具，但是这种进步带来了一个新的问题：当研究者可以整合更丰富的数据来源时，如何确保这些新增信息能够形成有效的结构增益，而非增加信息整合的复杂度与噪声风险？

除文本以外的多模态情感信息也开始受到关注（Vallarino, 2025; Yang et al., 2025），在通用情感计算领域，融合文本、图像、音频等多模态信息已被证明有助于更全面地识别情绪（Baltrušaitis et al., 2019; Han et al., 2021）。例如，通过分析语调（音调、语速、升调）、面

部表情以及话语内容，有助于更全面地刻画情绪特征（Das & Singh, 2023; Lai et al., 2023）。但在金融问题中，这类将多个模态结合起来的情绪研究仍相对匮乏。大多数研究还是使用单一文本数据，而不是使用图像或语音等多模态，少数研究融合了结构化的市场数据与文本情绪（可视为双模态）（Kiela et al., 2019; Wang et al., 2025），或者是结合了新闻和社交媒体多源文本情绪（Ngiam et al., 2011），而结合视觉甚至包括语调这样的非文本因素来进行金融情感分析的研究少之又少。已有少量研究开始尝试将文本、音频等多模态信息引入金融情感分析与趋势预测之中，如 Todd 等（2024a）、Wang 等（2025），但现有证据尚不足以支持多模态在不同任务中必然稳定优于单模态，其效果仍可能受到新增信息质量、任务相关性与融合方式的影响。这一现象的来源可能正是由于模态间缺乏有效的互补结构，信息的扩展反而带来更多的噪声，因此，在管理学研究中引入多模态信息时，需要特别关注不同模态之间所存在的结构性与条件性边界。

已有研究印证了上面的判断。Han 等（2021）在情感分析融合研究中指出，未经筛选的模态简单拼接可能导致模型关注偏移，需要通过相关性控制等方法确保各模态贡献互补信息。可见，多模态融合的功效并不取决于信息来源数量本身，模态间的互补性是多模态情感分析能否形成有效增量的关键之一。

总体而言，现有研究在多源数据融合上主要采用早融合和晚融合两种方式，早融合就是在输出层将多源数据进行直接拼接，而晚融合则是先对各个模态进行独立建模，最后在决策

层进行融合 (Muhammad et al., 2024), 前者有例如将文本情绪因子和市场因子合并输入到深度学习模型中 (Wang et al., 2025) 的例子。部分实验尝试通过引入注意力机制或者协同网络来提升模型的交互效率 (张飞鹏等, 2024; Sang & Li, 2024), 但是这些研究都普遍默认多模态能够提升模型的表现力, 却没有对融合有效性条件进行理论分析或实证检验 (如模态之间是否存在信息冗余, 是否存在互补性等), 正因为这个原因, 才导致现有的金融多模态工作在对模型性能的提高上有限甚至退化, 因此对模态互补性的研究是当前文献的主要不足, 也是本文重点探讨的核心命题。

综上所述, 现有文献的不足在于: 一方面, 金融情感分析大多局限于单一文本模态, 对于图像、语音等其他潜在情绪载体的作用缺乏研究; 另一方面, 已有少数尝试多模态融合的工作往往默认“信息越多越好”, 而较少深入探讨不同模态信息的重叠与干扰问题。当前研究尚缺乏对“信息模态的扩展与融合是否能够形成更有效的信息表征效果?”这一问题的系统检验。尤其是在模态信息存在冗余的情况下, 额外模态可能提供的边际增益甚微, 甚至因噪声引入导致模型过拟合。这种情形在文献中未有充分的实证讨论。为填补上述空白, 本文将多模态融合问题置于信息整合框架之中, 拟通过对比单模态与多模态情绪的融合效果, 检验“多模态并非必然带来更优的信息表征效果, 其有效性取决于模态间互补结构, 而非单纯的数量堆叠”这一命题, 从而丰富金融多模态情感分析领域的研究视角。通过系统的实证分析, 本文旨在揭示在多源信息环境下如何实现结构

性增益, 并为研究者在管理学研究中进行信息筛选、指标构造与研究设计提供理论依据。

三、研究方法与实验设计

(一) 研究总体框架

为系统检验“多模态并非必然带来更优的信息表征效果, 其有效性取决于模态间互补结构, 而非单纯的数量堆叠”这一核心命题, 本研究设计了两个模型, 一个单模态基线模型以及一个多模态模型, 意在通过对比两个模型的实验结果来验证命题的真实性。本研究构造了一个完整的实证对比流程, 主要流程包括数据收集与特征提取、情感分析模型训练、多模态情感分析、特征融合与预测建模、结果分析五个主要环节, 如图1所示。

首先, 收集本次实证分析的主要对象, 比亚迪在2019—2024年六年间相关的每日新闻和每月视频, 构建出包含文本、视频两种模态的多源数据集, 值得注意的是, 为节省计算资源, 视频这个模态的数据在后续被分解为图像以及语音两个模态, 同时, 由于视频是月度数据, 因此在情感分析后通过艾宾豪斯记忆曲线对这两个模态的日度数据进行插值。其次, 通过Tushare收集到比亚迪在2019—2024年期间的股票市场数据 (开盘价、收盘价、涨跌额等), 用于为LSTM模型提供市场特征。

在情感分析过程中, 对于文本的情感分析, 本研究首先训练了一个专门用于分析新闻情感的模型, 然后将每天与比亚迪相关的新闻进行情感分析, 得到每条新闻的情感得分 (Positive 趋向于1, Negative 趋向于0), 最后聚合每天的

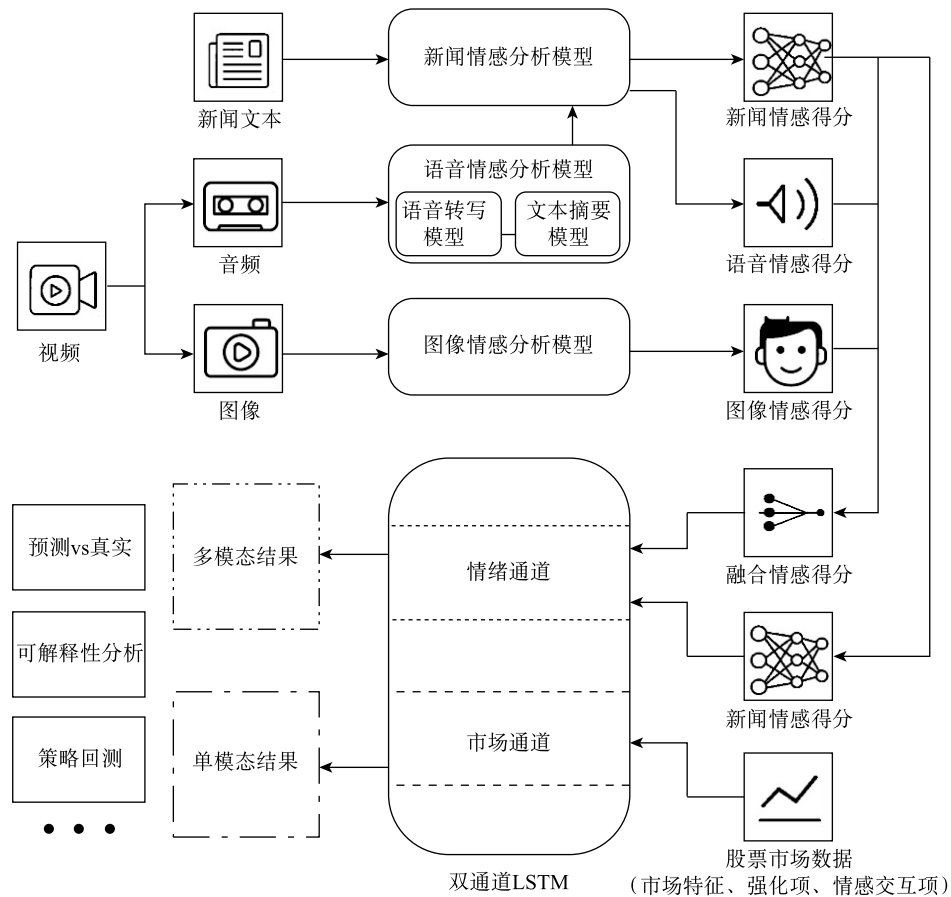


图1 方法论框架概述

情感得分，得到按时间顺序的“文本情感得分数据”（*people_daily_sentiment*）；对于图像情感分析，由于相关数据收集难度较大，直接采用训练好的 Vision Transformer 模型（trpakov/vit-face-expression）用于人物图像情感分析，通过对视频进行抽帧并识别人物脸部提取出人物图片，然后输入到图像情感分析模型中并取均值得到情感得分，后通过记忆曲线公式对其余日期进行插值，得到“人物图像情感分析数据”（*image_sentiment*）；对于语音情感分析，语音同样采集自视频，运用 OpenAI 的开源的语音转写模型 whisper 先将视频中的语音转化为文本，再由 mT5 摘要模型（csebuetnlp/mT5_multilingual-XLSum）对提取出的大文本进行摘要总结，最

后将摘要输入到自己训练的新闻情感分析模型中进行打分，同样使用记忆曲线公式对其余日期进行插值，得到“语音情感分析数据”（*audio_sentiment*）。

在预测中，本研究设计了一个双通道 LSTM 模型，一个市场通道（市场原有指标，*ma_3* 等强化项，情感交互项）、一个情绪通道，分别通过隐藏层，进行融合之后通过 Linear 进行映射，预测未来的股票走势。

在结果分析中，分别能够得到单模态和多模态的预测结果、可解释性分析结果以及策略回测这一实际应用场景的结果，通过对二者得到的结果进行对比，揭示“模态的堆叠不一定提高模型性能，模态互补性比模态数量更关键”

的实证结论。

(二) 互补性的形式化界定

直接检验不同模态特征之间的相关性是最基础、最常用的确定模态间是否互补的方法,通过相关性检验能够初步判断多模态信息间是否能够实现互补而非冗余。若相关性较高,说明二者在一定程度上存在共享的信息,可能会造成信息冗余,引入多余噪声;如果相关性较低,则意味着模态提供的是差异化信息结构,能够产生信息的增益。这也是多模态研究普遍认为“低相关性模态更可能带来性能提升”的理论基础。

但是低相关性并不一定意味着信息互补一定成立,其本质上只能揭示信息差异性,而非信息的实际增益。模态之间的信息是否真正与任务目标相关(任务相关信息增益),以及模型能否有效整合并利用这些差异化信号(融合机制可利用性)才是更深层的决定性因素。

为了增强理论的严谨性,本文将多模态的互补性形式化为三个必要条件,具体为:(1)模态间应具备较小的相关性,即信息非冗余;(2)需要具备任务相关信息增益;(3)融合机制可利用性。

