

# 大语言模型驱动的跨域属性级情感分析\*

李诗晨, 王中卿, 周国栋

(苏州大学 计算机科学与技术学院, 江苏 苏州 215008)

通信作者: 王中卿, E-mail: [wangzq@suda.edu.cn](mailto:wangzq@suda.edu.cn)



**摘要:** 属性级情感分析作为一种细粒度情感分析方法, 目前在许多应用场景中都具有重要作用. 然而, 随着社交媒体和在线评论的日益广泛以及各类新兴领域的出现, 使得跨领域属性级情感分析面临着标签数据不足以及源领域与目标领域文本分布差异等挑战. 目前已有许多数据增强方法试图解决这些问题, 但现有方法生成的文本仍存在语义不连贯、结构单一以及特征与源领域过于趋同等问题. 为了克服这些问题, 提出一种基于大语言模型 (large language model, LLM) 数据增强的跨领域属性级情感分析方法. 所提方法利用大模型丰富的语言知识, 合理构建针对跨领域属性级情感分析任务的引导语句, 挖掘目标领域与源领域相似文本, 通过上下文学习的方式, 使用领域关联关键词引导 LLM 生成目标领域有标签文本数据, 用以解决目标领域数据缺乏以及领域特异性问题, 从而有效提高跨领域属性级情感分析的准确性和鲁棒性. 所提方法在多个真实数据集中进行实验, 实验结果表明, 该方法可以有效提升基线模型在跨领域属性级情感分析中的表现.

**关键词:** 属性级情感分析; 大语言模型 (LLM); 迁移学习; 数据增强; 领域适应

**中图法分类号:** TP18

中文引用格式: 李诗晨, 王中卿, 周国栋. 大语言模型驱动的跨域属性级情感分析. 软件学报. <http://www.jos.org.cn/1000-9825/7156.htm>

英文引用格式: Li SC, Wang ZQ, Zhou GD. LLM Enhanced Cross Domain Aspect-based Sentiment Analysis. Ruan Jian Xue Bao/Journal of Software (in Chinese). <http://www.jos.org.cn/1000-9825/7156.htm>

## LLM Enhanced Cross Domain Aspect-based Sentiment Analysis

LI Shi-Chen, WANG Zhong-Qing, ZHOU Guo-Dong

(School of Computer Science and Technology, Soochow University, Suzhou 215008, China)

**Abstract:** As a fine-grained sentiment analysis method, aspect-based sentiment analysis is playing an increasingly important role in many application scenarios. However, with the ubiquity of social media and online reviews, cross-domain aspect-based sentiment analysis faces two major challenges: insufficient labeled data in the target domain and textual distribution differences between the source and target domains. Currently, many data augmentation methods attempt to alleviate these issues, yet the target domain text generated by these methods often suffers from shortcomings such as lack of fluency, limited diversity of generated data, and convergent source domain. To address these issues, this study proposes a method for cross-domain aspect-based sentiment analysis based on data augmentation from a large language model (LLM). This method leverages the rich language knowledge of large language models to construct appropriate prompts for the cross-domain aspect-based sentiment analysis task. It mines similar texts between the target domain and the source domain and uses context learning to guide the LLM to generate labeled text data in the target domain with domain-associated keywords. This approach addresses the lack of data in the target domain and the domain-specificity problem, effectively improving the accuracy and robustness of cross-domain sentiment analysis. Experiments on multiple real datasets show that the proposed method can effectively enhance the performance of the baseline model in cross-domain aspect-based sentiment analysis.

**Key words:** aspect-based sentiment analysis; large language model (LLM); transfer learning; data augmentation; domain adaptation

\* 基金项目: 国家自然科学基金 (62076175, 61976146)

收稿时间: 2023-08-31; 修改时间: 2023-12-06; 采用时间: 2024-01-05; jos 在线出版时间: 2024-05-29

情感分析是自然语言处理领域中的一个重要研究方向,其应用广泛,涉及舆情监测、社交媒体分析、产品评价等领域<sup>[1,2]</sup>.其中,情感分析基于分类粒度,可以被分为3个级别,分别是篇章级别的情感分类,句子级别的情感分类和属性级别的情感分类<sup>[3,4]</sup>.属性级情感分析的任务是对文本中的每个属性词进行情感分类,即判断该属性词所表达的情感是正面、负面还是中性.如图1所示,关于餐厅领域的评论中对于“上菜速度”这一属性词所表达的情感是正面的,对菜品的情感是负面的.然而,各个领域间的评论文本存在语义分布差异.以图1中电脑领域的评论为例,该评论中提到“电池消耗很快!”,其中“电池消耗”为属性词,“快”为该属性词的意见词,明显地传达对“电池消耗”属性的负面情绪.相对地,在餐饮领域的评论中,关于“上菜速度”这一属性词,“快”却表达了积极的情绪.因此,不同领域间的语义分布差异会影响现有属性级情感分析方法的表现,使其大多只能在特定领域或特定类型的数据上表现良好<sup>[5]</sup>.但是,在面对特定领域的属性级情感分类任务时,往往又缺乏足够的有标签数据<sup>[6]</sup>.同时,由于语料标注的成本高,获取大规模和高质量训练语料的难度很大,所以目前主流的属性级情感分析方法<sup>[7-9]</sup>难以适应不同领域或不同类型数据的属性级情感分析需求.因此,跨领域属性情感分类任务被提出.

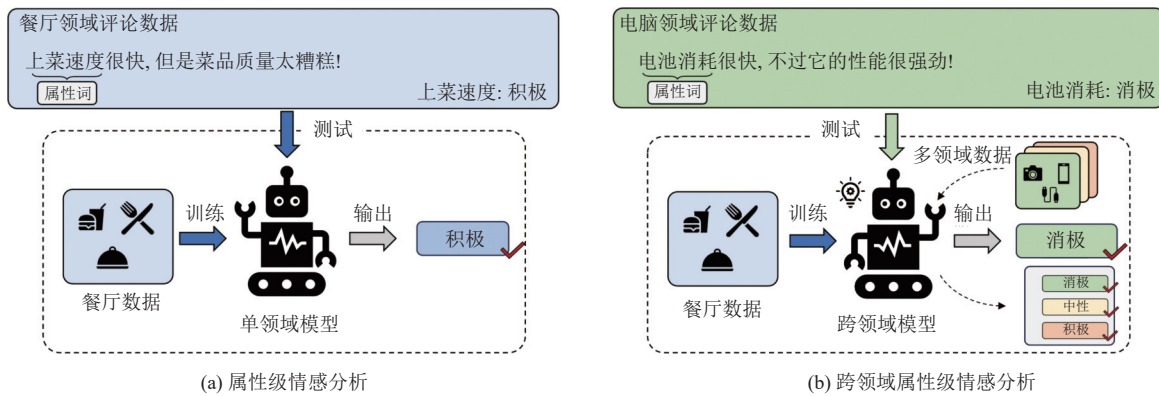


图1 跨领域与单领域属性级情感分析任务示例

跨领域情感分类存在两个难点,第1个难点是目标领域缺乏有标签数据,第2个难点是跨领域文本特征差异大<sup>[10]</sup>.为了缓解这些问题,一系列领域适应性 (domain adaptation) 的跨领域情感分类方法被提出<sup>[11-14]</sup>.在早期的工作中,研究者通过挖掘领域无关的特征信息来实现跨领域的情感迁移学习.近年来随着预训练模型的发展,研究者使用领域适应性的迁移学习方法并结合预训练模型来解决这一问题.同时,这类方法也被应用于跨领域的属性级情感分类方法中,如Wang等人<sup>[15]</sup>利用领域不变句法关系来抽取属性词与观点词,Gong等人<sup>[16]</sup>通过挖掘不同领域间不变的词性和句法依赖关系来解决端到端的跨领域属性级情感分类任务.这些方法可以适当缓解领域差异来增强模型的泛化性,但是,这些方法仅使用源领域中的有标签数据,无法缓解目标领域缺乏有标签数据的问题,从而导致这类方法缺乏对目标领域中特有属性词与情感词的敏感性<sup>[17]</sup>.

为了进一步解决这些问题,近期有研究者提出一种新的领域适应性方法,名为跨领域的数据增强<sup>[17]</sup>.以餐厅领域为源领域,电脑领域为目标领域,则不同结构跨领域数据增强方法如图2所示.其中,基于预训练语言模型 (pretrained language model, PLM) 的方法如图2(a)所示,这一类方法通常基于样例模板,通过区间掩码 (span mask) 方式遮盖源领域特定文本,再利用训练好的掩码模型或序列到序列模型,通过文本补全方式生成目标领域的有标签数据<sup>[18,19]</sup>.基于PLM的方法相较之前缓解领域差异性的方法可以改善目标领域缺乏有标签数据的困境,但是通过此类方法生成的目标领域文本仍有许多不足.首先,不同领域之间的语义和语法差异可能导致此类方法生成的文本不够自然流畅,或属性词情感标签错误,如图2(a)所示.其次,使用此类方法进行数据增强时,大多基于样例模板,难以保持目标领域文本的多样性.此外,模型在进行目标领域数据生成时,往往会受到源领域数据的束缚,导致生成的文本过于倾向于源领域的表达风格和特点.

