

基于多层级知识增强大模型的金融文本情感分析方法

万永智^{1,2}, 胡慧君^{1,2}, 张耀峰^{3,4†}

1. 武汉科技大学 计算机科学与技术学院, 湖北 武汉 430065;
2. 智能信息处理与实时工业系统湖北省重点实验室, 湖北 武汉 430065;
3. 湖北经济学院 数字金融创新湖北省重点实验室, 湖北 武汉 430205;
4. 湖北经济学院 新财经交叉学科研究院(数字经济学院), 湖北 武汉 430205

收稿日期: 2025-03-19 † 通信联系人 E-mail: yfzhang@hbue.edu.cn

基金项目: 国家社会科学基金重点项目(23ATJ005); 数字金融创新湖北省重点实验室开放基金(DFIK2024Y06)

第一作者: 万永智, 男, 硕士生, 现从事知识图谱方面的研究。E-mail: 1819298649@qq.com

摘要: 在情感分析领域, 基于大语言模型(LLM)的情感判断方法已得到广泛应用。然而, 金融文本通常包含复杂的情感特征, 且不同情感类别之间边界模糊, 导致 LLM 在需要准确区分情感类别时面临重大挑战, 限制了其在金融领域情感分析任务中的应用效果。因此, 通过构建金融情感知识图谱并结合少样本提示对 LLM 进行多层级知识增强, 解决情感边界模糊的问题。该图谱通过 LLM 进行半自动构建, 随后对其中的金融词汇进行情感标注。用提示模板将金融情感知识图谱中的情感词汇、情感强度、情感极性等多维度信息显性表示, 为 LLM 提供语义层次的支持, 再结合少样本提示以增强 LLM 在金融情感文本中的分析能力。在 Qwen2.5-7B、ChatGLM3-6B 和 ChatGLM4-9B 等 LLM 上对本文方法进行了验证。结果显示, 使用多层级知识增强的方法能够显著提升 LLM 的情感分析准确性, 相较于未使用该方法的 LLM, F1 平均提升 3.67 个百分点, 验证了所提方法的有效性, 为金融文本情感分析提供了新的解决思路与重要参考。

关键词: 金融文本情感分析; 知识增强; 大模型; 知识图谱; 少样本提示

中图分类号: TP391

文献标志码: A

文章编号: 1671-8836(2026)04-0531-12

Financial Text Sentiment Analysis Method Based on Multi-Level Knowledge-Enhanced Large Language Models

WAN Yongzhi^{1,2}, HU Huijun^{1,2}, ZHANG Yaofeng^{3,4†}

1. School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, Hubei, China;
2. Hubei Province Key Laboratory of Intelligent Information Processing and Real-Time Industrial System, Wuhan 430065, Hubei, China;
3. Hubei Key Laboratory of Digital Finance Innovation, Hubei University of Economics, Wuhan 430205, Hubei, China;
4. Interdisciplinary Research Institute in New Finance and Economics (School of Digital Economics), Hubei University of Economics, Wuhan 430205, Hubei, China

Abstract: In the field of sentiment analysis, large language models (LLMs) have been widely adopted for sentiment judgment. However, financial texts inherently involve complex sentiment features with blurred boundaries between distinct sentiment categories, presenting LLMs with substantial challenges in accurate classification. To address this issue, we propose a method that integrates a financial sentiment knowledge graph with few-shot prompting for multi-level knowledge enhancement. The graph is semi-automatically constructed using LLMs, and financial terms within it are annotated with sentiment labels. This information is explicitly represented through prompt templates, providing semantic-level support for the LLM. Experiments are conducted on Qwen2.5-7B, ChatGLM3-

引用格式: 万永智, 胡慧君, 张耀峰. 基于多层级知识增强大模型的金融文本情感分析方法[J]. 武汉大学学报(理学版), 2026, 72(4): 531-542. DOI: 10.14188/j.1671-8836.2025.0064.

WAN Yongzhi, HU Huijun, ZHANG Yaofeng. Financial Text Sentiment Analysis Method Based on Multi-Level Knowledge-Enhanced Large Language Models [J]. J. Wuhan Univ. (Nat. Sci. Ed.), 2026, 72(4): 531-542. DOI: 10.14188/j.1671-8836.2025.0064(Ch).

6B, and ChatGLM4-9B, demonstrating that our approach significantly improves sentiment analysis accuracy, with an average $F1$ increase of 3.67 percentage points over LLM that don't employ without this method. The results show that this method provides new solutions and important references for financial text sentiment analysis.

Key words: financial text sentiment analysis; knowledge enhancement; LLM(large language model); knowledge graph; few-shot prompting

0 引言

在金融领域,情感分析(Sentiment Analysis)已成为理解市场情绪、预测金融风险和指导投资决策的重要工具。早期研究主要依赖于词典方法,Loughran 和 McDonald^[1]开发的金融情感词典为情感词的分类和计量奠定了基础。随着深度学习的发展,各种基于深度学习的预训练语言模型(如 BERT^[2]、RoBERTa^[3]等)在情感分析任务中展现了更优的性能,这些模型利用上下文信息捕捉能力,显著提高了文本分析的准确性。Fatouros 等^[4]进一步探讨了大语言模型(Large Language Model, LLM)在金融情感分析中的潜力,结果表明 LLM 在该领域具有显著优势。

尽管如此,LLM 在金融情感分析任务中仍然面临情感边界模糊的问题,具体表现为文本情感倾向不明确、语境依赖性强、中立情感模糊及语言表达复杂等现象,而金融文本中大量专业术语与多元情感元素的存在,进一步加剧了 LLM 对文本整体情感倾向的判断难度,这类特殊语境使得情感边界的模糊性更易被放大,使 LLM 在缺少领域情感知识的背景下难以精准捕捉情感特征,大大降低 LLM 在该任务上的准确性和实用性。例如金融文本:“央行降息提振了市场信心,但滞胀风险引发担忧。企业盈利改善支撑股市,而去全球化趋势可能威胁长期增长。”这段文本很好地体现了金融文本中情感边界模糊的特点。积极与消极情感的混合、专业术语的多重含义以及情感强度的不对等性,使得 LLM 难以准确判断整体情感倾向。金融市场的情感往往是复杂且多变的,导致这种模糊性在金融文本中尤为常见。

为应对上述问题,本文提出融合知识图谱^[5](Knowledge Graph, KG)与少样本提示^[6](Few-shot Prompting)的多层级知识增强大模型的金融文本情感分析方法(Multi level Knowledge enhanced Financial text Sentiment analysis method, MKeFS)。参考 Pan 等^[7]的框架中提及三种 KG 与 LLM 结合的方式,本文采用了在 LLM 的输入阶段使用提示模板^[8]进行多层级知识增强。此外,考虑到金融领域情感知识

图谱的研究相对稀缺,现有的通用知识图谱主要聚焦于实体关系建模,而缺乏对情感信息的关注。为此,本文还构建了一个金融情感知识图谱,为金融文本情感分析提供关键知识支撑。该图谱融入了金融领域特有的情感知识,结合少样本提示增强模型的泛化能力,逐步引导 LLM 正确分类相似的情感特征。MKeFS 相较于现有方法,不仅弥补了单一知识增强方法的局限性,还开创性地实现了从不同层级对 LLM 进行的多方位知识增强。