也就是要满足:

$$\text{Complementarity} = C_1 \wedge C_2 \wedge C_3 \quad (1)$$

1. 信息非冗余性

多模态互补性的首要前提和第一判断标准就是模态间应存在较少的信息冗余,能够提供多样化的信息,在形式上可以表示为:

$$C_1: I(A; B) < \tau_r \quad (2)$$

在实际操作过程中,可以利用模态间的相关性作为操作化指标:

$$\hat{C}_1: |Corr(A, B)| < \tau_r \quad (3)$$

如果 A 和 B 两个模态间的相关性低于一定的阈值,则认为二者实现了信息非冗余,具备了模态互补性的可能性。类似的,对于 n 个模态来说,通过计算模态之间的成对相关性矩阵就可以获得整体的信息结构差异,从而判断是否具备模态互补的可能性。

2. 任务相关信息增益

在具备信息较少冗余的前提下,还需要具备对相关任务具有增益效果,具体来说可以表示为:

$$C_2: I(Y; B|A) > \tau_g \quad (4)$$

即在已知模态 A 后,模态 B 对预测目标 Y 是否提供“额外的任务相关信息”。

注意,这一条件属于后验条件,也就是判断实验在模态 A 的条件下引入模态 B 等是否产生了“性能增量”:

$$\hat{C}_2: Perf(A + B + \dots) - Perf(A) > \tau_p \quad (5)$$

3. 融合机制可利用性

除了 C_1 这一先验条件和 C_2 这一后验条件,还需要满足方法论层面的可利用性条件,也就是融合机制要能够有效整合模态间差异性信息,而不会引入多余噪声、对齐困难、过大提高训练成本等:

$$C_3: \Phi(A, B, \dots) \text{ is effective} \quad (6)$$

(三) 数据来源与预处理

1. 新闻文本数据

新闻文本数据来自“人民日报”历年的新闻,通过关键词检索收集比亚迪相关新闻数据,最终共计 10153 条新闻,本研究考虑到了股票市场的时效性,最终选择以日为单位进行后续数据的构建,每个交易日有多条新闻,在后面

经过自己训练的新闻情感分析模型进行情感分析后进行聚合并得到每日的文本情感得分 (*people_daily_sentiment*)。

2. 视频数据

视频来自权威主流媒体, 内容包含了比亚迪的发布会、发展会等, 共计 72 个视频, 原因是重要的发布会等不会每天都有, 因此采用月度数据, 后续通过艾宾豪斯记忆曲线进行日度数据的扩充。为了节约计算资源, 并验证“模态互补性比模态数量更关键”这一实证结论, 视频数据会分解为两个模态 (人物图片和语音), 其中图片数据通过利用 ffmpeg 进行逐帧采集 (每秒一帧), 然后利用人脸检测网络 (MTCNN) 进行人脸检测并进行裁剪, 经过 ViT 模型得到每个视频提取出的人物图片情感得分并取均值, 最后记忆曲线公式插值到每日, 得到日度数据 (*image_sentiment*); 语音数据通过 OpenAI 的开源语音转文本模型 whisper 进行转换, 然后输入到 mT5 的摘要模型中进行总结以缩短文本长度, 短文本输入到自己训练的情感分析模型中进行打分, 同样记忆曲线插值到每日 (*audio_sentiment*)。

3. 市场交易数据

运用股票数据接口 Tushare 获取“比亚迪”2019—2024 年的股票日线数据, 数据特征包括日期 *trade_date*、开盘价 *open*、最高价 *high*、最低价 *low*、收盘价 *close* 等, 这些数据经过筛选后与情绪得分拼接得到最终的输入到 LSTM 中的数据。在原有的股票市场信息的基础上, 增加强化项, 例如最近 3 日收盘价的平均值 *ma_3*、收益率 *return_1* 等。同时增加情感交互项, 对于多模态, 就将融合情感得分^① (*daily_sentiment_*

score) 与市场主要特征收盘价 *close*、成交量 *vol*、涨跌幅 *pct_chg* 进行交互项的构造。

具体来说, 强化项有如下几项: (1) 移动平均线 $MA_k(t)$ 即 *ma_k*: 最近 k 日收盘价的平均值, 如 $MA_3(t) = \frac{1}{3} \sum_{i=0}^2 P_{t-i}$, $MA_5(t) = \frac{1}{5} \sum_{i=0}^4 P_{t-i}$, $MA_{10}(t) = \frac{1}{10} \sum_{i=0}^9 P_{t-i}$; (2) 收益率 $return_1: r_t = \frac{P_t - P_{t-1}}{P_{t-1}}$; (3) 成交量变化率 $vol_chg: v_t = \frac{V_t - V_{t-1}}{V_{t-1}}$ 其中 V_t 是 t 日成交量; (4) 融合情绪 3 日均值: $SMA_3(t) = \frac{1}{3}(S_t + S_{t-1} + S_{t-2})$ 。

交互项有: $S_t \cdot P_t$ (情绪 × 收盘价)、 $S_t \cdot V_t$ (情绪 × 成交量) 和 $S_t \cdot R_t$ (情绪 × 涨跌幅)。

4. 数据对齐与缺失值处理

为保证多模态特征在时间维度上的一致性, 本研究采用以交易日为基准, 将三个模态独立的日度情感得分和日度融合情感得分与市场特征按交易日对齐, 休息日的情绪得分直接去除。若某日没有情绪得分, 则使用前一天的进行补齐。

(四) 文本情绪分析

本研究首先针对新闻文本构建金融情感分析模型。采用预训练模型 rbt3 (中文 BERT 全词掩码模型, 基本架构见图 2) 作为预训练模型, 对“比亚迪”相关的财经新闻进行情感分类。

^① 融合情感得分经过线性回归调权重, 得到最优的近似权重比是 1:1:3, 同时采用标准化后的三个模态情感得分, 也就是 $daily_sentiment_score = 0.2 \times image_sentiment_z + 0.2 \times audio_sentiment_z + 0.6 \times people_daily_sentiment_z$ 。具体求解过程详见附录 C.1。

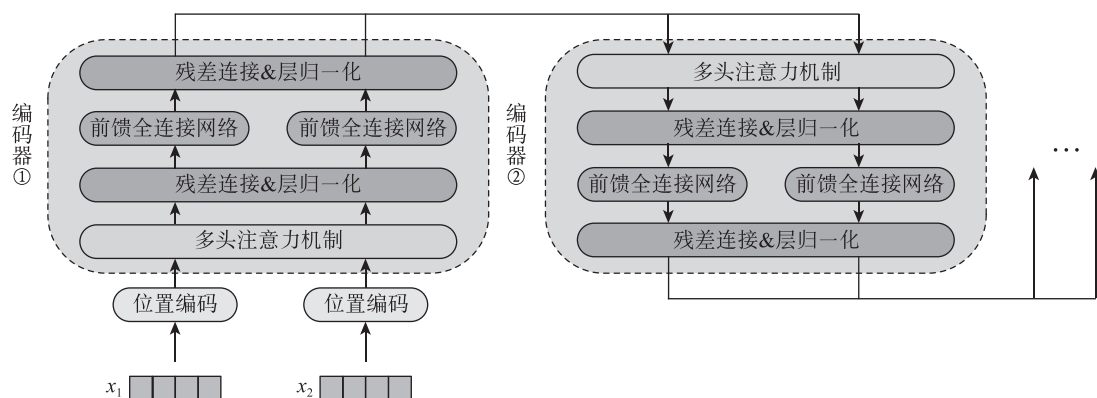


图2 Bert 基本框架 (Transformer Encoder - only)

由于手工标注大规模金融新闻数据代价高耗时久,因此在第二次情感分析模型训练前设计了弱监督标注机制获取训练样本:先利用少量已标注的金融新闻数据(Dataset - of - financial - news - sentiment - classification 金融新闻情感分析数据集)以及公开情感语料(ChnSentiCorp 酒店评论数据集)混合训练初始模型。然后将初始模型大概率预测的大量金融新闻样本作为伪标签数据进行再训练,从而扩充财经新闻情感分类的训练集并提高模型在金融领域的适应性。模型输入为新闻文本,输出为二分类概率分布,表示新闻的正面或负面情绪概率。训练过程中使用交叉熵损失函数优化参数,其二分类形式为:

$$L = -\frac{1}{N} \sum_{i=1}^N [y_i \log p_i + (1 - y_i) \log(1 - p_i)] \quad (7)$$

其中, $y_i \in \{0, 1\}$ 为真实情感标签(0 = 消极, 1 = 积极), p_i 为模型预测新闻为积极的概率。

训练完成后,对于每一条新闻,将模型输出的“积极”概率记为该新闻的情感得分 S (取值 0 ~ 1)。例如,若模型输出 $(p_{neg}, p_{pos}) = (0.2, 0.8)$,则 $S = 0.8$ 表示该新闻偏正面。与此同时,可以取 $\arg\max$ 来确定情感标签(例如

$S \geq 0.5$ 判为正面),本研究使用的是评分而非二分类标签。通过将一天内所有新闻的情感得分取均值,计算得到日情绪得分 $S_t^{text} = \frac{1}{M} \sum_{j=1}^M S_{t,j}$ (M 为 t 日新闻数量)。

对于文本情绪分析方法,关键步骤在于:

(1) 先利用大量公开数据集对中文预训练模型 rbt3 进行情感分析训练,得到初步的情感分析模型;

(2) 为缓解金融新闻人工标注成本高的问题,将新闻文本输入到初步训练的模型中进行情感分析并打上伪标签,仅保留置信度高于阈值 τ 的样本作为伪标签训练集,在该伪标签训练集上再次微调 rbt3;

(3) 利用得到的伪标签数据作为训练集训练得到最终的新闻情感分析模型。

(五) 其他模态情绪分析

1. 图像情绪分析

图像情绪分析的主要模型是基于 ViT 训练的 vit - face - expression 模型(ViT 基本框架见图3),专门用于人物的面部情感识别,最终可以输出属于七个情绪类别的概率分布:

$$P = \{p_{happy}, p_{surprise}, p_{neutrals}, p_{sad}, p_{angry}, p_{disgust}, p_{fear}\} \quad (8)$$

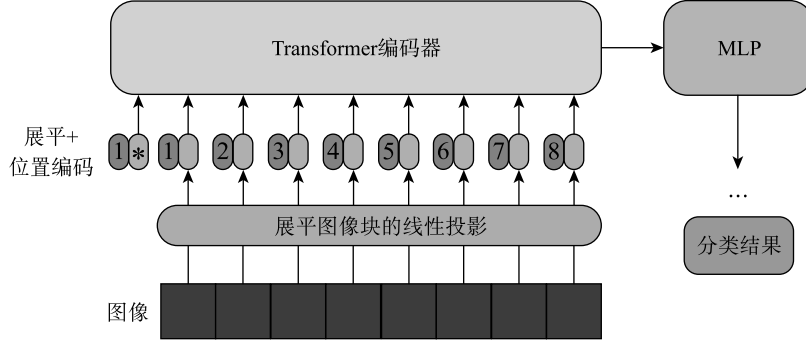


图3 ViT 基本框架

为了将离散概率结果聚合为一个最终图像情感得分，下面利用映射函数 $\phi: c \rightarrow [0, 1]$ 对每个类别进行了积极程度的数值映射，使最终的情感得分位于 $[0, 1]$ ：

$$\phi(c) = \begin{cases} 1.0, & c = happy \\ 0.8, & c = surprise \\ 0.5, & c = neutral \\ 0.2, & c = sad \\ 0.1, & c = angry \\ 0.0, & c = disgust \text{ 或 } fear \end{cases} \quad (9)$$