为了克服这些问题,本文提出了一种新的跨领域数据增强方法,即利用大语言模型 (large language model,

LLM) 进行数据增强. LLM 如 ChatGPT<sup>[20]</sup>和 Claude<sup>[21]</sup> 通过学习大规模文本数据, 拥有极其庞大的知识储备, 能够更好地理解自然语言内容的上下文信息, 并且通过简单文本指引即可生成高质量的文本用以缓解目标领域缺乏有标签数据的问题, 如图 2(b) 所示. 然而, LLM 虽然在文本生成方面具备强大的潜力, 但是仅通过简单文本指引 LLM 生成的文本过于单一, 同时可解析性较低, 为此需要构造合适的指引和指令来确保生成的文本满足预期的要求. 因此, 为了更好地利用 LLM 生成目标领域的有标签数据, 从而解决跨领域属性级情感分类, 本文提出一种基于 LLM 数据增强的跨领域属性级情感分析方法, 如图 2(c) 所示. 该方法第 1 步针对跨领域属性级情感分类任务, 合理构造指令语句用以引导 LLM 完成目标领域文本结构化生成任务; 第 2 步, 挖掘目标领域与源领域相似文本, 提取样例级别的文本生成关键词; 第 3 步, 通过上下文学习方式使用领域关联关键词, 引导模型高效生成目标领域有标签文本数据, 用以解决目标领域数据缺乏以及领域特异性问题, 从而有效提高跨领域情感分析方法的准确性和鲁棒性.

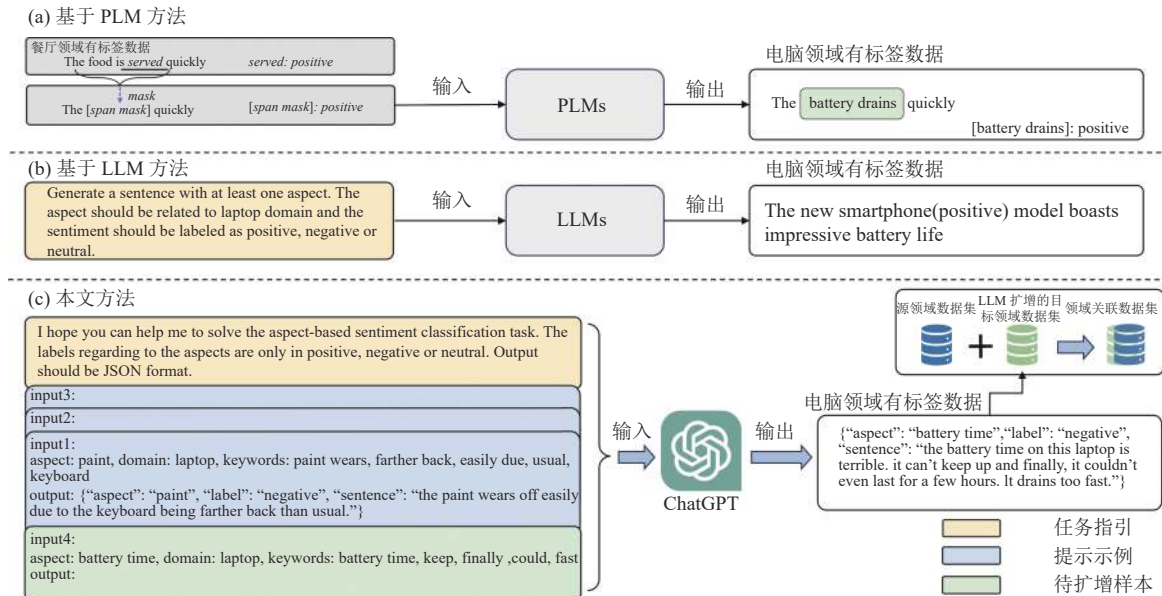


图 2 不同结构跨领域数据增强方法示意图

本文第 1 节介绍单领域与跨领域属性级情感分析的相关方法和数据增强技术研究现状. 第 2 节介绍本文提出的 LLM 与数据增强驱动的跨领域属性级情感分析方法. 第 3 节通过对比实验验证了本文所提方法的有效性. 第 4 节总结全文并提出未来工作方向.

## 1 相关工作

### 1.1 属性级情感分类

在社交媒体和在线评论平台的普及推动下, 人们能够在网络上轻易地发布和分享对产品、服务和事件等的观点与评价. 这种趋势促使了对细粒度情感分析方法, 如属性级情感分类 (aspect-based sentiment classification, ABSC) 的研究需求日益增长. 属性级情感分类旨在从文本中识别和分析特定属性词的情感极性<sup>[3,22]</sup>. 早期的属性级情感分类方法主要依赖于人工设计的模板来提取文本中与目标属性词相关的情感信息<sup>[23]</sup>. 然而, 这类方法的效果在很大程度上取决于人工设计特征的质量, 因此存在明显的局限性. 随着深度神经网络的进步, 当前的研究主要依赖于神经网络模型来处理属性级情感分类任务<sup>[24-26]</sup>. Tang 等人<sup>[22]</sup>提出了 TD-LSTM, 此方法通过使用两个 LSTM 模型分别编码属性词的上下文信息, 加强了模型对属性级情感分类任务的理解. 注意力机制在构建神经网络模型用以解决属性级情感分析中也得到了广泛应用. 例如, ATA-E-LSTM<sup>[27]</sup>利用基于注意力机制的 LSTM 模型

来完成属性级情感分类任务,而 IAN<sup>[28]</sup>和 MemNet<sup>[29]</sup>则通过使用多层注意力机制,融合属性词与上下文之间的语义关联,从而更好解决属性级情感分析任务。

在预训练模型的发展下,研究者们开始尝试将其应用于属性级情感分类任务。Li 等人<sup>[30]</sup>使用预训练模型叠加线性层,在多个属性级情感分析任务中取得了良好的效果。Huang 等人<sup>[31]</sup>则通过在预训练模型中融合句法信息,使用图卷积神经网络 (graph convolutional network, GCN) 来完成属性级情感分类任务。预训练生成模型如 T5<sup>[32]</sup>, BART<sup>[33]</sup>等的出现,启发了研究者们利用生成模型来处理 ABSC 任务。Zhang 等人<sup>[34]</sup>提出的 GAS 将传统的 ABSC 任务转化为生成任务,在多个实验集上展现出了优秀的表现。此外, Liu 等人<sup>[9]</sup>也使用生成模型进行属性类别情感分类,实验结果证明,相较于传统的分类模型,使用生成模型的表现更加优异。此外, Yan 等人<sup>[35]</sup>基于 BART 模型构建的统一生成方法,以及 Lu 等人<sup>[36]</sup>基于 T5 模型构建的统一信息抽取方法,在 ABSC 任务上均表现良好,但它们仅使用预训练生成模型,缺乏对结构化句法知识的利用。因此,为了解决这一问题, Fei 等人<sup>[37]</sup>将结构化句法知识与 UIE 结合,提出了 LasUIE 模型,在许多属性级情感分析任务中表现良好。

虽然目前提出的属性级情感分析方法可以在特定的领域取得较好的结果,但是不同领域间的语义分布差异以及新兴领域缺乏标签数据等问题会影响现有属性级情感分析方法的表现。因此,本文提出一种基于 LLM 数据增强的通用方法,通过合理生成目标领域的有标签数据,可以有效强化现有属性级情感分析方法在跨领域问题中的表现。

## 1.2 跨领域情感分类

跨领域情感分类是指在某个特定领域中训练情感分类模型,并将该模型应用于不同领域的文本数据中<sup>[5]</sup>。在实际应用中,由于不同领域的语言结构、词汇和文化背景的差异,单一领域的情感分析模型往往难以适应其他领域的文本数据<sup>[6]</sup>。因此,跨领域情感分析的研究意义在于提高模型的泛化能力和适应性,从而更好地满足现实应用的需求<sup>[38]</sup>。

现有的跨领域情感分析方法大多可以分为以下几类。挖掘领域无关的特征信息来实现跨领域的情感迁移学习的方法,如 Biltzer 等人<sup>[39]</sup>利用结构相关性学习算法 (structural correspondence learning, SCL) 学习不同领域之间的特征信息; Ziser 等人<sup>[40]</sup>通过挖掘不同领域中的共享 (枢纽) 特征 (pivot features) 特征来强化语言模型在跨领域情感分类中的表现。基于领域对抗学习的方法,这类方法利用领域对抗学习模块强化模型对于不同领域信息地判别,从而强化模型的泛化性。如 Du 等人<sup>[41]</sup>在领域预训练中使用对抗学习模块,从而强化模型对于领域信息识别。此外, Xue 等人<sup>[42]</sup>针对跨领域情感分析问题使用两组特征判别器结合对抗学习来实现目标域情感判别。基于半监督学习,这类方法利用无标签数据对预训练模型进行领域后训练,如 Zhou 等人<sup>[11]</sup>,针对情感分析任务设计模型损失函数,并使用大量无标签数据对模型进行训练,从而强化模型解决多领域情感分析任务的能力。虽然这些方法可以有效缓解领域差异给模型带来的性能损失,但是缺乏对于目标领域特有语义结构的学习。

## 1.3 跨领域的数据增强方法

数据增强可以有效缓解领域迁移所带来的问题,并帮助模型更好地适应不同领域的情感分类任务,从而提升模型的性能和泛化能力。传统的数据增强方法可简单分为 3 类<sup>[43]</sup>: 第 1 类为改写类 (paraphrasing),即通过改变句子的结构、单词的顺序或者使用同义词等方式来生成新的句子。第 2 类为噪声类 (noising),即通过在原始文本中添加噪声,如随机替换单词、删除单词或者交换单词位置等方式来生成新的句子。第 3 类为采样类 (sampling),即通过从原始文本中随机采样一些片段,如单词、短语或者句子等,然后将它们组合成新的句子。使用数据增强方法可以增加训练数据的多样性,从而提高模型的泛化能力。在跨领域情感分类任务中,数据增强可以缓解数据稀缺的情况,从而有效提高跨领域情感分析效果。但是,传统的数据增强方法存在语法错误、语义失真和上下文难捕捉等问题,而且依赖预定义规则,无法处理复杂任务。此外,这些方法仅局限于利用有标签的源领域数据,缺乏对目标领域无标签数据的利用,因此,这类方法同样缺乏对目标领域中特有属性词与情感词的敏感性。

近期研究者们提出一种新的领域适应性方法,名为跨领域的数据增强,通过直接生成目标领域的有标签数据用以解决目标领域缺乏有标签数据的问题。如 Yu 等人<sup>[19]</sup>基于掩码语言模型 (mask language model, MLM) 通过学

习源领域有标签数据基于文本重构生成目标领域有标签数据. Li 等人<sup>[44]</sup>基于预训练生成模型使用源领域中的有标签数据来生成具有细粒度注释的目标域数据, 从而完成跨领域情感分析. 此类方法可以提高模型对于目标领域中特有属性词与情感词的敏感性, 但是目前跨领域的数据增强方法生成的目标领域文本仍有许多不足, 如文本语义不连贯, 文本结构单一, 文本特征与源领域趋同等. 因此, 本文提出一种基于 LLM 数据增强的通用方法, 通过合理引导 LLM 生成目标领域的有标签数据, 同时保证生成数据的多样性和流畅性.