本文主要贡献如下:

1) 构建金融情感知识图谱,填补了现有金融知识图谱在情感信息建模方面的空白,并结合 LLM 解决了知识图谱更新困难的问题。

2) 提出 MKeFS 方法,通过融合知识图谱与少样本提示的多层级知识增强方法,解决金融文本中的情感边界模糊问题。

1 相关工作

1.1 金融文本情感分析

金融文本情感分析是自然语言处理(Natural Language Processing, NLP)领域的重要研究方向,其目标是通过对金融领域的文本数据进行情感倾向性分析,以辅助投资决策和市场预测。Todd 等^[9]对将情感分析技术应用于金融领域的现有文献进行了评估,并对未来研究方向进行了探讨,指出随着深度学习技术的发展,Transformer 架构和多模态分析(Multi-modal Analysis)将成为研究热点。近年来,LLM 在金融文本情感分析中的应用逐渐增多,取得了显著效果。例如 Zhang 等^[10]提出的 Instruct-FinGPT,通过创建一系列人工编写的指令,将金融情感分析数据集转换为指令格式的数据集,将通用 LLM 适配至金融情感分析任务,增强了 LLM 在金融情感分析中的表现。Qian 等^[11]提出的 Fino1 模型,使用推理增强方法在多个金融任务中实现了性能提升,超越了同等参数规模的其他模型。

尽管这些研究在模型性能优化上成果斐然,但多聚焦于算法改进与任务适配,却忽视了金融文本情感分析的深层挑战,缺乏对情感层面的系统性建

模分析。尤其在LLM应用场景下,金融文本所包含的大量专业术语、隐含语义和复杂语境,使得LLM难以精确捕捉情感特征,导致不同情感类别的边界模糊问题愈发突出。MKeFS方法主要是使用知识图谱补充多维度情感知识帮助LLM捕捉情感特征,减少情感特征模糊的问题,同时使用少样本提示增强LLM的泛化能力,整体上提高金融文本情感分析的效率以及准确性。

1.2 知识增强

知识增强^[12]是通过引入外部知识提升LLM对文本的理解能力,弥补LLM在特定领域的知识空缺,从而达到提升LLM性能的效果。知识增强的核心思想是利用结构化或非结构化的知识资源,弥补LLM在领域知识和上下文理解上的不足。外部知识有多种类型,本文使用知识图谱以及少样本提示进行知识增强。

知识图谱增强是通过将特定领域的实体、关系及其语义信息融入LLM,帮助LLM更好地捕捉文本中的隐含知识。例如GPT-RE^[13]方法,将知识图谱嵌入预训练模型中,结合知识图谱在结构化知识表达上的天然优势,有效发挥了LLM在文本生成和理解方面的优势,在关系抽取、检索、问答等多个任务上取得了显著成果。Sun等^[14]提出了KGT(Knowledge Graph Tuning)方法,将知识图谱中的实体和关系融入LLM中,利用知识图谱作为一个外部知识库来为LLM提供补充信息。然而,基于知识图谱的知识增强方法存在局限性:其依赖高质量的知识图谱构建,对图谱的覆盖度和实时性要求较高,可能导致知识更新滞后或特定领域适应性不足。

少样本提示增强是通过设计针对性的提示模板,在少量标注数据的情况下引导LLM生成更准确的预测。这种方法特别适用于金融领域标注数据稀缺的场景,能够显著降低对大规模标注数据的依赖。但其缺点在于提示模板的设计需要较强的领域专业知识,且对LLM的泛化能力要求较高,可能在不同任务或场景下表现不稳定。Yoran等^[15]提出了多思维链推理方法,将多个推理链组合成一个多链上下文,并将其与原问题一起输入元推理模块,通过少样本提示来构建提示模板。Wu等^[16]提出一个根据输入的文本,从知识图谱中检索出与问题相关的子图,随即将其重写为对问答有用的提示模板。少样本提示通过提供有限的标注数据示例,引导LLM快速适应下游任务,在提示模板的构造与优化中具有重要作用。该技术可以在缺乏标注

数据的条件下提升LLM在低资源任务中的表现能力。Sahoo等^[17]通过少样本提示有效增强LLM在问答、文本分类等任务上的泛化能力。

综上所述,知识图谱能够提供结构化的领域知识,但无法帮助LLM快速构建分类框架,并且其效果十分受限于知识图谱的质量。少样本提示可以有效引导LLM执行任务,增强LLM的适应性和泛化能力,但在缺乏足够知识支撑的情况下可能导致LLM在特定任务的表现不够稳定。在当前研究中,两者尚未形成紧密结合,难以充分利用各自优势。尤其在金融情感分析这样的专业领域,在面对情感边界模糊、知识覆盖度不足等问题时,使用单一方法进行增强更会放大自身的缺点。为此,本文探索了融合知识图谱与少样本提示增强LLM的策略。不同于现有研究使用单一方法对LLM进行增强,本文方法克服了其局限性,实现了知识图谱与少样本提示之间的相互增强与短板互补。二者结合可以利用知识图谱的结构化情感知识为LLM提供知识支撑,同时使用少样本提示为其提供多样化样例,更好地覆盖金融领域内可能出现的各种情感状态和表达方式。这样的紧密结合不仅弥补了金融领域知识覆盖度上的不足,同时也解决了少样本提示可能面临的领域专业知识要求高以及稳定性差的问题,从而构建了一个更加全面、精确且适应性强的金融情感分析框架。这种多层级的知识增强策略使得LLM能够在复杂多变的金融市场中更准确地捕捉和理解细微的情感变化。

2 金融情感知识图谱的构建

构建知识图谱的方法通常可以分为自顶向下和自底向上两种思路。自顶向下的构建方法主要是基于已有的理论框架或本体结构,从宏观层面逐步细化知识图谱的层次和关系。自底向上的构建方法是从实际数据出发,通过数据挖掘技术逐步发现知识之间关系,将具体的数据逐步汇总成结构化的图谱。

本文采用自顶向下和自底向上相结合的方法,图谱的构建架构如图1所示。整个构建流程以半自动化循环迭代的方式进行,主要任务包括本体构建、金融词典构建以及使用LLM进行词典扩充。

2.1 模式层

本文考虑了复用现有本体的可能性。大连理工大学信息检索实验室在经过全体教研室成员的努力整理和标注下开发了中文情感词汇本体库^[18]。

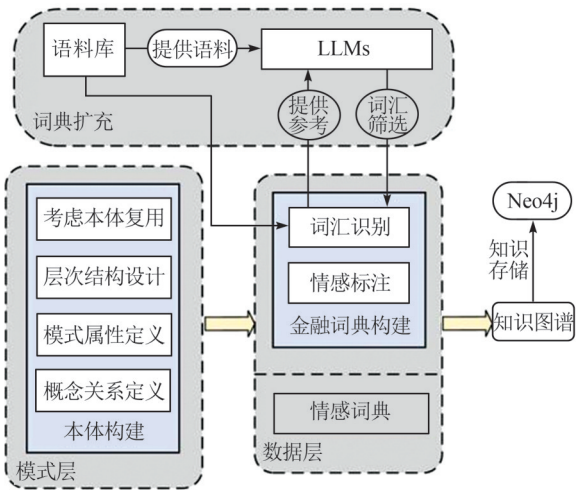


图1 知识图谱构建流程

Fig. 1 Knowledge graph construction process

该情感本体库对基本情绪进行了更为细致的分类，提供了情感方面的中文词汇以及对应的情绪分类和情感知识。为本文中的领域本体以及情感词汇提供了参考和依据。

结合对复用本体的分析和整理,本文将金融情感知识图谱分为两个大类:情绪类(emotion类)和词汇类(word类)。

定义属性与取值是构建本体中的重要步骤之一,涉及如何定义类的属性(Property)以及每个属性的取值范围(Value Range)。具体来说分为两步:定义类的对象属性(Object Property),定义类的数据属性(Data Property)。

