

基于大型语言模型文本简化的细粒度情感分析

王叶, 王中卿

引用本文

王叶, 王中卿. 基于大型语言模型文本简化的细粒度情感分析[J]. 计算机科学, 2025, 52(10): 258-265.

WANG Ye, WANG Zhongqing. [Text Simplification for Aspect-based Sentiment Analysis Based on Large Language Model](#) [J]. Computer Science, 2025, 52(10): 258-265.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于文本生成的多粒度评论情感分析](#)

Multi-grained Sentiment Analysis of Comments Based on Text Generation

计算机科学, 2025, 52(10): 239-246. <https://doi.org/10.11896/jsjcx.240800025>

[基于提示学习与超图的事件因果关系识别模型](#)

Event Causality Identification Model Based on Prompt Learning and Hypergraph

计算机科学, 2025, 52(9): 303-312. <https://doi.org/10.11896/jsjcx.240800121>

[结合预训练模型和数据增强的跨领域属性级情感分析研究](#)

Cross-domain Aspect-based Sentiment Analysis Based on Pre-training Model with Data Augmentation

计算机科学, 2025, 52(8): 300-307. <https://doi.org/10.11896/jsjcx.240900114>

[解耦知识蒸馏在文档级关系抽取中的应用](#)

Application of Decoupled Knowledge Distillation Method in Document-level Relation Extraction

计算机科学, 2025, 52(8): 277-287. <https://doi.org/10.11896/jsjcx.240600050>

[结合评价对象信息的评论摘要研究](#)

Study on Opinion Summarization Incorporating Evaluation Object Information

计算机科学, 2025, 52(7): 233-240. <https://doi.org/10.11896/jsjcx.240600144>

基于大型语言模型文本简化的细粒度情感分析

王叶 王中卿

苏州大学计算机科学与技术学院 江苏 苏州 215006

(xiaoluhuishuohua@163.com)

摘要 细粒度情感分析旨在识别句子中每个方面的情感极性。然而,现有研究大多忽视了评论文本中普遍存在的冗余信息,这些无关信息不仅增加了模型处理的复杂性,还可能导致模型无法准确捕捉原始文本中的情感元素。为解决这一问题,提出了一种将原始文本转化为简化子句的模型,以更简明的方式表达相同的情感观点。其基本思想是利用大型语言模型预识别文本中的方面词和意见词,再基于识别结果生成简化子句,并通过自我验证机制确保生成的子句满足情感一致性、相关性和简洁性。此外,所提模型结合原始文本和简化子句共同生成情感元素。在公开数据集 Restaurant 和 Laptop 以及 Phone 上,所提模型的表现均优于现有基线模型,证明简化子句在细粒度情感分析中具有重要的作用。

关键词 细粒度情感分析;文本简化;大型语言模型;自我验证;自然语言处理

中图分类号 TP391

Text Simplification for Aspect-based Sentiment Analysis Based on Large Language Model

WANG Ye and WANG Zhongqing

School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu 215006, China

Abstract Aspect-based sentiment analysis aims to identify the sentiment polarity of each aspect in a sentence. However, most existing approaches overlook the redundant and irrelevant information often present in review texts, which not only complicates model processing, but also hinders accurate sentiment element extraction. To address this issue, this paper proposes a model that transforms the original text into simplified clauses, expressing the same sentiment in a more concise manner. The key idea is to leverage a large language model to pre-identify aspect and opinion terms in the text, and then generate simplified clauses based on these identified sentiment elements. A self-verification mechanism is employed to ensure the generated clause satisfy three criteria: sentiment consistency, relevance, and conciseness. Furthermore, the model jointly uses both the original text and the simplified clauses to generate sentiment elements. Experimental results on public datasets—Restaurant, Laptop, and Phone, demonstrate that the model outperforms existing baselines, highlighting the significance of simplified clauses in aspect-based sentiment analysis.