每张图像的情绪得分计算公式为：

$$S_{image} = \sum_{c \in c} p_c \cdot \phi(c) \quad (10)$$

至于图片的来源，图片都是源自收集到的每月比亚迪发布会等的公开视频，通过抽帧以及 MTCNN 人脸识别获取多张人脸图像，对每张图像独立计算 S_{image} ，并取平均值作为该视频的图像情绪得分：

$$S_v^{image} = \frac{1}{N_v} \sum_{j=1}^{N_v} S_{image,j} \quad (11)$$

其中， N_v 表示该视频中被抽取并识别的人脸图像数量。

由于视频数据为月度频率，而后续需要日度数据来与市场特征数据进行拼接，因此采用

艾宾豪斯记忆曲线对图像情绪得分进行时间插值，得到每日图像情绪得分数据：

$$S_t^{image} = \sum_{k=1}^K \alpha_k \cdot S_{v_k}^{image} \cdot e^{-\beta(t-t_k)} \quad (12)$$

其中， $S_{v_k}^{image}$ 为第 k 条视频的图像情绪得分， t_k 为该视频发布日期， α_k 为归一化系数， β 为遗忘衰减参数，最终得到的 S_t^{image} 表示第 t 日的图像模态情绪得分序列。

2. 语音情绪分析

针对新闻视频中的音频模态，本研究构建了基于语音转写、长文本摘要和文本情绪识别相结合的间接语音情绪分析方法，语音情绪分析过程涉及三个模型，分别是语音转写模型 Whisper、摘要模型 mT5 以及文本情绪分析中自己训练的基于 rbt3 的文本情感分析模型。其中，Whisper 和 mT5 均是完整的 Transformer Encoder - Decoder 框架，而 rbt3 是 BERT 模型，它采用的是 Transformer Encoder - only 框架。

具体过程如下，首先利用 OpenAI 的开源语音转写模型 Whisper 对视频的语音内容进行自动转写（Automatic Speech Recognition, ASR）：

$$T_v = \text{Whisper}(A_v) \quad (13)$$

由于视频的大量冗余信息以及视频中复杂

的语音交流环境，直接进行情感分析显然会对结果造成不利的影响，因此，在中间加入了一个摘要模型 mT5 - multilingual - XLSum，对得到的长文本内容进行总结去噪声：

$$\tilde{T}_v = \text{mT5}(T_v) \quad (14)$$

随后，将摘要文本 $\tilde{T}_v$ 输入前述训练好的新闻情绪分析模型，计算该语音片段对应的情感得分：

$$S_v^{\text{audio}} = f_{\text{BERT}}(\tilde{T}_v) \quad (15)$$

同样的，由于是月度数据，需要用艾宾豪斯记忆曲线进行时间插值：

$$S_t^{\text{audio}} = \sum_{k=1}^K \alpha_k \cdot S_{v_k}^{\text{audio}} \cdot e^{-\beta(t-t_k)} \quad (16)$$

3. 多模态情感融合

在本研究中，情绪得分来自文本、图像和语音三个模态。本研究采用线性回归在训练集上求解三个模态的最优权重，从而固定融合权重比例。设训练集中共有 T 天数据，每天的真实目标情绪值为 y_t （由人工标注）。构建回归矩阵：

$$X = \begin{bmatrix} S_1^{\text{text}} & S_1^{\text{image}} & S_1^{\text{audio}} \\ \vdots & \vdots & \vdots \\ S_T^{\text{text}} & S_T^{\text{image}} & S_T^{\text{audio}} \end{bmatrix}, w = \begin{bmatrix} w_{\text{text}} \\ w_{\text{image}} \\ w_{\text{audio}} \end{bmatrix} \quad (17)$$

通过最小化残差平方和得到最佳近似权重比，且权重满足 $w_{\text{text}} + w_{\text{image}} + w_{\text{audio}} = 1$ 。

注意，在本研究中，由于很难获得大量人工标注的实际目标情绪值，因此仅采用少量样本进行手动线性回归，得到的图像、音频和文本模态的最优权重比是1:1:3，为了避免少量数据带来的偏差，还尝试了其他权重比，最终验证出确实若要得到多模态最佳结果，采用1:1:3

是合理的。

在得到权重之后，日度多模态情绪得分：

$$S_t^{\text{fuse}} = w_{\text{text}} z S_t^{\text{text}} + w_{\text{image}} z S_t^{\text{image}} + w_{\text{audio}} z S_t^{\text{audio}} \quad (18)$$

其中， $z S_t^{\text{text}}$ 、 $z S_t^{\text{image}}$ 、 $z S_t^{\text{audio}}$ 是标准化后的三个模态的情感得分。

(六) 双通道 LSTM 预测模型设计

1. 双通道设计

为了融合市场技术指标和媒体情绪信息，提高股价预测准确度，本研究中设计了双通道 LSTM 模型（dual-channel LSTM）。模型结构包括两个并行的 LSTM 网络通道：市场通道接收价格、成交量等市场特征序列，情感通道接收情绪得分序列。具体而言，两个通道均采用标准 LSTM 网络结构，通过输入门、遗忘门和输出门实现对时序信息的选择性保留与更新，各通道所采用的 LSTM 基本框架如图4所示。在每个交易日 t ，将过去 T 天的市场特征（包括强化项和交互项） $\{x_{t-T+1}^{(m)}, \dots, x_t^{(m)}\}$ 和情绪特征 $\{x_{t-T+1}^{(s)}, \dots, x_t^{(s)}\}$ 分别构成长度为 T 的时间序列输入两个并行的 LSTM 中。两个 LSTM 分别沿时间步迭代更新其隐藏状态 $\mathbf{h}_k$ 和记忆单元 $\mathbf{c}_k$ 。设 $\mathbf{h}_{k-1}$ 为上一时刻隐藏状态， $\mathbf{x}_k$ 为当前时刻输入向量，则 LSTM 在时间步 k 的更新公式为：

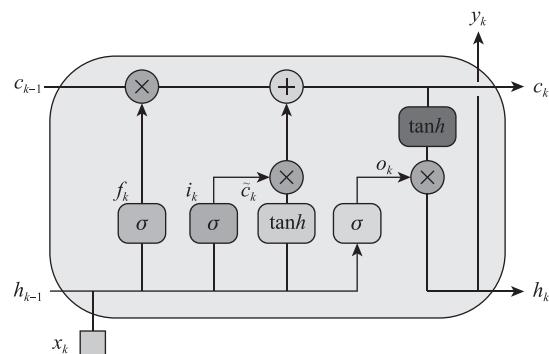


图4 单通道 LSTM 基本框架

$$\begin{cases}
\text{遗忘门: } f_k = \sigma(W_f [\mathbf{h}_{k-1}, \mathbf{x}_k] + b_f) \\
\text{输入门: } i_k = \sigma(W_i [\mathbf{h}_{k-1}, \mathbf{x}_k] + b_i) \\
\text{输出门: } o_k = \sigma(W_o [\mathbf{h}_{k-1}, \mathbf{x}_k] + b_o) \\
\text{候选记忆: } \tilde{\mathbf{c}}_k = \tanh(W_c [\mathbf{h}_{k-1}, \mathbf{x}_k] + b_c) \\
\text{状态更新: } \mathbf{c}_k = f_k \odot \mathbf{c}_{k-1} + i_k \odot \tilde{\mathbf{c}}_k \\
\text{隐藏状态: } \mathbf{h}_k = o_k \odot \tanh(\mathbf{c}_k)
\end{cases} \quad (19)$$

其中, $\sigma(\cdot)$ 是 Sigmoid 激活函数, $\tanh(\cdot)$ 是双曲正切函数, W_f, W_i, W_o, W_c 及 b_f, b_i, b_o, b_c 为各门控的权重矩阵和偏置向量, 符号 $\odot$ 表示按元素乘积。经过上述递归计算, 每个 LSTM 在最后一个时间步 T 输出其最终隐藏状态向量: 市场通道得到 $\mathbf{h}_T^{(m)}$, 情感通道得到 $\mathbf{h}_T^{(s)}$ 。融合机制采用向量拼接将这两个通道的隐藏向量进行合并:

$$\mathbf{h}^{fusion} = [\mathbf{h}_T^{(m)} \parallel \mathbf{h}_T^{(s)}] \quad (20)$$

即将长度为 d_m 的 $\mathbf{h}_T^{(m)}$ 与长度为 d_s 的 $\mathbf{h}_T^{(s)}$ 连接为新的特征向量 (在本实验中 $d_s = d_m / 2$, 即市场 LSTM 隐层维度 256 和情感 LSTM 隐层维度 128)。随后, 融合向量 $\mathbf{h}^{fusion}$ 通过输出层的前馈神经网络进行回归预测: 首先经过一个全连接隐含层将维度从 $d_m + d_s$ 变换到 d_m , 并使用 ReLU 激活函数及 Dropout 正则化, 最后经由线性层映射为标量 $\hat{y}_{t+1}$ 。这个输出 $\hat{y}_{t+1}$ 就是模型预测的下一交易日收盘价。模型训练以历史数据的实际收盘价 y_{t+1} 为目标, 采用均方误差损失函数来度量预测误差:

$$L = \frac{1}{N} \sum_{j=1}^N (\hat{y}_j - y_j)^2 \quad (21)$$

其中, N 为训练样本数量, $\hat{y}_j$ 和 y_j 分别为模型预测值和真实值。通过反向传播优化, 使用 Adam 优化器 (学习率为 $1e-4$) 迭代更新参

数, 使预测值逼近真实值。训练完成后, 在验证集上选择并保存损失最小的模型作为最终模型用于后续策略回测和解释。

值得注意的是, 在使用双通道 LSTM 时, 我们将两个通道的隐藏状态拼接来进行融合, 这种向量拼接的方式可能导致维度膨胀和过拟合风险, 特别是在模型的复杂度较高时, 特征维度的增加可能导致计算量的增加, 同时容易引发过拟合。因此, 在实验中我们引入了 Dropout 技术, 对网络中部分神经元进行随机丢弃, 以防止模型对某些特定特征的过拟合, 同时加入了早停机制, 避免模型的过度训练。