1) 对象属性

表示类与类之间的关系,通常连接两个实体或资源。在本文的金融情感知识图谱中,有三类对象属性分别为:emotion->word, word->emotion, word->word,如表 1 所示。

表 1 本体类间关系

Table 1 Relationship between ontology classes

关系名称	类间关系
多义词(wasHomonyms)	word->word
代表(actedOnBehalfOf)	word->emotion
相关(wasRelatedTo)	emotion->word

2) 数据属性

该图谱中的情绪类数据属性集表示为Property(emotion)={E_ID, E_Name, Tendency, Valence, Arousal, Dominance},如表 2 所示。

词汇类数据属性集表示为Property(word)={W_ID, W_Name, Pos, WordType, Polarity, Intensity, SecondPolarity, SecondIntensity},如表 3 所示。

表 2 情绪类数据属性描述

Table 2 Emotional data attribute description

数据属性	描述
E_ID	情绪编号
E_Name	情绪名称
Tendency	情感倾向:消极,中立,积极
Valence	愉悦度(1~9)
Arousal	激活度(1~9)
Dominance	支配度(1~9)

表 3 词汇类数据属性描述

Table 3 Vocabulary data attribute description

数据属性	描述
W_ID	词汇编号
W_Name	词汇名称
Pos	词性
WordType	词汇分类(金融,情感)
Polarity	情感极性(消极,中立,积极)
Intensity	情感强度(1,3,5,7,9)
SecondPolarity	辅助情绪的情感极性(消极,中立,积极)
SecondIntensity	辅助情绪的情感强度(1,3,5,7,9)

在本文构建的金融情感本体中,虽然复用了情绪领域本体,但仍缺乏金融方面的相关数据,所以本文所构建的本体在金融知识的覆盖面与深度上仍有所欠缺。但在可用性上,该本体基于对 LLM 进行增强的应用需求进行构建,可以较好地满足实际应用的需求。

2.2 数据层

本文的数据层包括金融数据和情感数据两个部分。情感数据来自大连理工大学信息检索实验室构建的情感词汇本体表,而金融数据则是通过词汇识别并对其进行情感标注实现。

1) 词汇识别

本文研究的金融数据主要依赖于爬虫技术,从多种金融网站中爬取相关词汇和术语,构建初版的金融词典,然后设计 Prompt 模板结合 LLM 从构建的语料库中识别和补充未在金融词典中的金融术语,从而实现半自动的词典更新。

2) 情感标注

金融情感知识图谱与其他事件级知识图谱不同,金融情感知识图谱的构建过程是对金融领域的特定词汇进行情感标注。这些词汇的情感意义和金融背景紧密相关,我们不仅关注它们的含义,还希望赋予这些词汇相关的情感倾向和金融属性。在此过程中,词汇本身被标注上了情感极性和情感强度,而这些标注实际上就已经构成了关系,形成

了知识三元组。本文构建的金融词典也是按照情感词汇本体表的情感格式进行标注,每个词条的格式为:{词语名、词性类别、情绪类别、情感极性、情感强度、辅助情绪的情感极性、辅助情绪的情感强度}。

目前金融情感知识图谱一共收录了7 048个金融词汇和27 418个情感词汇。并将这些词汇的情绪分类根据8种基本情绪划分到22种派生情绪中。

同时该金融情感知识图谱也有着完备的层次结构。如图2所示。

以模式层为模板概念,将具体的知识实例以类间关系(实体1,关系,实体2)和实体属性(实体,属性,属性值)的三元组形式进行存储。其中,(实体1,关系,实体2)形式的三元组有74 851个,(实体,属性,属性值)形式的三元组有170 344个。

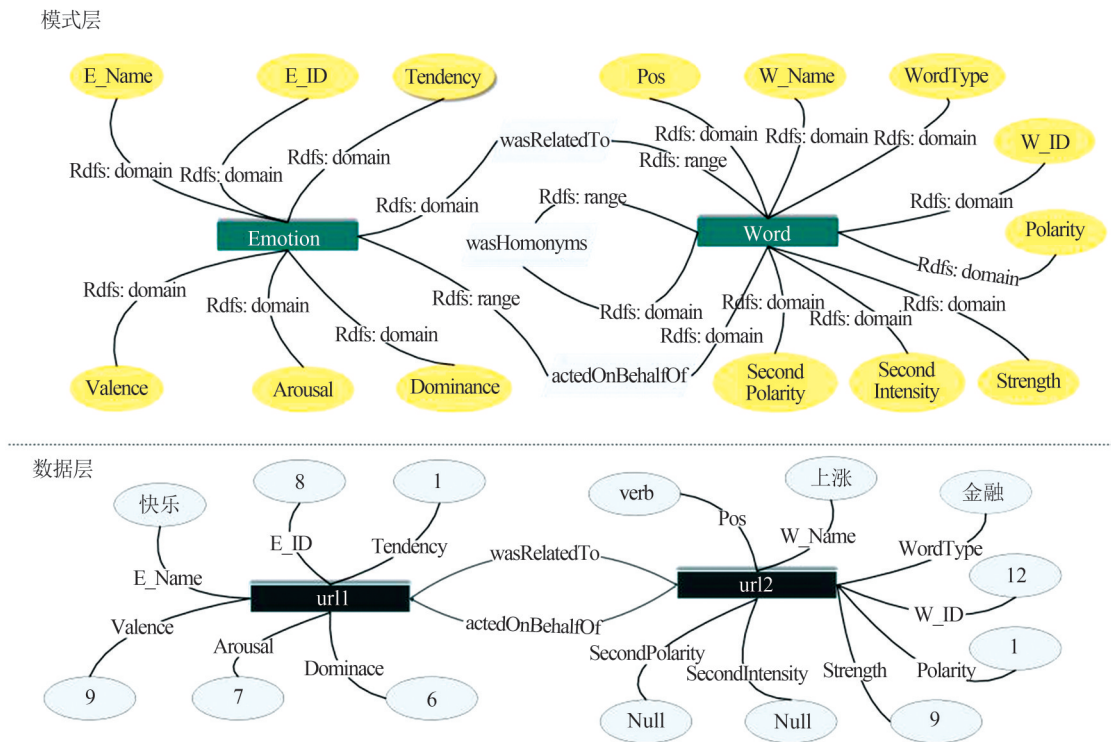


图2 金融情感知识图谱层次模型

在模式层中,绿色矩形框代表图谱内的本体类,用于对知识进行抽象分类;淡紫色菱形框代表类间关系,揭示类与类之间的关联;黄色椭圆框代表实体属性,用于描述本体类的特征。在数据层中,黑色矩形框代表具体实体数据,与之相连的灰色椭圆框为该实体对应的属性值。

Fig. 2 Hierarchical model of financial sentiment knowledge graph

In the schema layer, green rectangles represent the ontology classes within the knowledge graph, serving to conduct abstract classification of knowledge; light purple diamonds stand for inter-class relationships, revealing the associations between classes; and yellow ellipses denote entity attributes, which are used to describe the characteristics of ontology classes. In the data layer, black rectangles represent specific entity data, and the gray ellipses connected to them are the attribute values corresponding to the entity.