## 2 基于 LLM 与数据增强的跨领域属性级情感分类方法

本节将介绍本文提出的 LLM 数据增强驱动的跨领域属性级情感分析方法. 该方法第 1 步针对跨领域属性级情感分析任务, 合理构造指令语句用以引导 LLM 完成目标领域文本结构化生成任务; 第 2 步, 挖掘目标领域与源领域相似文本, 提取样例级别的文本生成关键词; 第 3 步, 通过上下文学习方式, 基于示例样本, 引导模型使用领域关联关键词引导模型高效生成目标领域有标签文本数据, 从而解决目标领域数据缺乏问题. 以餐厅领域作为源领域, 电脑领域作为目标领域, 则本文方法流程如图 3 所示.

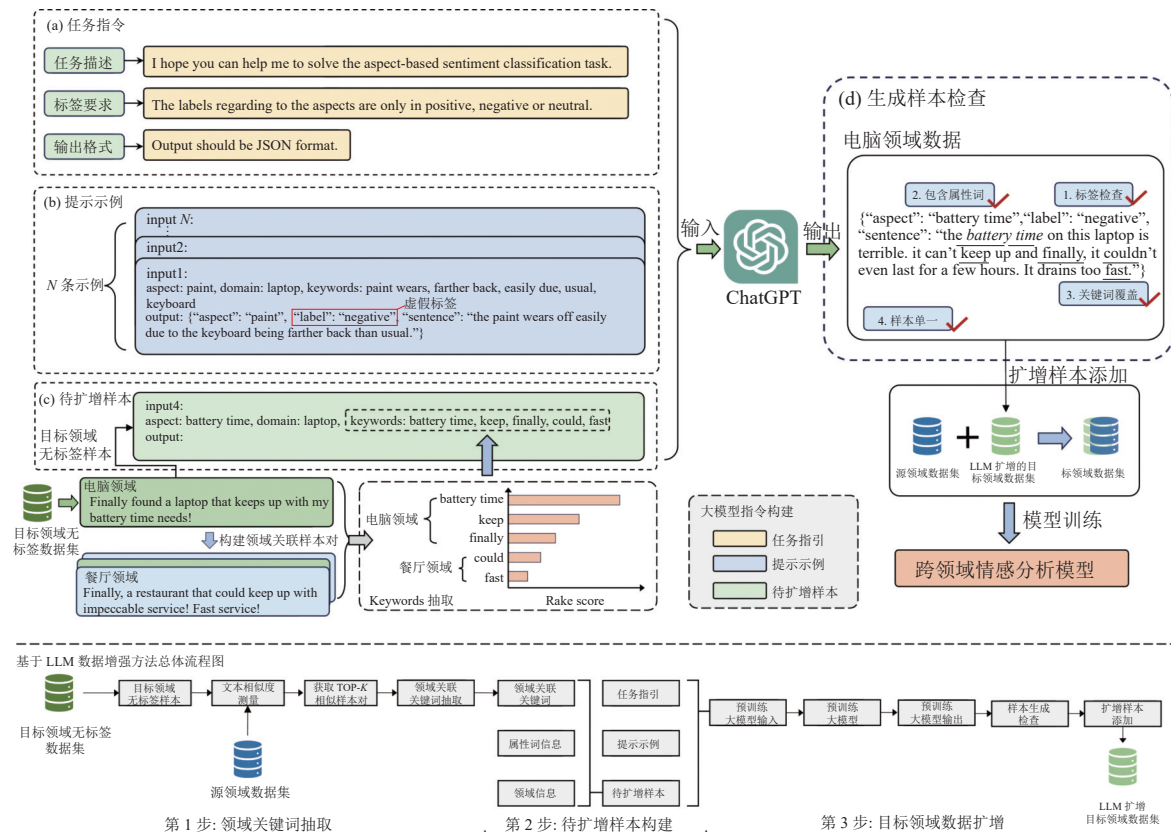


图 3 LLM 与数据增强驱动的跨域属性级情感分析方法流程图

### 2.1 LLM 与指令构建

LLM 通常是指包含百亿或更多参数量的语言模型<sup>[45]</sup>, 如 GPT-3<sup>[46]</sup>、ChatGPT<sup>[20]</sup>和 LLaMa<sup>[47]</sup>等. 这类模型通过学习大规模文本数据, 在许多自然语言处理任务中展现了强劲的性能. 相较于 PLM, LLM 的模型大小, 预训练数据量和总计算量都得到了显著扩展, 使得这类模型能够更好地理解自然语言内容的上下文信息, 并生成高质量的文本. 尽管 LLM 经过大规模语料库训练, 具备解决通用任务的能力, 但在执行特定任务时, 其能力可能不会显式展现. 因此, 为了确保生成的文本满足预期要求, 需要为特定任务设计合适的指引和指令, 对 LLM 进行合理的引导并

挖掘 LLM 解决特定任务的潜力.

本文旨在探讨利用 LLM 解决跨领域属性级情感分析任务. 因此, 本文针对跨领域属性级别情感分析任务, 合理构造指令语句用以引导 LLM 完成目标领域文本结构化生成任务. 其中, 本文设计了针对属性级别情感分析任务的数据生成指令“*I hope you can help me to solve the aspect-based sentiment classification task. The labels regarding to the aspect are only in positive, negative or neutral. Output should be JSON format.*”, 如图 3(a) 所示. 相较于先前工作直接生成自然语言样本<sup>[48,49]</sup>, 本文使用特定指令使得 LLM 输出结构化的 JSON 格式文本, 相比于非结构化文本, 结构化文本可以更方便地进行信息提取, 进而便于后续的模式训练. 同时, 本文还在指令中明确规定了生成样本的标签信息. 此外, 为了更好地在指令构建部分引导模型生成结构化数据, 本文针对跨领域属性级别情感分析任务, 设计了多个样例属性, 用以包含源领域与目标领域信息, 如属性词 (aspect word), 领域信息 (domain), 以及跨领域文本关键词信息 (keywords), 用以引导 LLM 高效生成目标领域样本.

## 2.2 样例级领域关联关键词抽取

如前文所述, 本文针对跨领域属性级别情感分析任务, 构造指令语句用以引导 LLM 完成目标领域文本结构化生成任务. 然而, 仅凭单一指令文本来引导 LLM 生成文本, 可能会导致生成结果的文本结构较为单一, 缺乏丰富性和多样性. 为了解决此问题, 本文在引导语句中添加领域相关的关键词, 以提高生成文本的质量和多样性. 这些领域关键词与源领域和目标领域的相关文本内容紧密相关, 能够引导 LLM 在生成文本时专注领域相关的特定知识和主题, 令其提供更具针对性的内容表达.

例如, 在图 3 中, 源领域为餐厅领域, 目标领域为电脑领域. 本文通过在引导语句中加入与源领域内容相关的关键词, 如“fast”, 以引导模型生成与之相关的目标领域评论信息, 从而有效构建源领域与目标领域文本关联, 进而缓解领域差异问题. 如图 3 中所示, 源领域餐厅评论中“fast”有积极的含义, 如果仅使用源领域数据训练模型, 模型由于缺乏对目标领域数据的学习, 针对电子设备领域中出现的用电速度快, 可能会做出错误的判断. 本文通过构建样例级领域关键词, 引导 LLM 产生与源领域关联的目标领域数据用于模型学习和训练, 可以有效缓解这一问题.

本文抽取样例级领域关键词可分为以下 3 步, 第 1 步使用 SimCSE<sup>[50]</sup> 计算待扩增目标领域样本与源领域所有样本文本相似度; 第 2 步, 获取源领域与目标领域待文本相似度 Top- $k$  样本对, 第 3 步针对这些样本对, 使用 RAKE 算法<sup>[51]</sup> 抽取关键词, 获取得分最高的 Top- $P$  个短语作为关键词. 具体流程如图 4 所示.