Keywords Aspect-based sentiment analysis, Text simplification, Large language model, Self-validation, Natural language processing

1 引言

细粒度情感分析 (Aspect-based Sentiment Analysis, AB-SA) 近年来受到了越来越多的关注,该任务旨在识别句子中的方面词 (Aspect Term)、观点词 (Opinion Term) 以及对应的方面类别 (Aspect Category) 和情感极性 (Sentiment Polarity)。例如,句子“This pizza is delicious(这披萨真好吃)”的方面词和观点词分别是“pizza(披萨)”和“delicious(好吃)”,对应的方面类别和情感极性分别是“food(食物)”和“positive(积极)”。

早期 ABSA 研究主要集中于单一情感要素的识别,如方

面词的挖掘^[1-2]和方面类别的辨识^[3],这些研究通常专注于对具体情感元素的提取,旨在提高系统对单一方面词或类别的识别精度。随着研究的深入,后期的研究则更多地关注于同时预测多个情感元素^[4-5]。一般而言,大多数 ABSA 任务被构建为序列级或令牌级的分类问题。然而,这类方法存在一个显著的局限性,即误差传播效应。具体来说,整体预测性能高度依赖于每个独立步骤的精确度,这意味着任何一步中的微小偏差都可能在后续流程中被放大,最终严重影响整体结果的准确性^[6]。因此,近年来,生成方法被引入以统一处理各种 ABSA 任务,并表现出优异的性能。常见的方法是生成特定格式的情感元素序列,从而利用标签语义如类索引^[7]、情感

到稿日期:2025-01-17 返修日期:2025-05-20

基金项目:国家自然科学基金(62076175,61976146)

This work was supported by the National Natural Science Foundation of China(62076175,61976146).

通信作者:王中卿(wangzq@suda.edu.cn)

元素序列^[8]、自然语言^[9-10]、结构化提取模式^[11]或意见树^[12]作为生成模型的目标。

生成方法虽然在情感分类任务中有了明显的性能提升,但是仍然有一些不足之处。

1)评论文本中常常包含大量无关或噪声信息,这不仅增加了模型的处理复杂性,还可能导致情感识别不准确。这些冗余信息干扰了模型对关键情感要素的捕获,容易导致模型无法有效识别真实的情感特征。

2)许多评论文本中的方面词和观点词往往以隐含方式存在,这为情感元素的提取带来了显著挑战。这类隐含的情感表达需要模型具备更强的上下文理解能力,以准确识别和解释文本中的潜在情感信息。

针对上述问题,本文提出将原文简化为子句的方法。简化子句通过去除复杂的句式结构、冗长的修饰成分以及无关的背景信息,显著降低了文本复杂度,减少了噪声干扰,使情感表达更加直接清晰。同时,简化子句还能有效揭示原文中隐含的情感元素,帮助模型更准确地识别和捕捉潜在情感信息。如图1所示,合格的简化子句保留了原文中的方面词和观点词,并以更简洁的形式表达相同的情感观点,从而便于提取情感元素。

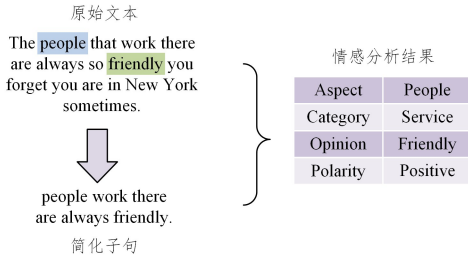


图1 文本简化举例

Fig. 1 Example of text simplification

本文提出的模型分为3个关键步骤:首先,利用大型语言模型(Large Language Model, LLM)对原始文本进行分析,识别文本中的方面词和观点词;其次,基于识别出的情感要素,生成包含这些关键信息的简化子句;最后,引入自我验证机

制,从情感一致性、相关性和简洁性3个维度对生成的简化子句进行评价和优化,确保其质量。在此基础上,模型结合原始文本与生成的简化子句,共同生成完整的情感元素四元组。

本文的贡献总结如下:

1)提出了一种新颖的文本简化方法,通过将原始评论文本转化为简化子句来解决冗余信息和隐含表达的问题;

2)设计了一个基于大型语言模型的文本简化框架,首先识别原文中的情感要素,继而生成简化子句,并引入自我验证机制确保生成子句满足情感一致性、相关性和简洁性;