2. 集成梯度可解释性分析

为了探究模型决策的依据, 本研究使用 Integrated Gradients (集成梯度) 方法对双通道 LSTM 模型进行可解释性分析。Integrated Gradients 是一种基于梯度的归因技术, 能够衡量输入特征对模型预测结果的贡献。其数学定义如下: 给定模型输出函数 $F(\mathbf{x})$ 和基线输入 $\mathbf{x}'$ (一般选取一个无信息输入, 例如本研究中使用全零特征序列作为基线 $\mathbf{x}' = \mathbf{0}$), 对于输入 $\mathbf{x}$, 第 i 维特征的集成梯度 IG_i 定义为沿着从基线 $\mathbf{x}'$ 到 $\mathbf{x}$ 直线路径积分梯度:

$$IG_i(\mathbf{x}) = (x_i - x'_i) \times \int_0^1 \frac{\partial F(\mathbf{x}' + \alpha(\mathbf{x} - \mathbf{x}'))}{\partial x_i} d\alpha \quad (22)$$

上述积分通过数值近似得到, 每个特征最终得到一个归因分数 IG_i , 其符号和大小分别表示该特征对预测结果影响的方向和强度。IG 方法满足完备性等性质, 即所有特征的归因和等于模型在输入 x 与基线 x' 输出之差: $\sum_i IG_i \approx F(\mathbf{x}) - F(\mathbf{x}')$, 从而保证归因分数的可解释

总量。

在本研究中，将训练好的双通道 LSTM 模型视为函数 F ，输入包括最近 T 日的市场和情绪特征序列。取基线输入为全零序列（表示“无市场波动且无情绪影响”的中性状态），对验证集中选取的样本应用 IG 算法。具体实现上，利用开源工具“Captum”的 Integrated Gradients 模块逐特征计算梯度并积分。由于双通道 LSTM 含有两个通道，因此对市场特征和情绪特征分别计算归因结果 $\mathbf{IG}^{(m)}$ 和 $\mathbf{IG}^{(s)}$ ，最终得到每个指标在时间序列上的贡献度，后又引入多样本来实现全局的特征重要性排序。

3. 投资策略设计与回测

(1) 买入信号构建：在获得股价预测结果后，构建一个简单的阈值交易策略并对其进行历史回测，以评估预测模型的实际应用价值。策略规则如下：如果模型预测下一日股价将上涨超过预设阈值 θ （例如 1%），则在当前收盘时产生买入信号，否则不交易。

具体实现中，应计算模型预测的下一日涨幅 $\hat{r}_{t+1} = \frac{\hat{P}_{t+1} - P_t}{P_t}$ （其中 P_t 为当日实际收盘价， $\hat{P}_{t+1}$ 为模型预测价），当 $\hat{r}_{t+1} \geq \theta$ 时视为有效信号。

一旦产生买入信号，策略将在下一交易日开盘建仓，并持有固定 N 天（如 20 个交易日）后卖出平仓。在持仓期间内即使再次出现信号也不额外买入（模拟一次交易周期内不重复建仓），等待持仓周期结束再继续观察后续信号。若当日无买入信号，策略空仓观望。该机制通过滑动窗口顺序扫描每一天的数据实现，模拟投资者在每个交易日根据昨日预测做出交易决

定的过程。

(2) 回测收益计算：每次交易的收益率按照买入持有期间的价格变动计算，具体公式为：

$$R_{\text{trade}} = \frac{P_{\text{sell}} - P_{\text{buy}}}{P_{\text{buy}}} \quad (23)$$

其中， P_{buy} 为买入价（信号触发后的下一交易日开盘价）， P_{sell} 为卖出价（持有 N 日后的收盘价）。

将所有历史交易的收益串联，可计算策略的累计收益曲线 $C_{\text{strategy}}(t)$ ：初始资金记为 1，每次交易收益率依次复合，例如第 i 笔交易收益率为 r_i 则 $C_{\text{strategy}} = \prod_i (1 + r_i)$ 。本研究将策略曲线与基准策略做对比，基准采用简单的“买入并持有”策略（即在回测期初始买入并一直持有比亚迪股票）。基准的累计收益可用股票价格相对起始价的比率表示： $C_{\text{bench}}(t) = P_t / P_{t_0}$ 。通过绘制策略与基准的收益曲线，可以直观比较模型辅助下的交易策略是否优于直接持股。

(3) 评估指标：为了全面评估交易策略的表现，本文采用了多维度的量化指标体系，包括：最终累计收益（净/毛）、年化收益率（净/毛）、胜率、平均盈亏表现、交易成本影响、夏普比率以及 Calmar 比率。通过上述指标，从收益能力、交易质量与风险调整回报等多个角度，对策略进行了系统性评估，评估了融合模型预测的交易策略相对于简单持有的优势和不足，为模型在实际投资中的应用提供参考依据，同时也为单模态和多模态在金融决策上的表现提供了一个实际的参考依据。

四、实证结果与分析

为避免时间序列自相关导致的样本泄露问

题，实验均采用时间顺序切分，以交易日为基准，使用过去 30 个交易日的情感数据和市场数据，这些数据作为滞后项，构造输入序列预测下一交易日收盘价，切分过程中不进行随机打乱，以确保数据的时序性和因果关系的合理性。

（一）模型预测性能比较

分别在单模态和多模态得到的对齐数据集下，用相同的模型框架，即双通道 LSTM 进行训练，并在验证集上进行预测，最终如表 1 所示，在单模态下，模型的最佳验证集损失为 39.6875；在多模态下，模型的最佳验证集损失为 61.1063，通过对比二者的验证集损失可以看出，在单模态下，模型的预测精度显然比多模

态的高。

表 1 验证集损失对比表

最佳验证集损失	
单模态（仅新闻文本）	多模态（文本 + 语音 + 图像）
Val Loss = 39.6875	Val Loss = 61.1063

同时，还对验证集进行了“预测曲线 vs 真实曲线”的可视化图像的对比，如图 5 和图 6 所示，虽然二者都能一定程度地描绘出与实际曲线近似的预测曲线，但是通过图像的直观对比，可以明显地看出单模态的预测更为精准，尤其是在后端，多模态预测曲线呈明显的平滑趋势，与实际结果严重不符。



图 5 单模态预测图像



图 6 多模态预测图像

结论：不管是从验证集损失还是可视化图像的直观观察，单模态的预测表现都要明显强于多模态的表现，这说明模态的简单堆叠（即使是线性回归进行融合）并未带来模型性能的提升，反而可能由于模态间信息同质化、互补性不足，会给表现较好的特征增加额外的噪声，使模型泛化能力减弱。这一结论验证了“多模态并非越多越好，其效果取决于模态间的互补性”。

（二）特征贡献度分析

本研究主要聚焦在特征贡献性上来深层次

判断单模态和多模态在预测任务中的主要差异。本实验采用 Integrated Gradients 方法对最优的双通道 LSTM 模型进行了特征重要性分析，以得到不同特征在时间序列上的贡献以及引入情感项是否真的对模型预测有提高作用。首先对于单模态，如图 7 所示，可以看到，情感交互项的贡献度都要高于对应的市场特征本身的贡献程度，说明引入单模态的新闻文本情感得分确实增强了模型的预测能力，帮助模型捕捉到价格波动中的情绪驱动因素。

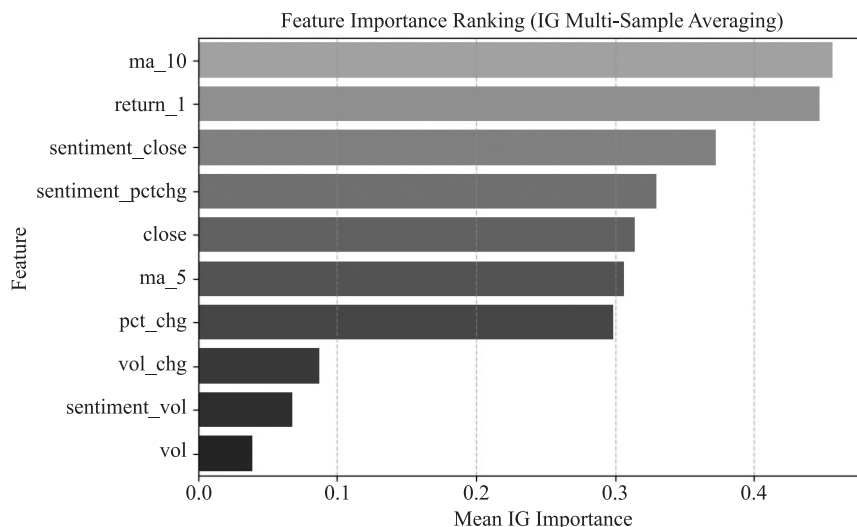


图7 单模态特征重要性排序

而对于多模态，从图 8 可以看出，融合了多模态情感得分的情感交互项的贡献程度远不及对应的市场特征本身，尤其是 *sentiment_pctchg* 即“情绪×涨跌幅”交互项对模型几乎没有贡献作用，增加的情感因素严重扰乱了特征本身的有效性。

通过上述单模态和多模态的情感对模型预测性能的影响可以看到，单模态文本情感对性能有提升，而多模态情感则对性能有抑制。表明多模态情绪信息未能形成有效互补（未满足

互补条件 C_2 ），削弱了关键特征的作用。

（三）网格搜索

由于主实验中的权重比例是利用人工打分的融合情绪得分进行线性回归得来，因此严重依赖人类的主观判断和评价标准，具有不确定性，为了减少这种现象对实验结论的影响，本节将回归确定的 1:1:3 作为初始权重，通过网格搜索进一步探究在不同权重比例下对模型性能的影响。采用步长为 0.2 进行网格搜索，通过图 9 可知，在多模态融合后，所有模型的

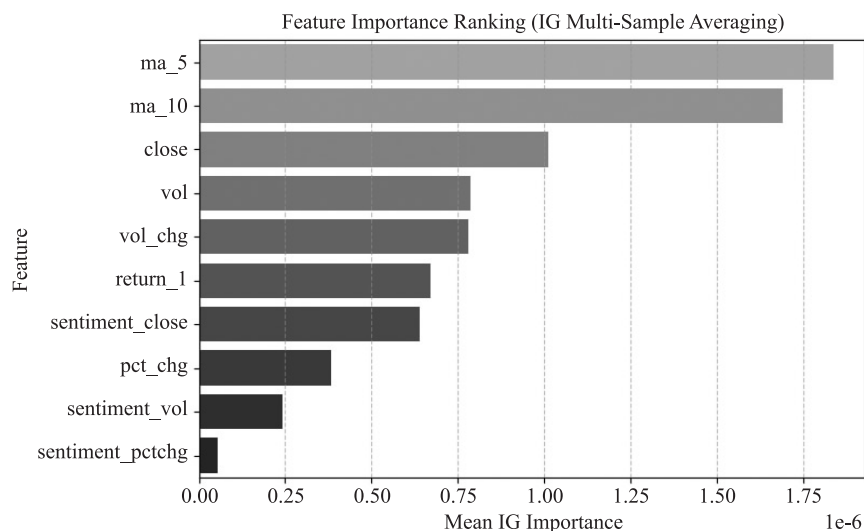


图8 多模态特征重要性排序

验证集损失均大于单模态的验证集损失 39.6875，同时可以观察到随着语音和图像权重逐渐减小、融合方式更加接近文本单模态时，模型的验证集损失越来越低并趋近于单模态损