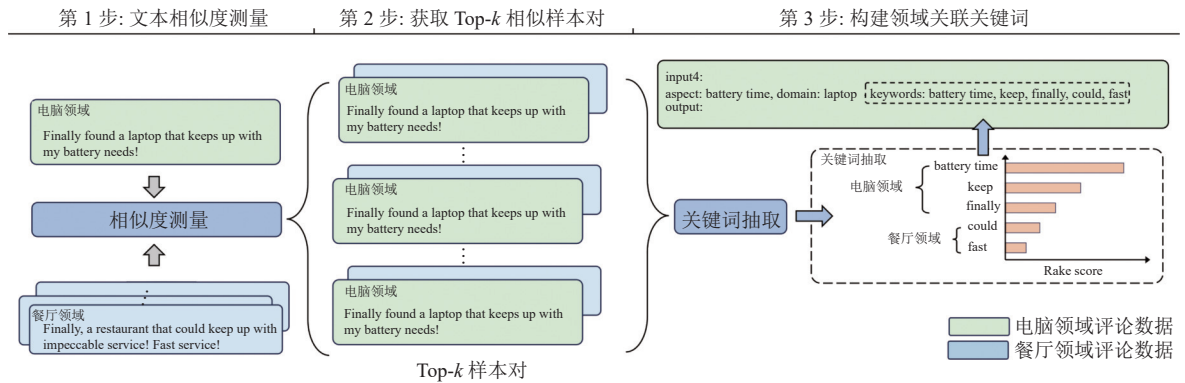


图 4 样例级领域关联关键词抽取流程图

具体来说, 针对有标签源领域数据  $\mathcal{D}^s = \{(x_i^s, y_i^s)\}_{i=1}^{N_s}$ , 以及无标签目标领域数据  $\mathcal{D}^t = \{(x_j^t)\}_{j=1}^{N_t}$ , 本文基于下式获取其对应的隐含层向量:

$$H_s = \text{SimEncoder}(x_1^s, x_2^s, \dots, x_i^s) \quad (1)$$

$$H_t = \text{SimEncoder}(x_1^t, x_2^t, \dots, x_j^t) \quad (2)$$

其中, *SimEncoder* 采用已训练好模型.

随后针对  $[h_1^s, h_2^s, \dots, h_i^s] \in H_s$ ,  $[h_1^t, h_2^t, \dots, h_i^t] \in H_t$  得到相似度矩阵  $G$ , 其中,

$$G_{(i,j)} = \text{Sim}(h_i^s, h_j^t) \quad (3)$$

$$\text{Sim}(h_i^s, h_j^t) = \frac{h_i^{sT} \cdot h_j^t}{\|h_i^s\| \|h_j^t\|} \quad (4)$$

得到源领域和目标领域文本相似度矩阵  $G$  后, 针对目标领域样本  $x_j^t$  计算其对应的源领域样本 Top- $k$  相似度, 即  $\text{Top}_k^{\text{sorted}} = \{G_{(i_1,j)}, G_{(i_2,j)}, \dots, G_{(i_k,j)}\}$ , 其中,  $G_{(i_1,j)} \geq G_{(i_2,j)} \geq \dots \geq G_{(i_k,j)}$ ; 随后, 针对  $x_j^t$  构建其对应的源领域最相似文本对集合  $a_j = (x_{i_1}^s, x_j^t), (x_{i_2}^s, x_j^t), \dots, (x_{i_k}^s, x_j^t)$ . 最终得到目标领域数据  $[x_1^t, x_2^t, \dots, x_j^t] \in D^t$  所对应的源领域 Top- $k$  样本对集合  $A_{\text{sim}} = a_1, a_2, \dots, a_j$ .

接下来的部分将介绍快速自动关键字提取 (rapid automatic keyword extraction, RAKE) 算法的应用和关键词的抽取过程. 该算法作为一种常用的关键词抽取算法, 通过计算每个单词与其他单词共同出现的频率作为其重要程度的评分, 从而能够自动从文本中提取关键词. 本文将此算法应用于源领域和目标领域的样本对集合  $A_{\text{sim}}$ , 以获取其关键词. 这些关键词可以反映出源领域和目标领域的特定知识和主题, 进而引导 LLM 更专注地生成与这些知识和主题相关的文本.

在 RAKE 算法的应用过程中, 本文针对目标领域样本  $x_j^t$ , 首先将源领域和目标领域的样本对即  $a_j$  作为输入, 然后算法会基于词频输出每个样本对的关键词及其相应的得分. 接着, 本文从中选取得分最高的 Top- $P$  个短语作为关键词集合  $k_i$ , 并将其添加到引导语句中, 以引导语言模型生成文本. 最后, 为了保证生成文本与当前样本的属性紧密相关, 本文将当前样本的属性词作为领域关键词添加到引导语句中, 通过这种方式使得语言模型在生成文本时将更加关注属性词, 从而生成更符合源领域和目标领域的文本.

### 2.3 上下文学习

随着 LLM 能力的不断提高, 上下文学习方法已成为使用 LLM 解决自然语言处理问题的新范式, 因其基于上下文指示命令以及少量的示例, 即可学习任务指示并完成文本预测<sup>[46]</sup>. 如图 3(b) 所示, 完整上下文学习指令可以形式化定义为:

$$C = \{I, (x_1^t, d, k_1, y_1^t), (x_2^t, d, k_n, y_2^t), \dots, (x_n^t, d, k_n, y_n^t)\} \quad (5)$$

其中,  $C$  表示本文使用完整指示命令,  $I$  为跨领域属性级情感分析任务指示命令文本,  $d$  为目标领域名称,  $n$  为添加示例数量,  $k_n$  为关键词集合,  $x_n^t$  为目标领域评论文本; 由于目标领域样本缺乏标签, 本文参考 Zhang 等人<sup>[52]</sup>的工作, 在指示命令中添加的  $y_n^t$  为虚假标签.

因此, 基于上下文学习的 LLM 可以定义为:

$$P(y_j | x) = f_{\text{LLM}}(y_j, C, x) \quad (6)$$

$$\hat{y} = \text{argmax}(P(y_j | x)) \quad (7)$$

其中,  $\hat{y}$  为 LLM 经由指令指导生成的目标领域有标签样本.

### 2.4 数据清洗

LLM 作为概率文本生成模型, 不可避免会生成低质量数据. 因此, 本文针对 LLM 生成的文本数据进行筛选、修复和处理, 以去除不必要的噪声、错误或无效数据, 确保增强后的数据质量和多样性, 如图 3(d) 所示. 具体来说, 本文采取以下措施自动筛选文本数据. 删除不符合 JSON 数据格式的样本. 删除标签错误的样本, 本文使用数据集仅包含 3 类样本标签, 即  $Y = \{\text{positive}, \text{neutral}, \text{negative}\}$ , 因此, 当生成样本的标签不在定义的标签集合中时, 样本将被从生成数据集中剔除. 删除文本中不包含指定属性词的样本, 确保生成的文本数据与指定的属性词相关. 删除关键词覆盖率小于  $\delta$  的样本, 确保生成的文本数据覆盖指定的关键词, 用以提高文本的多样性. 剔除重复样本, 确保生成的文本数据的多样性和独特性.

### 2.5 模型训练

本文采用预训练指令文本生成模型 T5 作为基础模型以验证本文方法有效性. T5 基于编码器-解码器结构用

以完成文本生成任务. 本文将属性级情感分类任务转化为指令指导的生成任务. 假设具有文本输入  $X = [I, x_1, x_2, \dots, x_m]$ , 首先基于 T5 模型的 Encoder 部分将输入文本  $X$  编码为  $H^e$ , 即:

$$H^e = T5Encoder([I, x_1, x_2, \dots, x_m]) \quad (8)$$

其中,  $H^e \in R^{m \times d}$ ,  $d$  为隐藏层维度,  $I$  为指令语句; 此外, 为了简化公式, 这里省略了自动添加的  $\langle \text{s} \rangle$  结束符.

假设具有  $X = [I, x_1, x_2, \dots, x_m]$ , 以及标签序列  $Y = [y_1, \dots, y_n]$ , 则生成模型的输出可以形式化的定义为:

$$P(Y|X) = \prod P(y_i|X, Y) \quad (9)$$

本文模型可以通过 T5 模型中的文本编码器和文本解码器来计算每一步的  $P = P(y|X, Y)$ . 具体来说, 通过上文所述步骤得到  $H^e$  后, 将其输入基于 T5 的文本解码器中, 通过  $H^e$  和当前时刻之前文本解码器的输出  $\hat{Y}_{<i>t-1</i>}$  来计算  $p_t$ , 即:

$$h_t^d = T5Decoder(H^e; \hat{Y}_{<i>t-1</i>}) \quad (10)$$

其中,  $h_t^d \in R^d$ ,  $h_t^d$  为 Decoder 最终输出的隐含层向量, 则:

$$P_t = \text{Softmax}(W_p h_t^d + B_p) \quad (11)$$

其中,  $W_p, B_p$  均为模型参数.

在得到生成词概率  $P_t$  后, 以交叉熵函数作为目标函数, 通过不断减小交叉熵损失函数的值来调节模型参数从而完成模型的训练.

$$\mathcal{L} = - \sum_{i=1}^N \sum_{j=1}^K y_i \log \hat{y}_i + \frac{\lambda}{2} |\theta_y|^2 \quad (12)$$

其中,  $i$  为数据样本的索引,  $j$  为词表索引,  $y_i$  为目标生成词,  $\hat{y}_i$  为预测生成词,  $N$  为样本总数,  $K$  为词表大小,  $\theta_y$  为模型参数,  $\lambda$  为 L2 正则参数.

### 3 实验

#### 3.1 实验数据与实验设置

本文使用 Restaurant, Service, Laptop, Device 这 4 个领域常用的英文属性级情感分析数据集<sup>[53-56]</sup>, 构建 10 个 (源领域, 目标领域) 数据集对; 同时, 使用扫地机器人, 香水, 牛奶, 智能开关领域的评论数据 (<https://github.com/blcunlp/CCD-ASQP>), 并去除其中无属性词的样本, 构建 4 个中文属性级情感分析数据集对; 用以验证本文方法在跨领域属性级情感分析任务中的有效性. 其中, 本文参考前文实验设定<sup>[17]</sup>, 删除了 (Device, Laptop) 和 (Laptop, Device) 两个英文评论领域对, 因为其评论内容过于接近, 难以测评模型跨领域属性级情感分析能力. 详细数据分析可见表 1.