3 MKeFS 方法

本文提出的MKeFS方法主要包括两个部分。第一个是提示模板的构建,通过人工设计硬模板,将多段内容有机结合,形成一个结构化的输入框架;第二个是多层级知识增强,将前一个模块中输出的提示模板传输给LLM,进行多层级增强后得到情感分类结果进行分析。研究框架如图3所示。

3.1 提示模板构建

在本文的研究中,提示模板的构建是MKeFS方法的关键设计环节。为了充分利用知识图谱中

的情感知识和少样本提示的优势,本文设计了一种人工构建的硬模板,构建框架如图4所示。该模板结合了情感结构化知识、任务描述和少样本示例,旨在引导LLM更准确地理解金融文本的情感倾向。具体而言,提示模板由四个部分组成{情感知识,待分析句子,任务介绍,少样本提示},其中,情感知识和少样本提示分别作为多层级知识增强中的知识层和示例层,在增强LLM的过程中起到了至关重要的作用。

含有句子-标签对的标注数据集 $D = \{(v_1, l_1),$

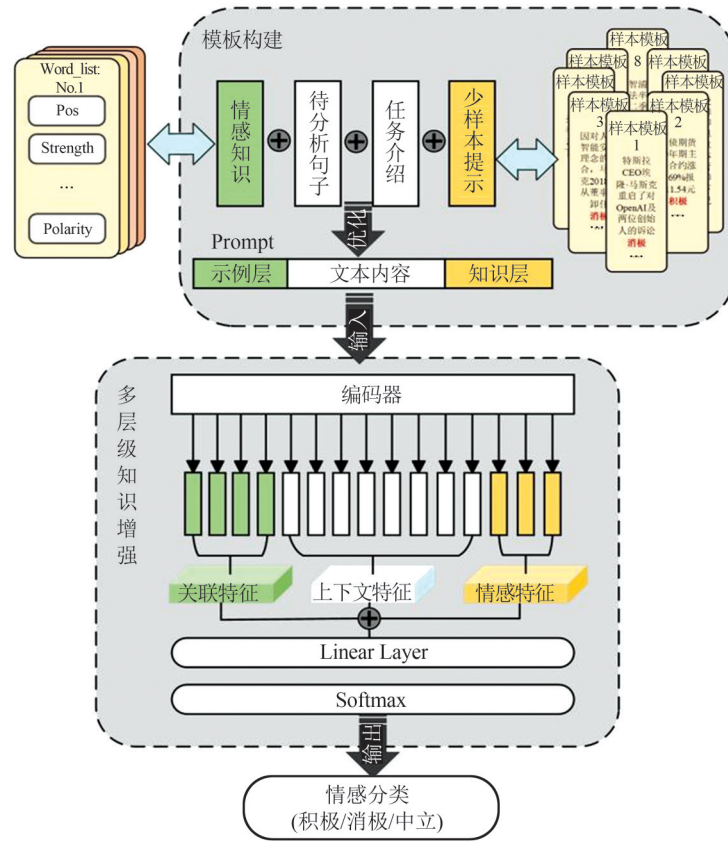


图3 增强大型语言模型情感分析流程图

Fig. 3 Enhanced sentiment analysis process of large language models

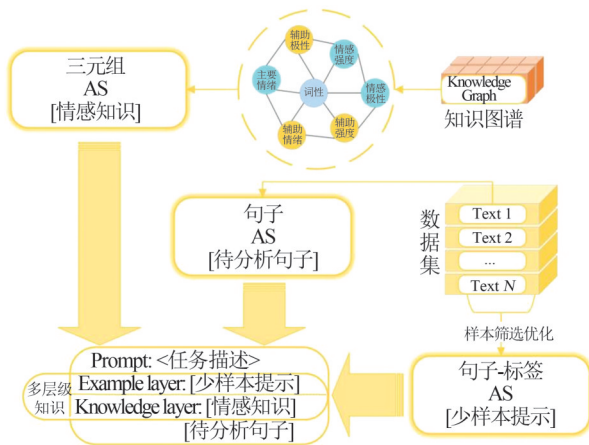


图4 提示模板构建

Fig. 4 Prompt template construction

$(v_2, l_2), \dots, (v_p, l_p), \dots, (v_n, l_n)\}$ 。其中, v 为句子; l 为该句子所对应的情感标签。金融情感知识图谱中包含 m 个单词。其中, 单词集合为 $T = \{w_1, w_2, \dots, w_q, \dots, w_m\}$ 。每个单词的知识集合可以表示为 $K = \{e_1, p_1, s_1, h, e_2, p_2, s_2\}$ 。其中, e_1 为主情绪; p_1 为主情绪的情感极性; s_1 为主情绪的情感强度; h 为词性; e_2 为辅助情绪; p_2 为辅助情绪的情感极性; s_2 为辅助情绪的情感强度。那么在给定句子 v_p 中提取到的单词

集合表示为:

$$T_p = \{w_x, \dots, w_y\}, T_p \subseteq T; w_x, w_y \in T \quad (1)$$

对于 T_p 中的每个单词 $w_i (w_i \in T_p)$, 可以从知识图谱中获取到相关的情感知识集合 $K(w_i)$, 表示如下:

$$K(w_i) = \{e_1(w_i), p_1(w_i), s_1(w_i), h(w_i), e_2(w_i), p_2(w_i), s_2(w_i)\} \quad (2)$$

最终从给定句子 v_p 中提取到的所有单词及其知识集合 K_p , 表示为:

$$K_p = \bigcup_{w_i \in T_p} K(w_i) = \{e_1(w_i), p_1(w_i), s_1(w_i), h(w_i), e_2(w_i), p_2(w_i), s_2(w_i)\} \quad (3)$$

从标注数据集 D 中选取 3 个标注示例作为少样本提示。每个示例包括一个金融文本句子及其对应的情感标签, 标签集合为 $L = \{\text{积极}, \text{消极}, \text{中立}\}$ 。本文的样本选择标准要求遵循标签覆盖广、领域代表性强、清晰度高等原则。其中, 标签覆盖度广是为了确保筛选的样本覆盖所有标签类别, 本文选择从每个标签类别中各选择一个代表性样本, 避免 LLM 因个例偏差导致对某些情感标签的识别能力缺失; 领域代表性强是为了使样本体现金融领域的特有事