失。该现象进一步说明多模态并不是越多越好，另外两个模态没有提供有效的增益（网络搜索的具体验证集损失值详见附录 C.2）。

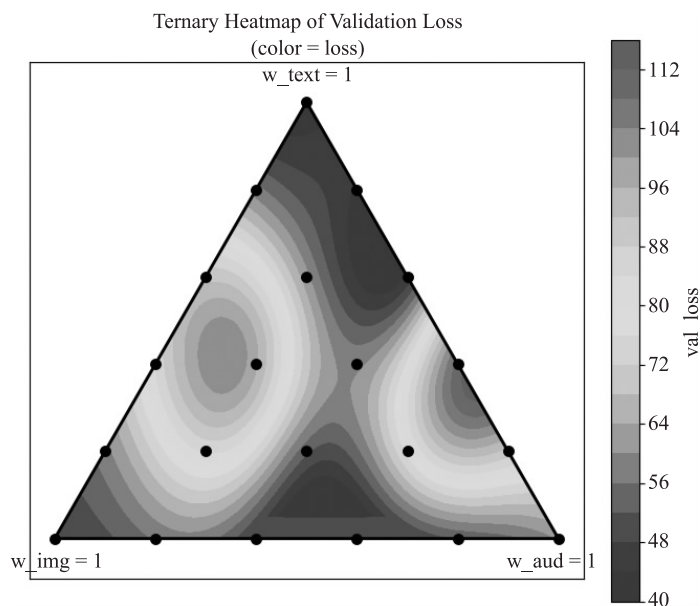


图9 验证集损失三元热力图

（四）可解释性分析

本节将针对上面两节的结果进行进一步的定量剖析，分析新增的两个模态究竟如何影响模型的性能，进一步揭示多模态融合未能产生

性能增益的原因。

从本部分（二）中的 IG 特征贡献度结果可知，对于多模态模型，将 IG 归因结果按特征所属进行分组后，可以观察到情绪交互项（*senti-*

ment_close、sentiment_pctchg、sentiment_vol) 的归因贡献度显著低于市场特征 (如 ma_5、ma_10、close 等), 并且通过数值可以观察到, 在引入交互项后也严重影响了其他指标的贡献程度 (例如, ma_10 从单模态贡献度的 0.46 左右骤降到了 $1.7e-6$), 说明引入多模态的情绪特征不仅不能带来信息增益、不能够被有效利用, 反而在梯度归因层面对市场特征形成明显的“信息压制”效应; 相比之下, 单模态模型的情感交互项的贡献度居于贡献度排行前列, 表明其在无干扰情形下能够被模型有效利用。

为了进一步验证 IG 归因结果与模型预测性能之间的结构性对应关系, 在本部分 (三) 中进行了网格搜索的权重调整, 得到不同模态权重下的验证集损失 (具体对照表见附录 C.2)。通过对比不同权重下的模型性能来看, 权重组合为 0:0:1 时, 验证集损失为 40.9, 在所有权重组合中表现最佳, 该权重比实际上就是单模态, 而随着语音模态和图像模态权重的增加, 验证损失大体上呈现显著恶化的情况 (如 0.6:0.2:0.2 对应 84.3; 0:0.6:0.4 对应 109.1), 这一现象与 IG 的特征贡献度结果高度一致, 进一步说明语音和图像两个模态不但未带来任何增益, 反而导致性能系统性下降, 说明其在表征空间中构成了典型的“干扰模态”。

综上所述, 多模态与市场模态的“交互项”表现出明显的负向协同结构, 当多模态情绪被引入后, 模型不能够有效利用多模态情绪并且对市场指标的梯度归因发生系统性塌陷, 而单模态却能够被有效利用, 说明三个模态的融合情绪会带来系统性的干扰, 不满足互补性条件

C_2 : 有效信息增益, 多模态间没有形成有效互补。

(五) 缺乏有效互补的信息整合在实践中的负面效应

本研究还利用单模态和多模态训练下的 LSTM 模型验证了简单的策略回测来验证实际价值, 最终结果也表明, 多模态的策略最终收益惨淡, 而单模态收益较好。

具体操作如下, 首先若预测下一交易日涨幅 $\hat{r}_{t+1} = \frac{\hat{y}_{t+1} - y_t}{y_t} \geq \theta$, 则认为存在上涨信号, 人为设定一个阈值 θ , 当信号超过阈值时则进行买入操作, 并持有特定的天数后卖出。本次模拟设定的阈值为 0.5%, 持有天数为 20 天。在附录 B.1 中展示了其中一组交易可视化图像。

从表 2 的回测结果来看, 单模态在收益水平、交易成功率以及风险调整收益等核心指标上均显著优于多模态策略。单模态能够更加稳定地识别股票趋势信号, 能够在维持较低风险下获得更高的收益表现。相反, 多模态的胜率显著偏低, 其表现甚至低于随机水平; 多模态在风险调整后的表现也相对有限, 其夏普比率为负, 显示在承担风险的情况下, 投资者却并未获得相应的补偿。

表 2 策略回测结果对比

策略回测结果		
绩效指标	单模态策略	多模态策略
交易频率相关		
总交易次数	12.75	8.80
盈利次数	7.15	4.19
胜率	56.10%	47.69%

续表

策略回测结果		
绩效指标	单模态策略	多模态策略
收益能力		
最终累计收益（净）	21.77%	5.63%
最终累计收益（毛）	28.87%	10.47%
平均盈利交易	8.71%	7.67%
平均亏损交易	-4.84%	-6.28%
年化收益率（净）	2.85%	0.44%
年化收益率（毛）	4.93%	1.39%
交易成本拖累	0.21%	0.09%
风险调整收益		
Calmar 比率（Net）	0.20	0.11
Calmar 比率（Gross）	0.32	0.14
夏普比率（Net）	0.12	-0.30
夏普比率（Gross）	0.26	-0.08

（六）消融实验

为了验证模态的具体干扰效果，本节通过消融实验来验证在文本信息的基础上仅加入图像信息或仅加入音频信息下模型的具体表现，具体的，我们通过模型在相同验证集上的损失值来量化模型的表现能力。通过图 10 可知（具体训练时间和验证集损失如表 3 所示），随着模态的增加，验证集损失不降反增，而在最终三种模态融合后，模型的验证集损失比单模态模型高了 53%，说明多模态的引入并未对预测任务形成有效增益，反而会由于引入了多余的干扰，使模型的预测能力减弱。

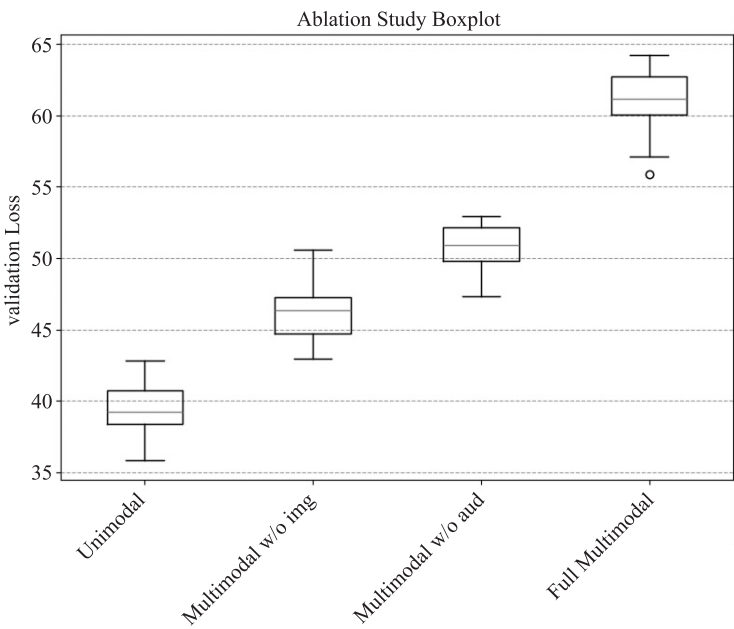


图 10 不同模态组合验证集损失

表 3 消融实验与融合策略比较

模型名称	训练时间	验证集损失	模型名称	训练时间	验证集损失
消融实验结果			融合策略比较		
Unimodal	51.1s	39.6875	LSTM Fusion	78.7s	46.8024
Multimodal w/o img	59.3s	46.8919	MLP Fusion	61.9s	47.9297
Multimodal w/o aud	60.8s	50.8469	Weight Fusion (Main)	58.4s	61.1063
FullMultimodal	58.4s	61.1063			

（七）融合策略比较

由于深度学习融合具有黑箱效应，因此在前面的主实验中采用的是回归方法确定权重，通过权重手动融合来实现更加透明的解释性分析。但是为了验证“多模态并非越多越好”这一现象来自模态本身，而非来自手动设定的权重所导致的偏差，我们引入了端到端的深度学习融合机制，完全避免了人为融合策略的影响，从而更准确地评估多模态信息在当前任务下的真实增益能力与鲁棒性。在本节的实验中，我

们新引入了两种融合方式，MLP 和 LSTM。

通过表 3 和图 11 可知，即使采用端到端的可学习融合结构，多模态的验证集损失仍然显著高于单模态，由此可以证明在本实验中，多模态融合效果不及单模态的现象并非偶然产生，也并非由手动权重融合造成的偏差所导致，而是多模态情绪特征在当前数据条件和特征构造方式下本身缺乏稳定增益，从而导致模型性能下降。

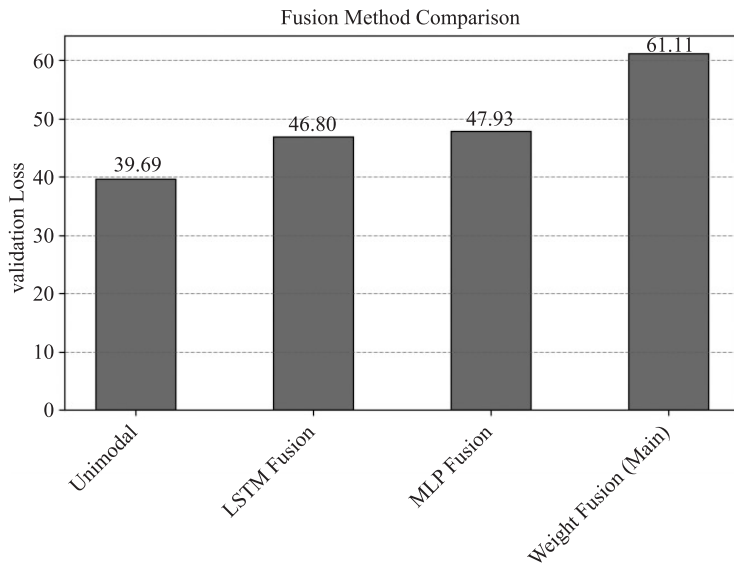


图 11 融合策略性能比较

（八）模态互补性分析

在第三部分（二）中，我们对模态的互补性进行了形式化的界定，若多模态之间具有互补性，需要具备三个必要条件：存在较少信息冗余、具备任务相关信息增益，以及可融合性。

首先，验证模态信息是否存在冗余，为了验证这一条件，我们对三种模态进行了相关性检验，相关性热力图如图 12 所示。从相关性热力图可以观察到三个模态之间的相关性较低，具有很少的信息冗余，满足 C_1 这一必要条件。