表 1 实验数据统计

领域名称	训练集	测试集	扩增数据
Device	1394	691	1378
Laptop	2297	631	2257
Restaurant	4284	2252	2574
Service	1840	886	1798
扫地机器人	3169	463	—
香水	855	533	831
牛奶	613	357	587
智能开关	615	369	600

本文所提出的基于 LLM 数据增强的跨领域属性级情感分类方法使用 ChatGPT 作为 LLM 模型, 通过 API 接口调用 gpt-3.5-turbo-0301 完成数据增强任务, 扩增数据如表 1 所示. 同时, 使用 T5-base 模型作为基准模型, 用以



验证 LLM 数据增强的有效性. 本文实验基于 GeForce RTX 3090 (24 GB) GPU 完成, 其中 Batch size 设置为 16, T5 模块的学习率设置为  $2E-4$ , 并使用 Adam<sup>[57]</sup> 优化器来优化和更新模型参数. LLM 生成中的 temperature 采样超参数设置为 0.2, 样本示例数量  $n$  设置为 3. 此外, 本文实验结果为随机初始化参数并重复实验 3 次后的平均结果, 并选用准确率 (accuracy, Acc) 作为评价指标.

### 3.2 对比方法与实验结果

为评估本文模型的有效性, 本文选取了 3 类基线模型进行实验对比, 结果见表 2. 其中, BERT<sup>[58]</sup>, T5<sup>[32]</sup>, BART<sup>[33]</sup> 为常用的基础预训练模型, ChatGPT<sup>[20]</sup> 与 LLaMa<sup>[47]</sup> 为 LLM, ADSPT<sup>[14]</sup>, PADA<sup>[59]</sup>, ACSC<sup>[9]</sup> 为近期较为强劲的跨领域情感分析模型, 以下是对这些对比方法的简要介绍.

表 2 本文模型与对比方法在不同领域对上的实验结果 (%)

Method	$D \rightarrow R$	$D \rightarrow S$	$L \rightarrow R$	$L \rightarrow S$	$R \rightarrow D$	$R \rightarrow L$	$R \rightarrow S$	$S \rightarrow D$	$S \rightarrow L$	$S \rightarrow R$	Avg
BERT	80.37	85.21	81.13	81.94	88.28	77.81	84.54	90.74	69.73	79.75	81.95
BART	82.33	86.46	82.73	86.00	90.16	77.50	86.68	91.03	70.84	81.35	83.51
T5	81.79	87.58	79.53	85.33	<u>93.92</u>	78.29	87.02	<u>93.49</u>	70.36	81.17	83.85
ChatGPT	<b>86.32</b>	83.63	<b>86.32</b>	83.63	90.30	<u>79.23</u>	83.63	<u>90.30</u>	<u>79.23</u>	<b>86.32</b>	<u>84.89</u>
LLaMa	81.86	<u>87.91</u>	81.04	<u>86.97</u>	92.03	74.81	86.57	92.76	73.18	81.43	83.86
ADSPT	82.86	<u>87.13</u>	84.81	83.30	91.17	78.45	83.86	91.46	71.47	82.37	83.69
PADA	83.28	87.23	84.66	85.20	91.10	78.34	<b>87.79</b>	93.04	69.78	82.93	84.34
ACSC	82.19	86.57	83.57	84.54	89.44	75.75	85.97	91.75	71.16	82.28	83.32
Ours	<u>85.70</u>	<b>87.81</b>	<u>86.10</u>	<b>87.36</b>	<b>94.36</b>	<b>82.73</b>	<u>87.70</u>	<b>93.78</b>	<b>80.03</b>	<u>85.75</u>	<b>87.13</b>

注: 加粗数据为当列最好结果, 下划线数据为当列次好结果

BERT, 即使用源领域训练数据对 Bert-base 模型进行微调, 随后使用 [CLS] 向量完成目标领域属性级情感分类任务. T5 与 BART 即使用 T5 与 BART 预训练生成模型, 经源领域训练数据微调后直接以自然语言输出目标领域无标签属性词的情感倾向. ChatGPT 是一种基于 GPT-3.5 架构的 LLM, 该模型通过大规模训练并经由指令调优, 在许多自然语言处理任务中都展现了强大的能力, 本文将其应用于跨领域属性级情感分析任务. LLaMa 是近期开源的 LLM, 在多个自然语言处理任务中都展现了作为基座模型的强大能力, 本文基于 Alpaca-LoRA (<https://github.com/tloen/alpaca-lora>) 使用源领域训练数据对 LLaMa-7B 进行指令调优, 用以完成跨领域属性级情感分析任务. ADSPT 采用软模板 (soft prompt) 获取领域特有特征并使用对抗学习方法获得领域不变特征, 从而完成跨领域情感分析任务. PADA 通过构建样例级模板建立源领域与模板领域特征, 并使用 T5 的编码器部分完成跨领域的情感分类任务. ACSC 基于模板使用预训练生成模型完成跨领域属性级情感分析工作.

从表 2 的实验结果中可以看出, BERT 模型在跨领域属性级情感分析任务上的表现相对较低, 这可能是由于其对于属性级别情感分析任务的理解较低, 因此领域泛化性较差. 然而, 预训练生成模型 T5 和 BART 的表现则相对较好. 因为生成模型可以基于自然语言文本来理解与解决属性级情感分析任务, 使得其在处理跨领域任务时有更好的泛化能力. 大语言模型 ChatGPT 和 LLaMa 的表现则超过了基础预训练模型 BERT, 并且与 T5 和 BART 的表现相当. 这可能是由于 LLM 能够更好地理解属性级情感分析任务, 从而拥有较好的领域泛化性能, 所以在多个领域对中表现优秀. 此外, 指令调优的方法也提高了 LLM 在跨领域属性级情感分析中的表现. 但在本文的实验中发现, LLM 对于中性情感的识别能力较低, 因此导致 LLM 的平均准确率下降. 针对跨领域情感分析任务的特定模型 ADSPT、PADA 和 ACSC 的表现也相当出色, 这些模型通过特定的策略来处理跨领域任务, 例如 ADSPT 使用软模板和对抗学习来获取领域不变特征, PADA 通过构建样例级模板来建立源领域与目标领域特征, ACSC 则通过使用预训练生成模型与模板来完成属性级情感分析任务. 此外, 本文方法通过合理构建针对跨领域属性级别情感分析任务的引导语句, 挖掘目标领域与源领域相似文本并通过上下文学习的方式使用领域关联关键词引导模型生成目标领域有标签文本数据, 有效缓解领域差异问题, 进而强化现有属性级情感分析方法在跨领域问题中的表现. 其中, 本文方法相较 ChatGPT, 在多个领域对上的平均准确率中提高了 2.3%, 相较于

基础预训练生成模型 T5 提高了 3.28%。表 2 中的实验结果表明, 本文所提出方法可以有效地解决跨领域属性级情感分类任务, 同时, 在多个数据集上的优秀表现也说明本文所提出的模型具有较强的泛化性。

### 3.3 中文数据集实验结果

为了验证本文方法在不同语言数据集上的有效性, 本节使用来自扫地机器人领域的中文评论数据作为各类方法的训练数据, 并分别使用来自牛奶, 香水以及智能开关领域的中文评论数据作为测试集。同时, 本节实验的基线模型 (<https://huggingface.co/bert-base-multilingual-uncased>, <https://huggingface.co/google/mt5-base>) 均使用多语言模型用以支持中文评论数据的训练与测试。表 3 实验结果表明, 本文方法在中文数据集中也有较强的有效性。如在香水领域评论数据作为测试集时, 由于训练集与测试集中的样本偏差较大, 导致基线模型的表现均较差。但是基线模型结合本文提出的基于 LLM 的数据增强方法, 通过高效引导大模型生成目标领域有标签数据, 可以有效缓解领域差异问题, 从而有效提高各类基线方法的表现。

表 3 中文数据集准确率实验结果 (%)

模型	牛奶	香水	智能开关
BERT	77.59	57.60	85.63
BERT+本文方法	86.27	81.98	92.68
T5	84.31	65.66	86.72
T5+本文方法	91.60	83.18	94.85

### 3.4 消融实验

为了进一步验证本文模型每个模块的有效性, 本节选用两个领域对的数据集对本文提出方法进行消融实验研究, 结果如表 4 所示。首先, 去除 LLM 数据增强部分后, 本文方法在两个领域对中的表现 (准确率) 均有较大下降, 其中  $S \rightarrow L$  领域对中的准确率下降了 9.94%。由此可以看出, 基于 LLM 的数据增强部分通过合理引导 LLM 生成目标领域有标签数据, 可以有效地缓解领域差异问题, 从而强化模型解决跨领域属性情感分析任务的性能。当去除针对 LLM 数据增强的上下文学习部分时, 即仅使用指令引导 LLM 生成目标领域有标签数据, 模型的性能也有大幅度的下降。因为缺乏上下文的示例指引, LLM 很难生成符合属性级情感分析任务的数据, 从而导致模型性能的大幅下降。此外, 去除关键词提示和领域标识符也会降低模型的性能, 因为, 关键词提示信息与领域标识符可以有效引导模型结合源领域信息, 生成符合目标领域表达特点的数据。同时, 样例级别的关键词提示可以有效提高文本生成的多样性, 从而缓解领域差异问题。因此, 本节消融实验也说明, 本文方法中不同模块对于模型解决跨领域属性级别情感分类任务均有帮助。

表 4 消融实验结果 (准确率 (%))

模型	$S \rightarrow L$	$R \rightarrow L$
本文方法 (T5)	80.03	82.73
-LLM数据增强	70.36	78.29
-上下文学习	70.81	78.90
-关键词提示	74.22	79.57
-领域标识符	78.71	81.15

### 3.5 不同数据增强方法适配基准模型实验

为了验证本文方法的普适性及有效性, 本节选取 3 个领域对, 将本文提出的基于 LLM 数据增强的方法与不同数据增强方法应用于多个不同架构的基准模型中进行比对。其中, ED<sup>A</sup><sup>[60]</sup>为传统数据增强方法, DA<sup>2</sup>LM<sup>[17]</sup>为基于预训练模型的数据增强方法。

表 5 中可以看出, 本文提出的基于 LLM 数据增强的方法可以有效地提升各类基准模型在跨领域属性级情感分析任务中的性能。其中, 在  $S \rightarrow L$  领域对中, 由于领域差距较大, 各类方法在该领域对的表现都不佳, 但在

基准模型中使用本文提出的基于 LLM 数据增强的方法,通过合理构建针对跨领域属性级别情感分析任务的引导语句,高效引导大模型生成目标领域有标签数据,可以有效缓解领域差异问题,从而有效提高各类方法的表现.此外,在  $D \rightarrow R$  与  $R \rightarrow L$  领域对中,本文方法相较其他数据增强方法,可以更加有效地提高基准模型的表现.因此,本节实验也说明,本文方法可以简洁有效地适配多种结构模型,令其更加有效地解决跨领域属性级别情感分类任务.