件类型,确保LLM可以很好地学习到金融领域的异性特征;清晰度高是为了保证样本的情感倾向必须足够明确,减少LLM学习噪声,使LLM可以从示例中学习正确且清晰的标签映射规则。本文最终选取的样本集合为 $E = \{(v_i, l_i), (v_j, l_j), (v_k, l_k)\}$,严格满足上述三点标准。其中 $l_i = \text{积极}$, $l_j = \text{消极}$, $l_k = \text{中立}$,确保示例覆盖所有情感类别。 $v_x (x \in \{i, j, k\})$ 是对应情感标签下选取的文本,该文本被要求具有很强的金融特征,且不存在易混淆的词汇以及语境。

(4)式清晰地定义了提示模板的完整构建过程。

$$\text{Prompt}(v_p, K_p, T, E) = \text{Concat}(T, K_p, v_p, E) \quad (4)$$

其中, T 是任务描述; v_p 是待分析句子; K_p 是待分析句子中辅助分析的情感知识; E 是少样本提示样本集合; Concat 表示将各部分内容拼接优化成一个完整的提示模板。

3.2 多层级知识增强

MKeFS方法采用了多层级知识提示方法对LLM进行增强。第一层是示例层,使用少样本示例定义任务范式。第二层是知识层,通过前文构建的知识图谱补充情感知识。Prompt模板可以影响LLM的注意力机制以及对输入序列的分数评估,模板中的关键向量会优先吸引注意力引导LLM将更多的计算资源放在模板中的知识信息上。MKeFS方法中,将知识图谱中的情感知识嵌入到Prompt模板中,再结合少样本提示组合形成多层级知识增强。模板中的知识会通过LLM的Token嵌入和Transformer层的自注意力机制编码,隐式地对注意力权重进行调整,让注意力机制自发地分配更多权重给金融情感相关词或特征。

注意力机制用(5)式表达:

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

其中, $\mathbf{Q}$ 为Query向量; $\mathbf{K}$ 为Key向量; d_k 为Key向量的维度; $\mathbf{V}$ 为Value向量。使用Query向量和Key向量计算输入序列中每个词的相关性之后对Value向量进行加权求和,生成注意力输出。在示例层中,使用少样本示例构建上下文学习模块,其核心作用是通过隐式任务范式定义和动态特征加权引导LLM在无参数修改的条件下适应目标任务。通过示例中的句子-标签对构建任务映射规则,增强目标文本中与示例相似词汇的注意力权重,帮助LLM理解输入文本与情感标签之间的对应关系。知识层是将结构化知识图谱嵌入模块,其核心作用是通过显性领域知识注入和语义约束,引导LLM在注

意力机制中优先关注相应金融词汇的情感特征。将构建好的提示模板输入LLM后,经过模型编码器转化为知识向量 E_{Prompt} ,如(6)式所示:

$$E_{\text{Prompt}} = \text{Encoder}(\text{Prompt}) = E_{\text{PromptE}} \oplus E_{\text{PromptK}} \quad (6)$$

其中, E_{PromptE} 是示例层编码的任务范式向量,负责提供结构映射规则; E_{PromptK} 是知识层编码的领域知识向量,负责提供多维度情感知识,从而达到情感特征增强。两者结合进而对 $\mathbf{Q}$ 和 $\mathbf{K}$ 进行修正,如(7)式所示:

$$\mathbf{Q}' = \mathbf{Q} + \gamma_E \cdot E_{\text{PromptE}}, \quad \mathbf{K}' = \mathbf{K} + \gamma_K \cdot E_{\text{PromptK}} \quad (7)$$

其中, γ_E 和 γ_K 是完全隐性表达的参数,分别用于衡量嵌入向量对 $\mathbf{Q}$ 和 $\mathbf{K}$ 的影响程度。当 $\gamma = 0$ 时,LLM完全忽略Prompt带来的多层次知识增强,等同于未引入任何知识的原始注意力机制;当 $\gamma \neq 0$ 时, $\mathbf{Q}'$ 和 $\mathbf{K}'$ 受到了多层次知识增强带来的信息,会因提示模板中的示例层和知识层影响到注意力权重分配。

示例层通过任务范式修正Query向量,定义任务解决范式,当待分析文本中出现与示例层相似的情况,LLM在生成每个词的Query向量时,会携带示例层体现出的任务解决逻辑,相当于在原始Query向量的基础上,叠加一个任务范式向量,使LLM优先搜索与示例结构匹配的关联特征。知识层通过提供多维度的情感知识修正Key向量,目标是增强LLM对金融领域特征的敏感度,使待分析文本中的金融词汇获得更高的Key向量相似度,从而在注意力计算中激活更高的注意力权重,相当于在原始Key向量的基础上,叠加一个金融情感领域知识向量,增加情感关键词的检索概率。通过这样的修正可以在无需修改LLM参数的情况下,仅通过提示模板间接对注意力权重进行调整,就可生成比在原始注意力机制上更高的注意力分数。将调整后的 $\mathbf{Q}'$ 和 $\mathbf{K}'$ 用于注意力权重的计算,最后得到注意力分数 a_{ij} ,如(8)式所示:

$$a_{ij} = \text{soft max} \left(\frac{\mathbf{Q}'(\mathbf{K}')^T}{\sqrt{d_k}} \right) = \text{soft max} \left(\frac{\left((\mathbf{Q} + \gamma_E \cdot E_{\text{PromptE}})_i \cdot (\mathbf{K} + \gamma_K \cdot E_{\text{PromptK}})_j \right)^T}{\sqrt{d_k}} \right) \quad (8)$$

LLM的注意力机制,能够提升Prompt模板中相关词汇的注意力分数。而情感特征的提取也依赖于对金融与情感相关词汇的特征聚合。通过提高情感词的注意力分数,LLM在计算特征时会赋予

这些词更大的权重,如(9)式所示:

$$H_{\text{final}} = \sum_{j=1}^n a_{ij} V_j \tag{9}$$

其中, H_{final} 为最终的特征表示,其计算结果融合了原始文本特征和知识图谱的金融情感知识; a_{ij} 为第 i 个位置对第 j 个输入 Token 的经过知识层和示例层修正后的注意力分数; V_j 为第 j 个 Token 的值向量。

再将 H_{final} 输入全连接层,将其映射到情感类别空间(三维对应“积极/消极/中立”)得到对应情感类别得分 S ,如(10)式:

$$S = W_o \cdot H_{\text{final}} + b_o, W_o \in \mathbb{R}^{3 \times d}, b_o \in \mathbb{R}^3 \tag{10}$$

其中, W_o 为权重矩阵; b_o 为偏置向量。随后通过(11)式计算每个情感类别的概率分布:

$$P(y = k|x) = \frac{\exp(S_k)}{\sum_{j=1}^3 \exp(S_j)}, k \in \{\text{积极, 消极, 中立}\} \tag{11}$$

其中, S_k 表示对应情感类别的分数; S_j 是作为求和符号的循环变量。分子表示当前情感类别的指数化分数,分母表示所有类别指数化分数总和。最后概率最高的类别作为最终预测得到 $\hat{y}$,如(12)式所示:

$$\hat{y} = \arg \max_k P(y = k|x) \tag{12}$$

通过融合知识图谱和少样本学习组成了多层级知识增强。其中示例层提供句子-标签对引导 LLM 构建相关映射,知识层提供外部情感知识强化金融领域词汇的情感特征,两者结合主导最终情感概率分布。