其次，验证模态是否有信息增益，为了量化信息增益，我们直接利用单模态模型以及多模态模型在验证集上的损失，从表 1 显示的结果上来看，单模态在三轮训练后的验证集损失为 39.6875，而多模态在训练三轮后的验证集损失达到了 61.1063。显然，在文本信息条件下引入语音和图像两个模态导致了模型的预测性能下降，并且从可解释性分析中也可以观察到多模态的引入并没有带来信息增益 [见第四部分（四）]，同时为了减少随机性对实验结果带来的

影响,通过消融实验更换模态组合、更换融合策略,以及网格搜索调整融合权重[见第四部分(三)(六)(七)中]都进一步验证了增加模态没有对模型产生增益效果,综上实验能够证明第二个条件 C_2 不成立。

最后,由于很难对融合效果直接进行量化比较,因此对于可融合性的验证,需要利用不同融合方法的泛化效果(即更换多种不同的模态融合方法,并观察最终的信息增益效果)、模型训练时间等进行间接验证。表3中的“融合策略比较”结果显示,虽然在训练时间上没有显著差异,但是即使切换为MLP、LSTM等融合方法,最终的多模态模型预测能力仍然不如单模态模型,说明条件 C_3 不成立,模态间不满足

可融合性。

综上所述,本文的三个模态仅满足条件 C_1 ,而不能够满足 C_2 和 C_3 ,也就是三个模态不具有互补性,因此可以得到结论:多模态数量不等于模型能力提升,模型能力的提升关键在于模态互补,而非简单的数量堆叠。

同时,结合第四部分(六)中的消融实验箱型图可以直观观察到边界条件的存在,在已知不满足互补性的条件下,在文本情绪模态的基础上逐渐增加另外两种模态,模型的验证集损失呈现逐渐增加的趋势,这进一步支持了边界效应的实证结论:模态间未能实现有效互补时,模态数量的增加反而可能导致性能退化。

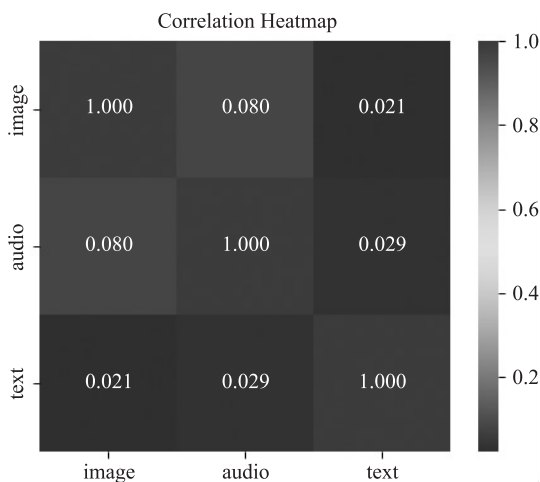


图12 三模态相关性矩阵

(九) 边界效应的经济和管理学意义

前文通过一系列对比实验对“多模态的应用整合可以改进信息表征效果”这一普遍隐含假设进行了实证检验与修正,并且基于本文提出的“互补性三条件”验证出了在本实验条件下模态之间不具备有效互补性,“互补性三条件”因而构成了边界效应的结构判定依据:当

模态之间不具备有效互补时,多模态扩展将难以形成任务相关的信息增量,反而可能削弱原有指标的代表效果。然而,这种差异的意义不仅体现在统计指标上,更体现在其对信息筛选、指标构造与研究设计所带来的方法论启示上。

1. 边界效应的经济后果:从统计差异到绩效代价

在第四部分（八）中，在“互补性三条件”框架下我们验证出了本实验条件下，模态间不具备有效互补。在此基础上，返回观察预测性能指标对比发现，多模态的损失为61.1063，而在只使用文本这一单一模态下，损失值为39.6875，验证集损失提高幅度超过50%。在策略回测方面，多模态的年化收益率（净）、最终累计收益等指标均低于单模态，同时多模态在夏普比率由单模态的正值转为负值，说明其在承担风险的情况下未能获得相应的经济补偿。

结果说明，多模态在不满足互补性的情况下，不仅不能带来边际收益的提升，反而会造成一定的经济代价：首先，预测精度的下降会带来对交易信号的误判；其次，模型风险调整能力的弱化会造成资本配置效率的下降；最后，新增模态所需要的信息成本并不能够转化为有效的绩效收益，形成资源投入的负回报。因此，本研究所揭示的边界效应并非技术层面的微弱差异，而是在互补性条件缺失下所呈现出的具有显著经济后果的结构性失效。

2. 边界效应的管理学意义：从信息扩展逻辑到结构诊断逻辑

从管理学视角看，本研究的核心贡献不仅在于修正了“多模态整合后的特征具有更有效的信息表征效果”的这一普遍假设，更在于为研究者界定了多源信息整合的结构边界，从而为管理学研究中的信息筛选、指标构造与研究设计提供了方法启示。

第一，在信息处理理论框架下，我们发现信息扩展并非单调提升绩效。在信息难以形成有效互补时，即使增加多维信息也不能够带来

明显的信息增益，甚至会导致信息的冗余，削弱原有指标的表征效果以及模型的泛化能力。

第二，在进行模态扩展时，信息的收集、数据的标注等会伴随着信息成本的显著提升。如果新增模态没有能够提供任务相关的信息增量，会形成“无效投资”，造成不必要的成本浪费。因此，在进行信息或模态扩展前，应先考察模态间的互补关系，避免基于“信息越多越好”的直觉逻辑进行资源投入。

第三，本文提出了“互补性三条件”作为边界效应的诊断工具，在正式引入新模态前，应该按以下顺序对将要新增的模态进行简要验证：（1）是否存在信息冗余；（2）是否提供可验证的任务相关增益；（3）融合机制是否具备可利用性。只有同时满足这三点，才可能会带来有效的信息增益。

总体而言，本研究将多模态问题从技术优化层面提升至信息整合边界的理论问题，强调多源信息整合应遵循“结构互补优先于数量扩展”的逻辑。这一结论不仅界定了多模态方法作为信息表征与指标构造策略的适用边界，也为管理学研究中如何审慎引入新增信息来源提供了理论启示。

（十）小结

本部分从预测性能比较（验证集损失，可视化图像对比）、可解释性分析、冗余信息在实践中的负面效应、消融实验、模态互补性分析等几个角度系统对比了单模态与多模态模型的实证结果。结果表明：

（1）多模态并未显著优于单模态，反而在模态简单堆叠下性能下降，说明模态数量的增加并不必然带来更优的信息表征效果；

(2) 文本情绪是本实验中最稳定且最具解释力的模态,通过“互补性三条件”分析出图像与语音模态未能与之形成有效互补,造成了特征干扰与学习负担增加;

(3) 策略回测结果进一步表明,这种性能差异已转化为现实决策层面的经济后果。

因此,多模态扩展的绩效效果具有明确的结构边界,其关键并不在于模态数量,而在于是否形成可利用的互补结构。

五、结论与启示

本研究采用单模态/多模态情感分析+双通道 LSTM 方法,结合单一股票的市场特征,对多模态信息整合作为一种信息表征策略的效果进行了系统分析。通过最终的预测结果、可解释性分析结果、策略回测结果等多个方面对单模态和多模态下模型的表现力进行综合性对比研究,更正了“模态的扩展与融合可以改进信息表征效果”这一普遍假设,并揭示了多模态信息整合的结构性边界条件,即“互补性三条件”:多模态并非必然带来更优的信息表征效果,其有效性取决于模态间互补结构,而非单纯的数量堆叠。

首先,对预训练模型 rbt3 进行了两次训练(基础训练+弱监督再训练),得到用于金融新闻情感分析的模型,然后利用它对收集到的“比亚迪”相关新闻文本数据进行情感打分,并聚合每天的情感得分;将视频数据进行抽帧加人脸识别,得到每个视频的人脸图像,输入到 ViT 模型中并聚合每个视频的人脸图像情感得分;利用 Whisper 对视频数据进行语音识别转

换,得到视频内容长文本,再通过 mT5 模型进行摘要获得便于情感分析的短文本,然后输入到训练后的 rbt3 模型中进行情感分析。图像和音频数据的情感得分需要通过艾宾豪斯记忆曲线进行插值,得到日度数据,然后用固定权重融合三种情感得分,四种情感得分与市场数据对齐,得到最终双通道 LSTM 的训练数据。

其次,构造双通道 LSTM 模型,并通过双通道分别输入市场特征数据、情感特征数据进行训练,最后再通过全连接层聚合结果,得到最终的预测模型。结合 IG 特征贡献度以及网格搜索调整权重对该实验进行了详细的可解释性分析。同时,通过策略回测的方法模拟了股票的买入、持有、卖出过程,验证了模型在实际运用中的价值。为了增强结论的严谨性以及互补性的进一步探索,通过消融实验以及更换不同融合策略来反复验证,结果均保持一致:多模态模型无法超越最佳单模态模型。综合所有对比实验,发现该现象并非由特定参数设定或样本随机性导致,而是源自数据结构与任务属性本身。

最后,在互补性分析中,进一步将实证结果置于“多模态互补性三条件框架”(非冗余性、相关信息增益、可融合性)之中进行系统解释,尽管相关性低常常被认为存在互补性,但实际上相关性低仅能说明信息非冗余,是互补性的必要条件而非充分条件。在本实验中虽然相关性较低,满足互补性条件一,但是在引入另外两个模态后未能产生信息增益同时无论采用哪种融合方式都不能实现有效融合,因此不满足互补性条件二和条件三,也就不具备互

补性。同时结合消融实验的箱型图可知，在不满足有效互补性的条件下，随着模态的增加，损失逐渐增大，这一趋势是边界效应存在的直观体现，即在模态间不具备有效互补性的情况下，简单堆叠模态不仅无法提升模型性能，反而会导致性能退化。

本实验最终成功得出结论：多模态并非必然带来更优的信息表征效果，“模态互补性”而非“模态数量”才是影响多模态整合能否形成有效增量的关键机制。

这一结论也界定了多模态方法在金融与管理研究中的适用边界：当多模态信号之间存在冗余信息或即使非冗余但任务相关性弱、质量差异较大或难以实现有效融合时，采用单一高质量模态往往比强行叠加多模态更具效率。反之，只有在多模态信号同时具备差异性、增量性与可整合性时，多模态策略才可能带来信息增益。