表 5 不同数据增强方法适配基准模型实验结果(准确率(%))

模型	数据增强方法	$S \rightarrow L$	$D \rightarrow R$	$R \rightarrow L$
BERT	—	69.73	80.37	77.81
	EDA	71.31	80.03	77.30
	DA <sup>2</sup> LM	74.20	81.44	78.12
	本文方法	<b>76.39</b>	<b>82.99</b>	<b>78.99</b>
T5	—	70.36	81.79	78.29
	EDA	72.33	81.35	78.70
	DA <sup>2</sup> LM	75.87	82.74	79.40
	本文方法	<b>80.03</b>	<b>85.70</b>	<b>82.73</b>
ADSPT	—	71.47	82.86	78.45
	EDA	73.01	82.69	79.11
	DA <sup>2</sup> LM	76.48	83.18	79.38
	本文方法	<b>78.17</b>	<b>84.36</b>	<b>80.12</b>
LLaMa	—	73.18	81.86	74.81
	EDA	74.20	81.37	75.22
	DA <sup>2</sup> LM	75.88	82.55	76.38
	本文方法	<b>78.91</b>	<b>84.05</b>	<b>79.02</b>

### 3.6 LLM 生成文本分析

不同领域间的文本相似度差异会极大影响各类模型解决跨领域属性级情感分析任务的性能,因此本节通过计算不同领域间文本相关性(Jaccard similarity score)<sup>[61]</sup>,从而验证本文基于 LLM 数据增强的方法可以有效地提高源领域与目标领域的文本相似度.如图 5 所示,本文使用 5 个(源领域,目标领域)领域对组合进行实验.其中,分别使用源领域数据与目标领域数据,以及源领域数据混合本文基于 LLM 进行数据增强的数据,进行文本相似度计算.从图 5 的结果中可以看出,本文基于 LLM 数据增强的方法在多个不同领域对中都可以有效地提高领域文本相似性,说明本文方法对于弥合不同领域之间的文本差距非常有效.

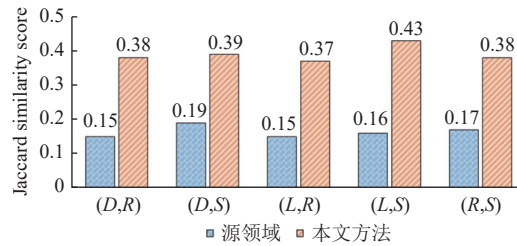


图 5 LLM 生成文本相似度分析

为了进一步探究 LLM 文本生成质量,本文对 LLM 所产生的低质量数据进行详尽分析.实验结果如表 6 所示,其中数据格式不匹配为生成数据未能满足本文设定的 JSON 格式要求;而标签不一致,则是指在 LLM 生成的目标领域评价数据中,所含情感标签并非指定的情感标签.此外,其他类型的错误包括属性词遗漏、关键词缺失等.实验结果显示,在 LLM 生成的低质量数据中,数据格式不匹配和标签错误占有较高比例.然而,与待扩增的样本总量相比,LLM 错误生成的数据比例相对较低,这表明本文提出的基于 LLM 的数据增强策略具备较高的鲁棒性.

表 6 LLM 生成数据质量分析

领域名称	待扩增数据	数据格式不符	标签不一致	其他错误	扩增数据
Device	1394	7	8	1	1378
Laptop	2297	20	11	9	2257
Restaurant	2600	8	13	5	2574
Service	1840	18	13	11	1798
香水	855	10	8	6	831
牛奶	613	7	11	8	587
智能开关	615	6	4	5	600

### 3.7 生成样例分析

在图 6 的样例分析中, 本文以餐厅领域为源领域, 电脑领域为目标领域, 展示了本文利用 LLM 进行数据增强的样例.

原始电脑领域评论文本: { <b>"aspect"</b> : "microsoft software", <b>"sentence"</b> : "Everything is so easy to use, mac software is just so much simpler than <i>microsoft software</i> ."}	
<b>LLM 生成指令</b>	<b>LLM 输出文本</b>
<b>Instruction:</b> I hope you can help me to solve the aspect-based sentiment classification task. The labels regarding to the aspects are only in positive, negative or neutral. Output should be JSON format.	<b>LLM Output:</b> { <b>"aspect"</b> : "microsoft software", <b>"label"</b> : "negative", <b>"sentence"</b> : "The <i>Microsoft software</i> is not as user-friendly as the Mac software; it could be much simpler."}
<b>Examples:</b> <b>input 1:</b> aspect: 18-inch, domain: laptop, keywords: inch laptop, sized, 18 output: {"aspect": "18-inch", "label": "positive", "sentence": "this is an over-sized, 18-inch laptop."}	
<b>input 2:</b> aspect: cords, domain: laptop, keywords: headphones plus network connection, cords going every direction, power plus cords coming, cords coming, messy setup output: {"aspect": "cords", "label": "negative", "sentence": "cords coming out the right for power plus cords coming out front for headphones plus network connection on left make for a very messy setup with cords going every direction."}	
<b>input 3:</b> aspect: usb connect, domain: laptop, keywords: usb connect, can, use, compatible output: {"aspect": "usb connect", "label": "neutral", "sentence": "I had a usb connect but, I can't use it because it is not compatible."}	
<b>input 4:</b> aspect: microsoft software, domain: laptop, keywords: much simpler, microsoft software, mac software, <i>friendly</i> output:	

图 6 样例分析

首先, 针对跨领域属性级别情感分析任务, 本文合理构造指令语句 (instruction) 用以引导 LLM 完成目标领域文本结构化生成任务, 并且在指示语句中明确指出情感标签类别与输出格式; 其次, 通过挖掘目标领域与源领域相似文本, 提取样例级别的文本生成关键词, 如 input4 中包含的关键词, friendly 来源于餐厅领域, 其余关键词来源于电脑领域; 接下来, 本文利用这些领域关联关键词, 通过上下文学习的方式引导模型高效生成目标领域有标签文本数据. 通过图 6 可以看出, 基于本文设计构造的指令, LLM 可以基于关键词信息准确地输出目标领域文本内容, 领域信息, 属性词信息和其情感信息, 同时具有较高的文本流畅性. 因此, 通过本文提出的基于 LLM 数据增强方法可以有效地缓解目标领域缺乏标签数据的问题, 从而帮助模型更好地解决跨领域属性级别情感分析任务.

## 4 总结

针对跨领域属性级情感分析中不同领域之间的语义分布差异以及新兴领域缺乏标签数据的难题,本文提出一种基于 LLM 与数据增强的跨领域属性级情感分析方法.该方法旨在克服目标领域数据匮乏和领域特异性的问题,通过合理构建针对跨领域属性级别情感分析任务的引导语句,挖掘目标领域与源领域相似文本,并通过上下文学习的方式使用领域关联关键词引导模型生成目标领域有标签文本数据,有效解决目标领域数据缺乏以及领域特异性问题.此外,本文在多个数据集中进行实验分析,实验结果表明本文方法在多个数据集上的实验中显著提升了基线模型的性能,并且能够轻松融合到目前表现优异的基线模型中并提高其性能.

本文下一步的研究方向主要聚焦于更加精细控制的文本生成方法,以减轻 LLM 可能产生的幻觉现象,从而提高生成文本的质量.本文计划利用这种改进的文本生成方法,完成更为复杂的情感分析任务,例如情感三元组或四元组的抽取任务.这将进一步推动情感分析领域的发展,使 LLM 能够更全面地理解和分析文本中的情感信息.