4 实验与结果分析

4.1 数据集

使用爬虫在东方财富网(<http://finance.eastmoney.com>)中爬取 3 126 篇相关文章以构建数据集。为确保数据集在金融领域的广泛性和代表性,该数据集涵盖了 8 个金融粗粒度事件:政策变化,信用事务,股票事务,金融事务,业务操作,股东操作,管理事务,融资投资。

在数据预处理阶段,对收集到的数据进行清洗,将文章中与金融内容相关的头两段内容作为语料,同时对具体语料进行清洗以获得纯净的金融语料。再通过人工标注的方式,对语料中重要的金融事件进行标注,筛选得到符合条件的金融事件。

得到大量的金融事件提及后,制定情感标注规范,以达到多名标注人对情感分类的标准统一,再

对金融事件语料进行情感分类。最后得到的数据集共有 7 044 个金融事件,其中,积极的事件有 3 480 条,消极的事件有 2 208 条,中立的事件有 1 356 条。详细内容如表 4 所示。

表 4 数据集具体分布
Table 4 Specific distribution of data sets

金融粗粒度事件名称	积极	中立	消极
政策变化	0	201	0
信用事务	3	59	101
股票事务	1 542	139	1 274
金融事务	1 415	357	613
业务操作	70	33	44
股东操作	173	211	149
管理事务	0	216	16
融资投资	277	140	11

4.2 评价指标

在情感分类任务中,本文使用一组标准的评估指标衡量模型的分类效果,包括准确率(Accuracy)、查准率(Precision)和 $F1$ 。这些指标基于混淆矩阵计算,如表 5 所示,该矩阵展示了模型在不同情感类别上的分类结果。其中, TP_k : 真实为类别 k 且预测为类别 k 的样本数。 FN_{kj} : 真实为类别 k 但预测为类别 j 的样本数。

表 5 混淆矩阵
Table 5 Confusion matrix

真实类别	预测类别		
	积极(1)	中立(2)	消极(3)
积极(1)	TP_1	FN_{12}	FN_{13}
中立(2)	FN_{21}	TP_2	FN_{23}
消极(3)	FN_{31}	FN_{32}	TP_3

准确率(Accuracy)是最基本的分类性能指标,表示模型正确分类的样本占总样本的比例。查准率($Precision_k$)为模型预测为某类别的样本中实际为该类别的比例,反映模型对正类的识别能力。 $F1$ ($F1_k$)是查准率和查全率的调和平均数,是一个综合考虑查准率和查全率的指标,尤其在类别不平衡的情况下比准确率更具参考意义。这些指标的计算公式如下:

$$Accuracy = \frac{TP_1 + TP_2 + TP_3}{N} \tag{13}$$

$$Precision_k = \frac{TP_k}{TP_k + \sum_{j \neq k} FN_{jk}} \tag{14}$$

$$F1_k = \frac{2 \times Precision_k \times Recall_k}{Precision_k + Recall_k} \tag{15}$$

其中, N 为总样本数; $j, k \in [1, 3]$ 且分别对应积极, 中立, 消极。最后本文采用了加权平均法, 按类别样本量加权平均。

4.3 实验结果

4.3.1 实验设置

本文在单台工作站上完成全部实验, 配置为 16 vCPU Intel(R) Xeon(R) Platinum 8481C 以及 NVIDIA RTX 4090D(24 GB)。服务器操作系统为 Ubuntu 20.04, CUDA 12.1, Python 3.10.8, PyTorch 2.1.0。通用大模型均使用 Hugging Face Transformer 库加载, 统一 $\text{temperature}=0.3$ 和 $\text{top}_p=0.8$ 。

4.3.2 对比实验

为了验证 MKeFS 方法的有效性, 本文通过在多个开源大模型 (Qwen2.5-7B^[19]、ChatGLM3-6B^[20]、ChatGLM4-9B^[21]) 上对比有无使用该方法的指标分数, 评估了 MKeFS 方法对情感分类效果的提升, 同时使用了 FinBERT 等金融领域专业模型 (FinBERT^[22]、FinBERT-FP) 进行横向对比。进行对比的模型为:

Qwen2.5-7B^[19]: 该模型是由阿里云开发的大型预训练语言模型, 是 Qwen 系列的一部分。该模型具有 70 亿个参数, 能提供强大的自然语言处理能力, 支持多种任务和应用场景。

ChatGLM3-6B^[20]: 该模型是智谱 AI 推出的一个大型预训练语言模型, 它基于 Transformer 架构, 并针对中文场景进行了优化。该模型具有 60 亿个参数, 旨在提供高质量的文本生成、对话理解以及多种自然语言处理任务的支持。

ChatGLM4-9B^[21]: 该模型是智谱 AI 推出的 GLM-4 系列大型预训练语言模型, 相比于 ChatGLM3 具有更大的参数规模, 性能表现也更优异。

FinBERT^[22]: 该模型是一种基于 BERT 的金融领域情感分析模型, 通过在金融领域语料上进行预训练, 能够很好地处理金融文本中的专业语言。

FinBERT-FP: 该模型是在 Financial PhraseBank 数据集上微调的 FinBERT 模型, 专用于金融领域的情感分析任务, 可以很好的适应本文的对比实验。

实验结果如表 6 所示。从表 6 可以看出, 采用本文提出的 MKeFS 方法后, 各模型的情感分类性能均实现了显著提升, 使用了 MKeFS 方法后的 LLM 与未使用 MKeFS 方法的 LLM 相比, $F1$ 平均提升了 3.67 个百分点。ChatGLM4-9B 表现最为突出, 其各项指标均达到了最高水平, 与上一代的 ChatGLM3-6B 相比, 水平提升尤为显著。此外,

表 6 对比实验结果

Table 6 Comparative experiments results

模型名称	Accuracy	Precision	$F1$
Qwen2.5-7B	78.73	85.23	81.85
Qwen2.5-7B(MKeFS)	84.99	85.09	85.04
ChatGLM3-6B	78.65	76.23	77.42
ChatGLM3-6B(MKeFS)	82.50	82.15	82.32
ChatGLM4-9B	83.02	84.49	83.75
ChatGLM4-9B(MKeFS)	87.05	86.30	86.67
FinBERT	81.84	84.50	83.15
FinBERT-FP	83.06	85.14	84.09

注:“(MKeFS)”表示该模型采用了 MKeFS 方法进行数据增强。

Qwen2.5-7B 在与 ChatGLM3-6B 参数量相近的情况下, 也展现出了明显的指标优势, 这表明两个 LLM 之间存在较大的性能差距。但值得注意的是, 尽管 ChatGLM3-6B 的基础性能相对较低, 但在使用本文方法后, 其 $F1$ 提升幅度超过了性能最高的 ChatGLM4-9B, 这充分证明了 MKeFS 方法对性能较弱模型的优化效果更为显著。