从管理实践来看，这一发现对研究者在高信息复杂度场景中如何使用多模态方法具有直接启示。研究者在引入多模态数据时，应更加关注模态间的互补性而非数量本身，不应简单以“信息越多越好”为前提，应当充分了解各个模态之间的信息关联性，尽量减少冗余信息的加入，确保模态互补以达到提高模型性能与解释力的目的。具体来说，多模态只有在满足一定的互补性条件下才能产生实际效益，首先要利用相关性或相似性指标判断模态之间是否存在过度重叠的冗余信息，在满足没有过多冗余信息这一条件后，再从引入的多模态中挑选出较为重要的单一模态，分别做一次单模态引入和多模态引入实验，通过对比实验验证多模

态是否能够带来可观的解释力增量或风险识别能力的提升。只有当新增模态确实能在解释力或风险识别层面提供关键增益时，并能够在技术上实现可整合性，多模态整合才是值得投入的方向。换言之，多模态方法的价值不取决于信息源多少，而取决于这些信息源能否共同构成有效的互补结构。

六、实验不足及展望

本研究仅研究了“比亚迪”这一单一股票，这种研究设计虽然有助于在相对稳定的情境下观察不同信息模态之间的结构关系，从而更清晰地识别多模态信息整合的机制，但缺乏对其他领域股票的横向比较，其通用性需要进一步加强；本实验的弱监督再训练用的是第一次训练后的模型，即基础模型打出的伪标签数据，缺乏人工标注的更为优质的数据，因此最终的情感分析模型还有待提高；由于图片和音频的训练成本和资源来源有限，没有训练更为适配本实验的模型，而是采用来自 huggingface 等的开源模型，且音频模型没有使用专门用于音频情感分析的模型，未来还需进一步探索；本实验采用的仅为双通道 LSTM，并未加入注意力机制等，在面对更复杂数据和深层关系时可能表现欠佳。

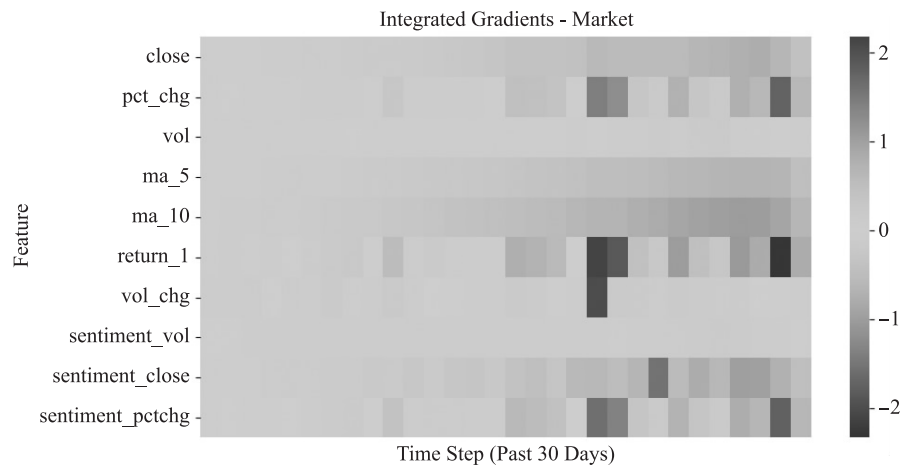
七、附录

（一）A.1 单模态两个通道特征在时间序列上的贡献作用

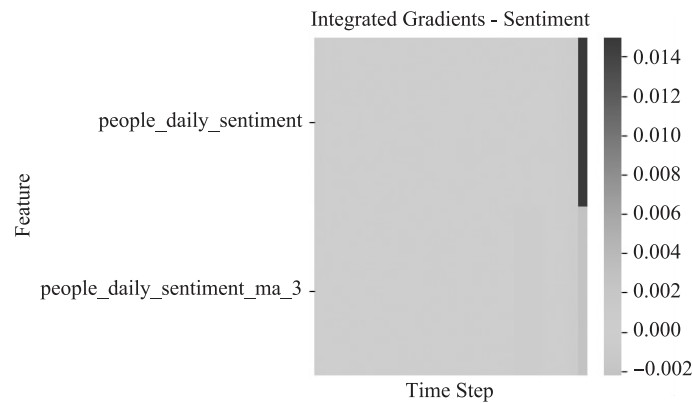
附图1和附图2表示的是单模态两个通道

在时间序列上的贡献作用，首先可以看到在附图 1 市场通道中，收盘价 *close* 的作用从时间步开始一直到最后都有着明显的贡献作用，符合基本逻辑；而未加入情感交互项的成交量 *vol* 显然没有加入情感交互项的 *sentiment_vol* 在时间步

的贡献持久，说明加入情感因素确实能够提高对模型的贡献度。其次，在附图 2 情感通道中，可以看到单一的情感因素只在时间步的最后起到一定的作用，符合在看到新闻时的反应逻辑。



附图 1 单模态市场通道时间序列贡献图

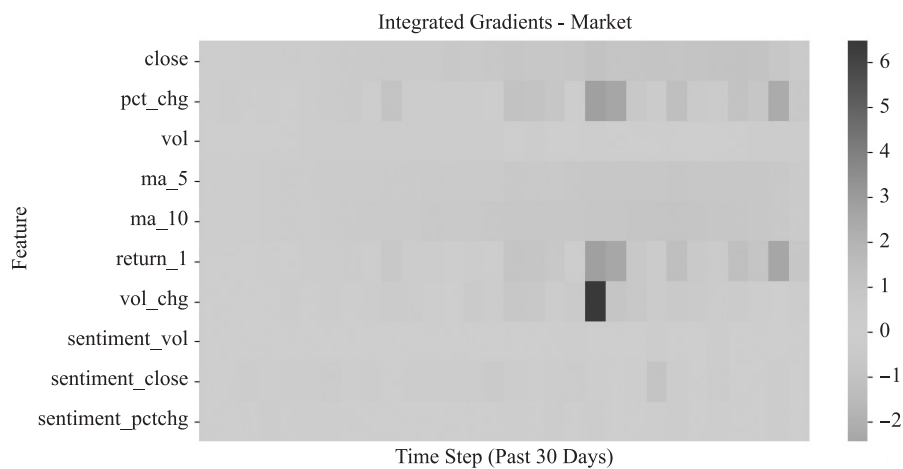


附图 2 单模态情绪通道时间序列贡献图

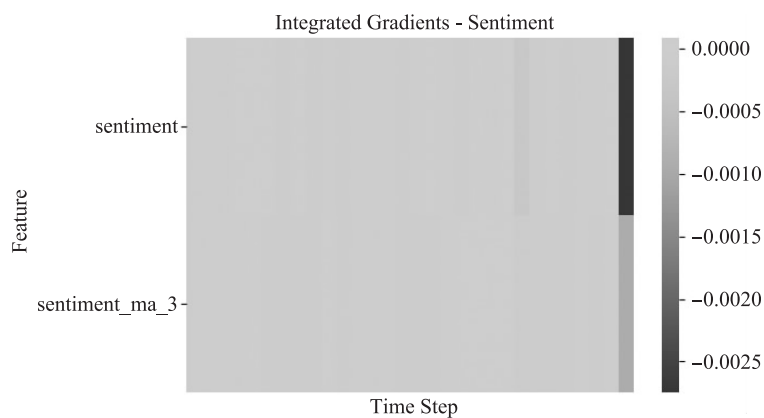
（二）A.2 多模态两个通道特征在时间序列上的贡献作用

附图 3、附图 4 表示的是多模态两个通道在时间序列上的贡献作用，通过两个图可以直观

地看到增加模态数量后融合项和单独情绪项的贡献度很低，说明多模态在整个时间序列上都没有正向影响。

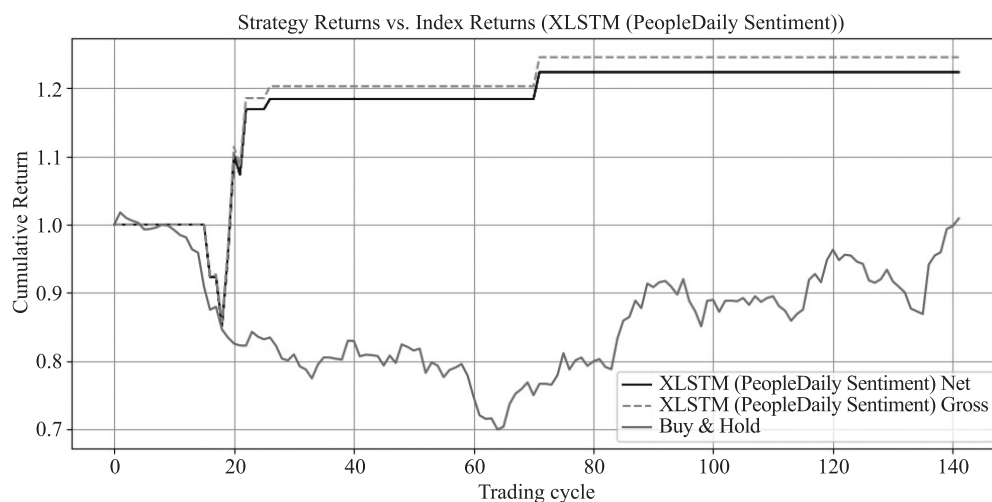


附图3 多模态市场通道时间序列贡献图

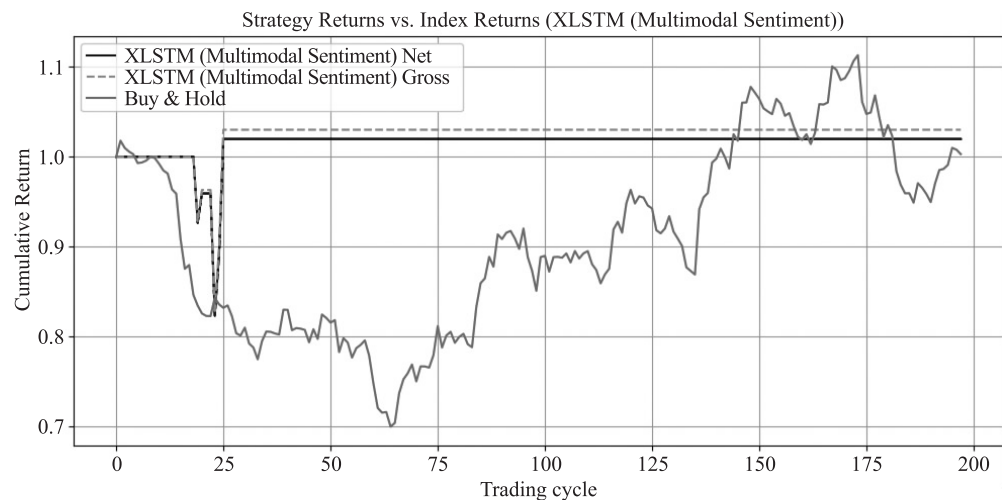


附图4 多模态情绪通道时间序列贡献图

(三) B.1 单模态和多模态策略回测“策略回报率与基准回报率”图像 (附图5 和附图6)



附图5 单模态“策略回报率与基准回报率”



附图 6 多模态“策略回报率与基准回报率”