### References:

- [1] Liu B. Sentiment Analysis and Opinion Mining. Cham: Springer, 2012. 1–167. [doi: 10.1007/978-3-031-02145-9]
- [2] Hu MQ, Liu B. Mining and summarizing customer reviews. In: Proc. of the 10th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. Seattle: ACM, 2004. 168–177. [doi: 10.1145/1014052.1014073]
- [3] Liu HY, Chatterjee I, Zhou MC, Lu XS, Abusorrah A. Aspect-based sentiment analysis: A survey of deep learning methods. IEEE Trans. on Computational Social Systems, 2020, 7(6): 1358–1375. [doi: 10.1109/TCSS.2020.3033302]
- [4] Zhao YY, Qin B, Liu T. Sentiment analysis. Ruan Jian Xue Bao/Journal of Software, 2010, 21(8): 1834–1848 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/3832.htm> [doi: 10.3724/SP.J.1001.2010.03832]
- [5] Zhao CJ, Wang SG, Li DY. Research progress on cross-domain text sentiment classification. Ruan Jian Xue Bao/Journal of Software, 2020, 31(6): 1723–1746 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/6029.htm> [doi: 10.13328/j.cnki.jos.006029]
- [6] Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans. on Knowledge and Data Engineering, 2010, 22(10): 1345–1359. [doi: 10.1109/TKDE.2009.191]
- [7] Cai HJ, Xia R, Yu JF. Aspect-category-opinion-sentiment quadruple extraction with implicit aspects and opinions. In: Proc. of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th Int'l Joint Conf. on Natural Language Processing (Vol. 1: Long Papers). Association for Computational Linguistics, 2021. 340–350. [doi: 10.18653/v1/2021.acl-long.29]
- [8] Chen CH, Teng ZY, Wang ZQ, Zhang Y. Discrete opinion tree induction for aspect-based sentiment analysis. In: Proc. of the 60th Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers). Dublin: Association for Computational Linguistics, 2022. 2051–2064. [doi: 10.18653/v1/2022.acl-long.145]
- [9] Liu J, Teng ZY, Cui LY, Liu HM, Zhang Y. Solving aspect category sentiment analysis as a text generation task. In: Proc. of the 2021 Conf. on Empirical Methods in Natural Language Processing. Punta Cana: Association for Computational Linguistics, 2021. 4406–4416. [doi: 10.18653/v1/2021.emnlp-main.361]
- [10] Xia R, Zong CQ. A POS-based ensemble model for cross-domain sentiment classification. In: Proc. of the 5th Int'l Joint Conf. on Natural Language Processing. Chiang Mai: Asian Federation of Natural Language Processing, 2011. 614–622.
- [11] Zhou J, Tian JF, Wang R, Wu YB, Xiao WM, He L. SentiX: A sentiment-aware pre-trained model for cross-domain sentiment analysis. In: Proc. of the 28th Int'l Conf. on Computational Linguistics. Barcelona: Int'l Committee on Computational Linguistics, 2020. 568–579. [doi: 10.18653/v1/2020.coling-main.49]
- [12] Karouzou C, Paraskevopoulos G, Potamianos A. UDALM: Unsupervised domain adaptation through language modeling. In: Proc. of the 2021 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2021. 2579–2590. [doi: 10.18653/v1/2021.naacl-main.203]
- [13] Ben-David E, Rabinovitz C, Reichart R. PERL: Pivot-based domain adaptation for pre-trained deep contextualized embedding models. Trans. of the Association for Computational Linguistics, 2020, 8: 504–521. [doi: 10.1162/tacl\_a\_00328]
- [14] Wu H, Shi XD. Adversarial soft prompt tuning for cross-domain sentiment analysis. In: Proc. of the 60th Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers). Dublin: Association for Computational Linguistics, 2022. 2438–2447. [doi: 10.18653/v1/2022.acl-long.174]
- [15] Wang WY, Pan SJ. Recursive neural structural correspondence network for cross-domain aspect and opinion co-extraction. In: Proc. of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne: Association for Computational Linguistics, 2018. 2171–2181. [doi: 10.18653/v1/P18-1202]

- [16] Gong C, Yu J, Xia R. Unified feature and instance based domain adaptation for aspect-based sentiment analysis. In: Proc. of the 2020 Conf. on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2020. 7035–7045. [doi: [10.18653/v1/2020.emnlp-main.572](https://doi.org/10.18653/v1/2020.emnlp-main.572)]
- [17] Yu JF, Zhao QK, Xia R. Cross-domain data augmentation with domain-adaptive language modeling for aspect-based sentiment analysis. In: Proc. of the 61st Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers). Toronto: Association for Computational Linguistics, 2023. 1456–1470. [doi: [10.18653/v1/2023.acl-long.81](https://doi.org/10.18653/v1/2023.acl-long.81)]
- [18] Yang LY, Yuan LF, Cui LY, Gao WY, Zhang Y. FactMix: Using a few labeled in-domain examples to generalize to cross-domain named entity recognition. In: Proc. of the 29th Int'l Conf. on Computational Linguistics. Gyeongju: Int'l Committee on Computational Linguistics, 2022. 5360–5371.
- [19] Yu JF, Gong CG, Xia R. Cross-domain review generation for aspect-based sentiment analysis. In: Proc. of the 2021 Association for Computational Linguistics: ACL-IJCNLP 2021. Association for Computational Linguistics, 2021. 4767–4777. [doi: [10.18653/v1/2021.findings-acl.421](https://doi.org/10.18653/v1/2021.findings-acl.421)]
- [20] Ouyang L, Wu J, Jiang X, Almeida D, Wainwright CL, Mishkin P, Zhang C, Agarwal S, Slama K, Ray A, Schulman J, Hilton J, Kelton F, Miller L, Simens M, Askell A, Welinder P, Christiano PF, Leike J, Lowe R. Training language models to follow instructions with human feedback. In: Proc. of the 36th Annual Conf. on Neural Information Processing Systems. New Orleans, 2022.
- [21] Bai YT, Jones A, Ndousse K, *et al.* Training a helpful and harmless assistant with reinforcement learning from human feedback. arXiv:2204.05862, 2022.
- [22] Tang DY, Qin B, Feng XC, Liu T. Effective LSTMs for target-dependent sentiment classification. In: Proc. of the 26th Int'l Conf. on Computational Linguistics: Technical Papers. Osaka: The COLING 2016 Organizing Committee, 2016. 3298–3307.
- [23] Zhang B, Xu D, Zhang H, Li MZ. STCS lexicon: Spectral-clustering-based topic-specific Chinese sentiment lexicon construction for social networks. IEEE Trans. on Computational Social Systems, 2019, 6(6): 1180–1189. [doi: [10.1109/TCSS.2019.2941344](https://doi.org/10.1109/TCSS.2019.2941344)]
- [24] Bao XY, Jiang XT, Wang ZQ, Zhou GD. Cross-lingual aspect-level sentiment classification with graph neural network. Ruan Jian Xue Bao/Journal of Software, 2023, 34(2): 676–689 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/6667.htm> [doi: [10.13328/j.cnki.jos.006667](https://doi.org/10.13328/j.cnki.jos.006667)]
- [25] Feng C, Li HH, Zhao HY, Xue Y, Tang JY. Aspect-level sentiment analysis based on hierarchical attention and gate networks. Journal of Chinese Information Processing, 2021, 35(10): 128–136 (in Chinese with English abstract). [doi: [10.3969/j.issn.1003-0077.2021.10.015](https://doi.org/10.3969/j.issn.1003-0077.2021.10.015)]
- [26] Yan JF, Shao XH. Aspect-level sentiment analysis based on graph convolutional network. Journal of Chinese Information Processing, 2022, 36(10): 135–144 (in Chinese with English abstract). [doi: [10.3969/j.issn.1003-0077.2022.10.015](https://doi.org/10.3969/j.issn.1003-0077.2022.10.015)]
- [27] Bao LX, Lambert P, Badia T. Attention and lexicon regularized LSTM for aspect-based sentiment analysis. In: Proc. of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop. Florence: Association for Computational Linguistics, 2019. 253–259. [doi: [10.18653/v1/P19-2035](https://doi.org/10.18653/v1/P19-2035)]
- [28] Ma DH, Li SJ, Zhang XD, Wang HF. Interactive attention networks for aspect-level sentiment classification. In: Proc. of the 26th Int'l Joint Conf. on Artificial Intelligence. Melbourne: AAAI Press, 2017. 4068–4074.
- [29] Chen P, Sun ZQ, Bing LD, Yang W. Recurrent attention network on memory for aspect sentiment analysis. In: Proc. of the 2017 Conf. on Empirical Methods in Natural Language Processing. Copenhagen: Association for Computational Linguistics, 2017. 452–461. [doi: [10.18653/v1/D17-1047](https://doi.org/10.18653/v1/D17-1047)]
- [30] Li X, Bing LD, Zhang WX, Lam W. Exploiting BERT for end-to-end aspect-based sentiment analysis. In: Proc. of the 5th Workshop on Noisy User-generated Text. Hong Kong: Association for Computational Linguistics, 2019. 34–41. [doi: [10.18653/v1/D19-5505](https://doi.org/10.18653/v1/D19-5505)]
- [31] Huang BX, Carley KM. Syntax-aware aspect level sentiment classification with graph attention networks. In: Proc. of the 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int'l Joint Conf. on Natural Language Processing. Hong Kong: Association for Computational Linguistics, 2019. 5469–5477. [doi: [10.18653/v1/D19-1549](https://doi.org/10.18653/v1/D19-1549)]
- [32] Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M, Zhou YQ, Li W, Liu PJ. Exploring the limits of transfer learning with a unified text-to-text Transformer. The Journal of Machine Learning Research, 2020, 21(1): 140.
- [33] Lewis M, Liu YH, Goyal N, Ghazvininejad M, Mohamed A, Levy O, Stoyanov V, Zettlemoyer L. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In: Proc. of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2020. 7871–7880. [doi: [10.18653/v1/2020.acl-main.703](https://doi.org/10.18653/v1/2020.acl-main.703)]
- [34] Zhang WX, Li X, Deng Y, Bing LD, Lam W. Towards generative aspect-based sentiment analysis. In: Proc. of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th Int'l Joint Conf. on Natural Language Processing. Association for Computational Linguistics, 2021. 504–510. [doi: [10.18653/v1/2021.acl-short.64](https://doi.org/10.18653/v1/2021.acl-short.64)]