使用经过 MKeFS 方法增强后的 Qwen2.5-7B 和 ChatGLM3-6B 横向对比 FinBERT 以及经过金融数据集微调后的 FinBERT-FP, 可以看出 MKeFS 方法的优势在于能够显著提升通用大模型在金融情感分析任务上的性能表现, 使其超越专为金融领域设计的专业模型。从数据对比来看, 未经 MKeFS 方法增强的 Qwen2.5-7B 和 ChatGLM3-6B 在准确率上均落后于 FinBERT, 而经过 MKeFS 方法优化后, 它们的准确率却能反超 FinBERT, 其中, 经过 MKeFS 方法优化后的 Qwen2.5-7B 甚至反超了经过大量金融数据微调后的 FinBERT-FP。这表明 MKeFS 方法不仅能弥补通用 LLM 在专业领域的先天不足, 还能通过知识增强等技术充分释放大模型的潜力, 最终在综合性能上领先金融领域专业模型。

4.3.3 消融实验

为了验证本文方法的多层级知识增强里不同层级对模型的提升有效性, 在各种实验参数保持不变的前提下, 选择在 ChatGLM4-9B 模型上进行了消融实验。实验结果如表 7 所示。

由表 7 可以观察到只保留示例层的方法比采取零样本提示的方法得到的准确率和 $F1$ 更低, 但结合知识层却可以得到最高分, 说明本文将知识图谱与少样本提示进行结合是至关重要的。单一靠少样本学习, 会很容易给模型制定一个错误的打分标准, 从而无法做到动态分析, 对模型的能力进行了限制。但在加入了知识图谱的知识后可以依靠图

表 7 消融实验结果

Table 7 Ablation results of different methods

方法名称	Accuracy	Precision	F1
ChatGLM4-ZS	83.02	84.49	83.75
ChatGLM4-KG	84.63	85.36	84.99
ChatGLM4-FS	79.26	85.61	82.31
ChatGLM4-MKeFS	87.05	86.30	86.67

注:ChatGLM4-ZS 表示不使用任何增强手段,采取零样本提示的方法进行实验;ChatGLM4-KG 表示在本文方法的基础上去掉示例层,只保留知识层进行实验;ChatGLM4-FS 表示在本文方法的基础上去掉知识层,只保留示例层进行实验;ChatGLM4-MKeFS 表示使用本文方法,融合示例层和知识层进行实验。粗体数字表示最优结果。

谱中的多维度情感知识对文本进行情感分析,再结合少样本提示之后可以更好的把握对情感特征的捕捉。为了更直观的看出不同方法的判断差异,本文详细的列出了 4 种方法的混淆矩阵结果进行对比,如图 5 所示,混淆矩阵为一个 12×3 的矩阵,可以看作 4 种方法对应的 3×3 矩阵,每个方法从左上到右下的对角线数据表示预测正确的标签,其他数据为预测错误的标签,从中可以看出每种方法的预测标签分布。

由图 5 可以看出,方法 ChatGLM4-ZS 与方法 ChatGLM4-KG 的混淆矩阵相似度极高,这表明二者预测标签的分布基本一致,即 ChatGLM4 对各类别的判断标准在知识注入前后是一致的。结合表 7 可知,在保留知识层的方法 ChatGLM4-KG 中,ChatGLM4 的分类准确性比零样本提示的方法 ChatGLM4-ZS 有明显提升,意味着知识图谱注入的多维度情感知识,精准修正错误预测的样本,将原本误判的标签更正为了正确标签,且未引入新的误判模式导致其他类别准确率下降。这表明金融情感知识图谱的注入未破坏 LLM 原有的分类逻辑框架,同时对正确预测数量产生了实质性提升,从而实现了对 LLM 在该任务表现上的全方位优化。

而方法 ChatGLM4-FS 表现最差的原因有两个:1) 积极类非常容易与中立类混淆(占主要错误),这说明 ChatGLM4 在识别积极情感时倾向于将其错误的分类为中立;造成这种错误的原因可能是因为

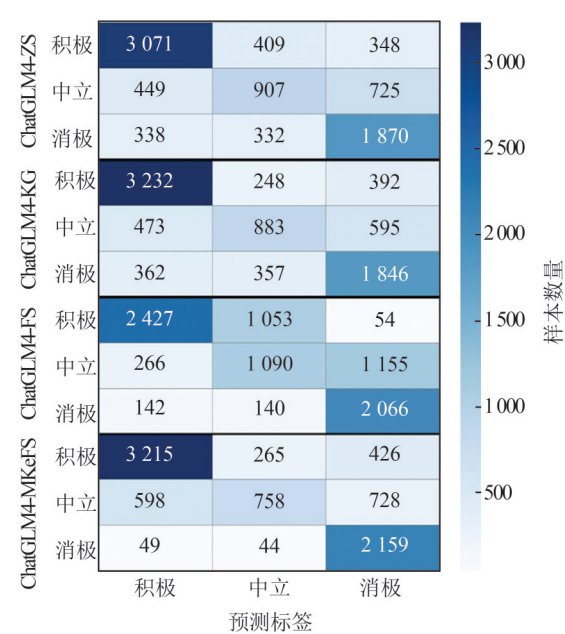


图 5 消融实验的混淆矩阵热力图

Fig. 5 Confusion matrix heatmap of ablation experiment

在少样本提示中,模型错误地混淆了积极和中立的情感特征,使其对积极类的预测标准偏高;2) 中立类也很容易与消极类混淆,且比例也明显较高;这说明少样本提示使模型对中立类的边界模糊,进而将中立情感的部分消极特征误解为完全消极。总的来说,单一的进行少样本提示,会导致模型的泛化能力减弱。这是因为在少样本提示过程中引入了特定样本的偏差,导致模型在整体数据分布上无法更广泛的理解分类标准,尤其在中立类和积极类上错误明显。

4.3.4 实例分析

为更直观地揭示方法间性能差异的内在机制,本文通过实验样例进行深入剖析,以验证上述方法表现差异较大的原因,具体分析结果如表 8 所示。

从表 8 可以看出,采用零样本提示的 ChatGLM4-ZS 方法、基于少样本学习的 ChatGLM4-FS 方法及本文提出的 MKeFS 方法在面对复杂语义场景时,呈现出显著的分类差异。以实例 1 中的事件为例,辞职虽然是一个偏消极的事件,但结合语境判断更倾向于一个中立的情感,ChatGLM4-FS 却

表 8 实例分析

Table 8 Sample analysis

实例编号	实验样例	分类结果			真实标签
		ChatGLM4-ZS	ChatGLM4-FS	MKeFS	
1	陈卓贤自公告日起辞任非执行董事及董事会副主席。	中立	消极	中立	中立
2	2023 年,中建信息的营业收入同比增长 2.93%。	积极	中立	积极	积极
3	两市合计成交 6 199 亿元,较昨日增加近 280 亿元。	中立	中立	积极	积极

错误的放大了其部分消极特征将其判定为完全消极,印证了该方法在中立类与消极类的分类边界上存在显著模糊性。实例2这一积极的表述,ChatGLM4-FS却误判为中立情感,进一步验证了其在情感分类边界识别上的缺陷。此外,在处理实例3这类隐含市场活跃度提升的复杂语句时,ChatGLM4-ZS与ChatGLM4-FS均错误判定为中立情感,而MKeFS方法则凭借引入的多维度情感特征,精准捕捉到交易量增长背后的积极信号,将其正确分类为积极情感。这充分表明,通过融合丰富的情感语义知识,MKeFS方法能够有效解决传统方法在情感分类边界模糊问题上的局限性,显著提升LLM在复杂语义场景下的分类准确率,为金融文本情感分析的实际应用提供了更可靠的解决方案。