（四）C.1 初始权重确定过程

由于数据量的限制，人工标注大量数据在短时间内并不现实，因此我们从中随机选取部分对齐数据，对这部分数据进行人工打分得到融合情绪值（允许人工的打分范围为 0~1，正向的趋近于 1，负面的趋近于 0，允许的最小步长为 0.1）。将人工打分得到的融合情绪值作为被解释变量，模型对三个模态的打分作为解释变量进行线性回归拟合，得到最终的手动权重约比为 0.221:0.198:0.582，约为 1:1:3。

（五）C.2 网格搜索不同权重下的性能

附表 1 显示了在网格搜索的权重组合上进行情绪融合，然后对模型进行训练并得到的在验证集上的损失。

附表 1 网格搜索不同权重比较

图像权重	音频权重	文本权重	验证集损失
0	0	1	40.9352
0	0.2	0.8	42.9598
0	0.4	0.6	44.5939
1	0	0	49.2675
0.2	0.8	0	52.0538
0.4	0.4	0.2	52.9118

续表

图像权重	音频权重	文本权重	验证集损失
0.8	0	0.2	58.5047
0.2	0.4	0.4	58.8479
0	1	0	60.1925
0.2	0.2	0.6	61.1063
0.6	0.2	0.2	84.3478
0	0.8	0.2	90.2988
0.4	0.2	0.4	96.3307
0.2	0.6	0.2	77.1845
0	0.6	0.4	109.121

接受编辑：程聪

收稿时间：2025 年 10 月 9 日

接收时间：2026 年 5 月 12 日

作者简介

刘壮，博士，现任职于东北财经大学金融科技学院（School of FinTech, Dongbei University of Finance and Economics），职称为副教授。于大连理工大学获得工学博士学位。主要研究方向包括金融欺诈、金融风险管理、投资策略与市场分析等。相关成果发表于 ACL、IJCAI、EMNLP、KDD、CIKM、ECAI 等国际会议，以

及 *ACM Transactions on Intelligent Systems and Technology* 等期刊。

畅一鸣 (通讯作者, E-mail: 548599911@qq.com), 现为东北财经大学金融科技学院 (School of FinTech, Dongbei University of Finance and Economics) 硕士研究生。主要研究方向包括人工智能、金融文本分析与管理决策分析等。

参考文献

- [1] 部慧、解峥、李佳鸿、吴俊杰:《基于股评的投资者情绪对股票市场的影响》,《管理科学学报》,2018年第21卷第4期。
- [2] 何勇、焦丽、杨艺、祝怡菲:《AI大模型赋能金融市场量化投资?基于另类数据与传统金融数据的研究》,《计量经济学报》,2024年第4卷第3期。
- [3] 林建浩、张一帆、陈良源等:《基于新闻情绪的机器学习交易策略》,《计量经济学报》,2022年第2卷第4期。
- [4] 孟勇、任梦、赵心:《行业资产的 Black-Litterman 模型配置研究》,《数量经济技术经济研究》,2022年第1期。
- [5] 石善冲、朱颖楠、赵志刚、康凯立、熊熊:《基于微信文本挖掘的投资者情绪与股票市场表现》,《系统工程理论与实践》,2018年第38卷第6期。
- [6] 徐鹏、尚维:《基于趋势情感映射的物价舆情词典及舆情指数构建研究》,《系统工程理论与实践》,2022年第42卷第12期。
- [7] 许雪晨、田侃:《一种基于金融文本情感分析的股票指数预测新方法》,《数量经济技术经济研究》,2021年第12期。
- [8] 姚加权、冯绪、王赞钧、纪荣嵘、张维:《语调、情绪及市场影响:基于金融情绪词典》,《管理科学学报》,2021年第24卷第5期。
- [9] 尹海员、吴兴颖:《投资者日度情绪、订单流不平衡与股票流动性》,《中国管理科学》,2023年第31卷第5期。
- [10] 于孝建、刘国鹏、刘建林等:《基于LSTM网络和文本情感分析的股票指数预测》,《中国管理科学》,2024年第32卷第8期。
- [11] 张宸瑞:《基于深度学习文本情绪挖掘股市相关性研究》,中国科学院科技论文预发布平台,DOI: 10.12074/202205.00035V2。
- [12] 张飞鹏、徐一雄、陈曦、周勇:《基于新闻文本情绪的区间值股票回报预测研究》,《计量经济学报》,2024年第4卷第1期。
- [13] 张宗新、吴钊颖:《媒体情绪传染与分析师乐观偏差——基于机器学习文本分析方法的经验证据》,《管理世界》,2021年第37卷第1期。
- [14] 周颖刚、唐诚蔚、林哲晖:《市场情绪、投资因子与资产定价——基于新闻与社交媒体的对比分析》,《计量经济学报》,2024年第4卷第3期。
- [15] 朱莉、刘向丽、杨晓光:《投资者情绪影响股指期货现货市场的价格动态关系吗?》,《中国管理科学》,2022年第30卷第4期。
- [16] Baltrušaitis, T., Ahuja, C., & Morency, L. P. 2019. Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41 (2): 423–443.
- [17] Das, R., & Singh, T. D. 2023. Multimodal sentiment analysis: a survey of methods, trends, and challenges. *ACM Computing Surveys*, 55 (13s): 270: 1–270: 38.
- [18] de la Rica Escudero, A., Garrido-Merchán, E. C., & Coronado-Vaca, M. 2024. Explainable post-hoc portfolio management financial policy of a deep reinforcement learning agent. *arXiv preprint, arXiv: 2407.14486*.
- [19] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. *arXiv pre-*

print, arXiv: 1810.04805.

[20] Du, K. K. W., Xing, F. Z., Mao, R., & Cambria, E. 2024. Financial sentiment analysis: techniques and applications. *ACM Computing Surveys*, 56 (9): 1–42.

[21] Eppler, M. J., & Mengis, J. 2004. The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines. *MIS Quarterly*, 28 (2): 325–344.

[22] Fazlija, B., & Harder, P. 2022. Using financial news sentiment for stock price direction prediction. *Mathematics*, 10 (13): 2156.

[23] Gite, S., Khatavkar, H., Kotecha, K., Srivastava, S., Maheshwari, P., & Pandey, N. 2021. Explainable stock prices prediction from financial news articles using sentiment analysis. *PeerJ Computer Science*, 7: e340.

[24] Han, W., Chen, H., Gelbukh, A., Zadeh, A., Morency, L. – P., & Poria, S. 2021. Bi – bimodal modality fusion for correlation – controlled multimodal sentiment analysis. arXiv preprint, arXiv: 2107.13669.

[25] Jiang, H., Miao, Z., Liu, Q., et al. 2020. Financial news annotation by weakly – supervised hierarchical multi – label learning. *Proceedings of FinNLP Workshop @ IJCAI 2020*.

[26] Kiela, D., Bhooshan, S., Firooz, H., & Tettgine, D. 2019. Supervised multimodal bitransformers for classifying images and text. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 1457–1466.

[27] Lai, S., Hu, X., Xu, H., Ren, Z., & Liu, Z. 2023. Multimodal sentiment analysis: a survey. *Displays*, 80: 102563.

[28] Li, T., Liu, Z., Shen, Y., Wang, X., Chen, H., & Huang, S. 2023. MASTER: market – guided stock transformer for stock price forecasting. arXiv preprint, arXiv: 2312.15235.

[29] Liagkouras, K., & Metaxiotis, K. 2025. A hybrid long short – term memory with a sentiment analysis system for stock market forecasting. *Electronics*, 14 (14): 2753.

[30] Liang, P. P., Zadeh, A., Morency, L. P., & Salakhutdinov, R. 2018. Multimodal language analysis with recurrent multistage fusion. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*; 150–161.

[31] Moradi – Kamali, H., Rajabi – Ghazlou, M. – H., Ghazavi, M., Soltani, A., Sattarzadeh, A., & Entezari – Maleki, R. 2025. Revisiting financial sentiment analysis: a language model approach. arXiv preprint, arXiv: 2502.14897.

[32] Muhammad, D., Ahmed, I., Naveed, K., & Bendechache, M. 2024. An explainable deep learning approach for stock market trend prediction. *Heliyon*, 10 (21): e040095.

[33] Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., & Ng, A. Y. 2011. Multimodal deep learning. In *Proceedings of the 28th International Conference on Machine Learning*; 689–696.

[34] Ouf, S., El Hawary, M., Aboutabl, A., & Adel, S. 2024. A deep learning – based LSTM for stock price prediction using Twitter sentiment analysis. *International Journal of Advanced Computer Science and Applications*, 15 (12): 207–216.

[35] Sang, S., & Li, L. 2024. A novel variant of LSTM stock prediction method incorporating attention mechanism. *Mathematics*, 12 (7): 945.

[36] Simon, H. A. 1971. Designing organizations for an information – rich world. In M. Greenberger (Ed.), *Computers, Communications, and the Public Interest*; 37–72. Baltimore: Johns Hopkins Press.

[37] Todd, A., Bowden, J., Su, Y., & Cummins,

M. 2024a. A multimodal sentiment classifier for financial decision making. *SSRN Preprint*, No. 5038989.

[38] Todd, A., Bowden, J., & Moshfeghi, Y. 2024b. Text – based sentiment analysis in finance: synthesising the existing literature and exploring future directions. *Intelligent Systems in Accounting, Finance and Management*, 31 (1): e1549.

[39] Usmani, S., & Shamsi, J. A. 2021. News sensitive stock market prediction: literature review and suggestions. *PeerJ Computer Science*, 7: e490.

[40] Vallarino, D. 2025. An AI – enhanced forecasting framework: integrating LSTM and transformer – based sentiment for stock price prediction. *Journal of Economic Analysis*, 4 (3): 1 – 15.

[41] Wang, H., Xie, Z., Chiu, D. K. W., & Ho, K. K. W. 2025. Multimodal market information fusion for stock price trend prediction in the pharmaceutical sector. *Applied Intelligence*, 55: 77.

[42] Wang, S. 2023. A stock price prediction method based on BiLSTM and improved transformer. *IEEE Access*,

11: 104211 – 104223.

[43] Yang, J., Li, P., et al. 2025. Multi – sensor temporal fusion transformer for stock performance prediction: an adaptive Sharpe ratio approach. *Sensors*, 25 (3): 976.

[44] Yeo, W. J., van der Heever, W., Mao, R., Cambria, E., Satapathy, R., & Mengaldo, G. 2023. A comprehensive review on financial explainable AI. arXiv preprint, arXiv: 2309.11960.

[45] Zadeh, A., Liang, P. P., Poria, S., Vij, P., & Morency, L. P. 2018. Multimodal language analysis in the wild: CMU – MOSEI dataset and interpretable dynamic fusion graph. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*: 2236 – 2246.

[46] Zhang, P., Harris, R. D. F., & Zheng, J. 2024. GNN – based social media sentiment analysis for stock market forecasting and trading. *SSRN Preprint*.

[47] Zhang, Q. 2023. Portfolio construction with news sentiment using large language model. Master's thesis, University of Technology Sydney.