- [35] Yan H, Dai JQ, Ji T, Qiu XP, Zhang Z. A unified generative framework for aspect-based sentiment analysis. In: Proc. of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th Int'l Joint Conf. on Natural Language Processing. Association for Computational Linguistics, 2021. 2416–2429. [doi: [10.18653/v1/2021.acl-long.188](https://doi.org/10.18653/v1/2021.acl-long.188)]
- [36] Lu YJ, Liu Q, Dai D, Xiao XY, Lin HY, Han XP, Sun L, Wu H. Unified structure generation for universal information extraction. In: Proc. of the 60th Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers). Dublin: Association for Computational Linguistics, 2022. 5755–5772. [doi: [10.18653/v1/2022.acl-long.395](https://doi.org/10.18653/v1/2022.acl-long.395)]
- [37] Fei H, Wu SQ, Li JY, Li BB, Li F, Qin LB, Zhang MS, Zhang M, Chua TS. LasUIE: Unifying information extraction with latent adaptive structure-aware generative language model. In: Proc. of the 36th Annual Conf. on Neural Information Processing Systems. New Orleans, 2022.
- [38] Zhao GY, Lv CG, Fu GH, Liu ZL, Liang CF, Liu T. Domain specific sentiment words based attention model for cross-domain attribute-oriented sentiment analysis. Journal of Chinese Information Processing, 2021, 35(6): 93–102 (in Chinese with English abstract). [doi: [10.3969/j.issn.1003-0077.2021.06.010](https://doi.org/10.3969/j.issn.1003-0077.2021.06.010)]
- [39] Blitzer J, Dredze M, Pereira F. Biographies, Bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In: Proc. of the 45th Annual Meeting of the Association of Computational Linguistics. Prague: Association for Computational Linguistics, 2007. 440–447.
- [40] Ziser Y, Reichart R. Task refinement learning for improved accuracy and stability of unsupervised domain adaptation. In: Proc. of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: Association for Computational Linguistics, 2019. 5895–5906. [doi: [10.18653/v1/P19-1591](https://doi.org/10.18653/v1/P19-1591)]
- [41] Du CN, Sun HF, Wang JY, Qi Q, Liao JX. Adversarial and domain-aware BERT for cross-domain sentiment analysis. In: Proc. of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2020. 4019–4028. [doi: [10.18653/v1/2020.acl-main.370](https://doi.org/10.18653/v1/2020.acl-main.370)]
- [42] Xue QM, Zhang W, Zha HY. Improving domain-adapted sentiment classification by deep adversarial mutual learning. In: Proc. of the 34th AAAI Conf. on Artificial Intelligence, the 32nd Innovative Applications of Artificial Intelligence Conf., and the 10th AAAI Symp. on Educational Advances in Artificial Intelligence. New York: AAAI Press, 2020. 9362–9369. [doi: [10.1609/aaai.v34i05.6477](https://doi.org/10.1609/aaai.v34i05.6477)]
- [43] Li BH, Hou YT, Che WX. Data augmentation approaches in natural language processing: A survey. AI Open, 2022, 3: 71–90. [doi: [10.1016/j.aiopen.2022.03.001](https://doi.org/10.1016/j.aiopen.2022.03.001)]
- [44] Li JJ, Yu JF, Xia R. Generative cross-domain data augmentation for aspect and opinion co-extraction. In: Proc. of the 2022 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Seattle: Association for Computational Linguistics, 2022. 4219–4229. [doi: [10.18653/v1/2022.naacl-main.312](https://doi.org/10.18653/v1/2022.naacl-main.312)]
- [45] Zhao WX, Zhou K, Li JY, Tang TY, Wang XL, Hou YP, Min YQ, Zhang BC, Zhang JJ, Dong ZC, Du YF, Yang C, Chen YS, Chen ZP, Jiang JH, Ren RY, Li YF, Tang XY, Liu ZK, Liu PY, Nie JY, Wen JR. A survey of large language models. arXiv:2303.18223, 2023.
- [46] Brown TB, Mann B, Ryder N, *et al.* Language models are few-shot learners. In: Proc. of the 34th Int'l Conf. on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2020. 159.
- [47] Touvron H, Lavril T, Izacard G, Martinet X, Lachaux MA, Lacroix T, Rozière B, Goyal N, Hambro E, Azhar F, Rodriguez A, Joulin A, Grave E, Lample G. LLaMa: Open and efficient foundation language models. arXiv:2302.13971, 2023.
- [48] Jiao WX, Wang WX, Huang JT, Wang X, Shi SM, Tu ZP. Is ChatGPT a good translator? Yes with GPT-4 as the engine. arXiv:2301.08745, 2023.
- [49] Wang ZZ, Xie QM, Ding ZX, Feng Y, Xia R. Is ChatGPT a good sentiment analyzer? A preliminary study. arXiv: 2304.04339, 2023.
- [50] Gao TY, Yao XC, Chen DQ. SimCSE: Simple contrastive learning of sentence embeddings. In: Proc. of the 2021 Conf. on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2021. 6894–6910. [doi: [10.18653/v1/2021.emnlp-main.552](https://doi.org/10.18653/v1/2021.emnlp-main.552)]
- [51] Rose S, Engel D, Cramer N, Cowley W. Automatic keyword extraction from individual documents. In: Berry MW, Kogan J, eds. Text Mining: Applications and Theory. Chichester: Wiley, 2010. 1–20. [doi: [10.1002/9780470689646.ch1](https://doi.org/10.1002/9780470689646.ch1)]
- [52] Zhang ZS, Zhang A, Li M, Smola A. Automatic chain of thought prompting in large language models. In: Proc. of the 11th Int'l Conf. on Learning Representations. Kigali: OpenReview.net, 2023.
- [53] Pontiki M, Galanis D, Pavlopoulos J, Papageorgiou H, Androutsopoulos I, Manandhar S. SemEval-2014 task 4: Aspect based sentiment analysis. In: Proc. of the 8th Int'l Workshop on Semantic Evaluation. Dublin: Association for Computational Linguistics, 2014. 27–35. [doi: [10.3115/v1/S14-2004](https://doi.org/10.3115/v1/S14-2004)]
- [54] Pontiki M, Galanis D, Papageorgiou H, Manandhar S, Androutsopoulos I. SemEval-2015 task 12: Aspect based sentiment analysis. In: Proc. of the 9th Int'l Workshop on Semantic Evaluation (SemEval 2015). Denver: Association for Computational Linguistics, 2015.

- 486–495. [doi: 10.18653/v1/S15-2082]
- [55] Pontiki M, Galanis D, Papageorgiou H, Androutsopoulos I, Manandhar S, Al-Smadi M, Al-Ayyoub M, Zhao YY, Qin B, De Clercq O, Hoste V, Apidianaki M, Tannier X, Loukachevitch N, Kotelnikov E, Bel N, Jiménez-Zafra SM, Eryiğit G. SemEval-2016 task 5: Aspect based sentiment analysis. In: Proc. of the 10th Int'l Workshop on Semantic Evaluation. San Diego: Association for Computational Linguistics, 2016. 19–30. [doi: 10.18653/v1/S16-1002]
- [56] Toprak C, Jakob N, Gurevych I. Sentence and expression level annotation of opinions in user-generated discourse. In: Proc. of the 48th Annual Meeting of the Association for Computational Linguistics. Uppsala: Association for Computational Linguistics, 2010. 575–584.
- [57] Kingma DP, Ba J. Adam: A method for stochastic optimization. In: Proc. of the 3rd Int'l Conf. on Learning Representations. San Diego, 2015.
- [58] Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional Transformers for language understanding. In: Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1 (Long and Short Papers). Minneapolis: Association for Computational Linguistics, 2019. 4171–4186. [doi: 10.18653/v1/N19-1423]
- [59] Ben-David E, Oved N, Reichart R. PADA: Example-based prompt learning for on-the-fly adaptation to unseen domains. Trans. of the Association for Computational Linguistics, 2022, 10: 414–433. [doi: 10.1162/tacl\_a\_00468]
- [60] Wei J, Zou K. EDA: Easy data augmentation techniques for boosting performance on text classification tasks. In: Proc. of the 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int'l Joint Conf. on Natural Language Processing (EMNLP-IJCNLP). Hong Kong, China: Association for Computational Linguistics, 2019. 6382–6388. [doi: 10.18653/v1/D19-1670]
- [61] Ioffe S. Improved consistent sampling, weighted Minhash and L1 sketching. In: Proc. of the 10th IEEE Int'l Conf. on Data Mining. Sydney: IEEE, 2010. 246–255. [doi: 10.1109/ICDM.2010.80]

#### 附中文参考文献:

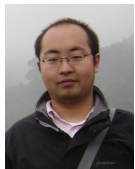
- [4] 赵妍妍, 秦兵, 刘挺. 文本情感分析. 软件学报, 2010, 21(8): 1834–1848. <http://www.jos.org.cn/1000-9825/3832.htm> [doi: 10.3724/SP.J.1001.2010.03832]
- [5] 赵传君, 王素格, 李德玉. 跨领域文本情感分类研究进展. 软件学报, 2020, 31(6): 1723–1746. <http://www.jos.org.cn/1000-9825/6029.htm> [doi: 10.13328/j.cnki.jos.006029]
- [24] 鲍小异, 姜晓彤, 王中卿, 周国栋. 基于跨语言图神经网络模型的属性级情感分类. 软件学报, 2023, 34(2): 676–689. <http://www.jos.org.cn/1000-9825/6667.htm> [doi: 10.13328/j.cnki.jos.006667]
- [25] 冯超, 黎海辉, 赵洪雅, 薛云, 唐婧尧. 基于层次注意力机制和门机制的属性级别情感分析. 中文信息学报, 2021, 35(10): 128–136. [doi: 10.3969/j.issn.1003-0077.2021.10.015]
- [26] 闫金凤, 邵新慧. 基于图卷积网络的特定方面情感分析. 中文信息学报, 2022, 36(10): 135–144. [doi: 10.3969/j.issn.1003-0077.2022.10.015]
- [38] 赵光耀, 吕成国, 付国宏, 刘宗林, 梁春丰, 刘涛. 基于领域特有情感词注意力模型的跨领域属性情感分析. 中文信息学报, 2021, 35(6): 93–102. [doi: 10.3969/j.issn.1003-0077.2021.06.010]



李诗晨(1998—), 男, 硕士生, 主要研究领域为自然语言处理, 情感分析.



周国栋(1967—), 男, 博士, 教授, 博士生导师, CCF 杰出会员, 主要研究领域为自然语言处理.



王中卿(1987—), 男, 博士, 副教授, CCF 专业会员, 主要研究领域为自然语言处理, 情感分析.