5 结 语

针对大模型对金融情感分析任务的情感边界模糊的问题,本文提出了构建一个金融情感知识图谱并将其嵌入进提示模板中进行第一层级的知识增强,为大模型提供多维度的词汇情感信息,同时在提示模板中加入少量标注样本进行第二层级的知识增强,为大模型提供分类参考。实验结果表明,本文方法采用多层级知识增强可以有效地提高大模型在金融情感分析任务上的效率与准确性。

然而,本研究的方法也存在一定的局限性。当前主要依赖于提示模板的方式进行增强,手段相对单一,可能无法全面应对复杂、多变的金融情感分析场景。例如,同时处理多个在不同的语境有不同含义的金融词汇时,现有方法可能表现不佳。此外,由于金融领域本身的动态性和专业性,导致静态的金融情感知识图谱难以适应全新的市场情况和术语。

未来的研究应致力于探索更多样化的提示方法和其他增强技术,以进一步提升LLM的表现。同时,考虑在LLM训练阶段直接使用知识图谱进行增强,可以更深层次地挖掘金融情感特征,提高LLM对不同类型金融情感的理解能力。

参考文献:

- [1] LOUGHRAN T, MCDONALD B. When is a liability not a liability? textual analysis, dictionaries, and 10-ks [J]. *The Journal of Finance*, 2011, **66**(1): 35-65. DOI: 10.1111/j.1540-6261.2010.01625.x.
- [2] DEVLIN J, CHANG M W, LEE K, *et al.* BERT: Pre-training of deep bidirectional transformers for language understanding [EB/OL]. [2024-12-20]. <https://aclanthology.org/N19-1423.pdf>.
- [3] LIU Y H, OTT M, GOYAL N, *et al.* RoBERTa: A Robustly Optimized BERT Pretraining Approach [EB/OL]. [2019-07-26]. <https://arxiv.org/abs/1907.11692>.
- [4] FATOUROS G, SOLDATOS J, KOUROUMALI K, *et al.* Transforming sentiment analysis in the financial domain with ChatGPT [J]. *Machine Learning with Applications*, 2023, **14**: 100508. DOI: 10.1016/j.mlwa.2023.100508.
- [5] JI S X, PAN S R, CAMBRIA E, *et al.* A survey on knowledge graphs: Representation, acquisition, and applications [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, **33**(2): 494-514. DOI: 10.1109/TNNLS.2021.3070843.
- [6] BROWN T, MANN B, RYDER N, *et al.* Language models are few-shot learners [EB/OL]. [2020-07-22]. <https://arxiv.org/pdf/2005.14165>.
- [7] PAN S R, LUO L H, WANG Y F, *et al.* Unifying large language models and knowledge graphs: A roadmap [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2024, **36**(7): 3580-3599. DOI: 10.1109/TKDE.2024.3352100.
- [8] BOZKURT A. Tell me your prompts and I will make them true: The alchemy of prompt engineering and generative AI [J]. *Open Praxis*, 2024, **16**(2): 111-118. DOI: 10.55982/openpraxis.16.2.661.
- [9] TODD A, BOWDEN J, MOSHFEGHI Y. Text-based sentiment analysis in finance: Synthesising the existing literature and exploring future directions [J]. *Intelligent Systems in Accounting, Finance and Management*, 2024, **31**(1): e1549. DOI: 10.1002/isaf.1549.
- [10] ZHANG B Y, YANG H Y, LIU X Y. Instruct-FinGPT: Financial sentiment analysis by instruction tuning of general-purpose large language models [EB/OL]. [2023-06-22]. <https://arxiv.org/abs/2306.12659>.
- [11] QIAN L F, ZHOU W P, WANG Y, *et al.* Fino1: On the transferability of reasoning enhanced LLM to finance [EB/OL]. [2025-02-27]. <https://ui.adsabs.harvard.edu/abs/2025arXiv250208127Q/abstract>.
- [12] YU W H, ZHU C G, LI Z T, *et al.* A survey of knowledge-enhanced text generation [J]. *ACM Computing Surveys*, 2022, **54**(11s): 1-38. DOI: 10.1145/3512467.
- [13] WAN Z, CHENG F, MAO Z Y, *et al.* GPT-RE: In-context learning for relation extraction using large language models [C]//Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, Singapore: Association for Computational Linguistics, 2023: 3534-3547. DOI: 10.18653/v1/2023.emnlp-main.214.
- [14] SUN J W, DU Z X, CHEN Y R, *et al.* Knowledge graph

- tuning: Real-time large language model personalization based on human feedback [EB/OL]. [2024-05-30]. <https://arxiv.org/abs/2405.19686>.
- [15] YORAN O, WOLFSON T, BOGIN B, *et al.* Answering questions by meta-reasoning over multiple chains of thought [C]//Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. Singapore. Stroudsburg: ACL, 2023: 5942-5966. DOI: 10.18653/v1/2023.emnlp-main.364.
- [16] WU Y K, HU N, QI G L, *et al.* Retrieve-Rewrite-Answer: A KG-to-Text Enhanced LLM Framework for Knowledge Graph Question Answering[EB/OL]. [2023-09-20]. <https://arxiv.org/abs/2309.11206>.
- [17] SAHOO P, SINGH K A, SAHA S, *et al.* A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications [EB/OL]. [2024-02-05]. <https://rotmandigital.ca/wp-content/uploads/2024/09/A-Systematic-Survey-of-Prompt-Engineering-in-Large-Language-Models.pdf>.
- [18] 徐琳宏, 林鸿飞, 潘宇, 等. 情感词汇本体的构造[J]. 情报学报, 2008, 27(2): 180-185. DOI: 10.3969/j.issn.1000-0135.2008.02.004.
- XU L H, LIN H F, PAN Y, *et al.* Constructing the affective lexicon ontology[J]. *Journal of the China Society for Scientific and Technical Information*, 2008, 27(2): 180-185. DOI: 10.3969/j.issn.1000-0135.2008.02.004(Ch).
- [19] YANG A, YANG B S, ZHANG B C, *et al.* Qwen2.5 Technical Report[EB/OL]. [2024-12-19]. <https://arxiv.org/abs/2412.15115>.
- [20] DU Z X, QIAN Y J, LIU X, *et al.* GLM: General language model pretraining with autoregressive blank infilling[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg: ACL, 2022: 320-335. DOI: 10.18653/v1/2022.acl-long.26.
- [21] ZENG A H, XU B, WANG B W, *et al.* ChatGLM: A Family of Large Language Models from GLM-130B to GLM-4 All Tools[EB/OL]. [2024-06-18]. <https://arxiv.org/abs/2406.12793>.
- [22] DOGU A. FinBERT: Financial sentiment analysis with pre-trained language models [EB/OL]. [2019-08-27]. <https://arxiv.org/abs/1908.10063>.

□