3)在3个公开数据集即 Laptop, Restaurant 以及 Phone 上进行实验,结果证明了本文模型的先进性和可解释性。

2 相关工作

2.1 细粒度情感分析

细粒度情感分析的研究经历了从单一子任务到复杂任务组合的发展历程。早期研究主要集中在单个情感元素的预测上。例如,Liu等^[1]采用递归神经网络和词嵌入进行方面项提取,Tang等^[13]利用长短时记忆网络预测给定方面项的情感极性,Bu等^[14]则专注于中文评论中的方面类别检测。随着研究的深入,学者们开始考虑情感元素之间的内在关联性,逐渐探索多个情感元素的联合提取。Chen等^[15]研究了方面词和观点词的成对提取;Zhang等^[16]致力于方面词及其对应情感极性的联合预测;Peng等^[4]提出了方面情感三元组提取任务,其能够同时抽取文本中的方面词、观点词和情感极性。近年来,预训练的编码器-解码器语言模型在 ABSA 研究中取得了显著进展,研究者尝试将类索引^[7]、自然语言^[9]或所需的情感元素序列^[8]作为生成模型的目标。其中,Bao等^[12]深入探讨了情感元素间相关性的重要性,提出了一种观点树生成模型,通过树状结构联合检测所有情感元素;Gou等^[17]则通过整合不同序列生成的情感元素,实现了多角度的情感分析;Yu等提出了基于集合预测的方法,以提高三元组提取的准确性^[18]。此外,基于方面的少样本情感分析也逐渐成为研究热点^[19-20]。

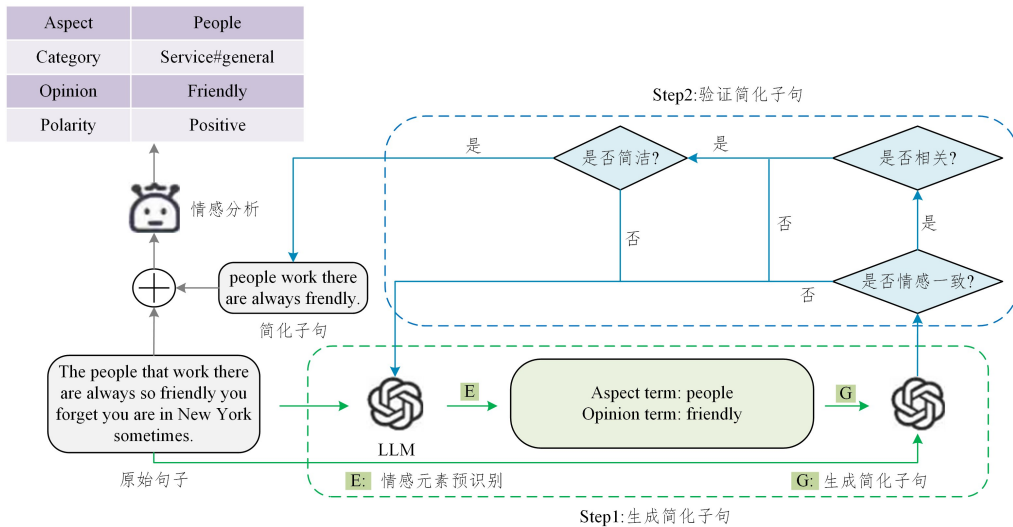


图2 模型整体框架

Fig. 2 Overall architecture of the model

2.2 文本简化

文本简化是自然语言处理中的一个重要研究方向,其目标是将复杂文本转换为更简单的版本,同时保留原文的核心语义,以提高文本的可读性和可理解性^[21]。现有的文本简化研究主要可以分为词汇简化和句法简化两大类。

词汇简化主要关注词汇层面的转换,通过将文本中的复杂词汇替换为更容易理解的同义词来提高文本的可读性^[22-23]。这种方法不仅能够降低文本的词汇复杂度,减少读者的理解障碍,而且有助于提升信息传递的效率和准确性。句法简化则着重于降低语法结构的复杂性,例如将复杂的从句结构分解为简单句,或将被动语态转换为主动语态^[24]。文本简化技术经历了从基于规则的早期方法向基于深度神经网络的高级技术的显著发展。早期方法依赖预定义的规则和词典,通过词汇替换或句法转换实现文本简化,但在应对复杂语言结构时性能有限^[25]。近年来,深度学习模型显著提升了文本简化的能力,生成式预训练模型通过序列到序列的架构实现了上下文感知的文本生成,进一步提高了简化文本的流畅性和可读性^[26-28]。相比之下,基于深度学习的方法能够更好地适应不同领域和文本类型的简化需求。最新研究还结合信息损失控制与多任务学习,优化了简化过程中的信息保留和文本适应性,从而更好地平衡文本简化与信息完整性之间的关系^[29-31]。

然而,在情感分析领域,特别是细粒度情感分析中,文本简化技术的应用研究仍然相对有限。受到 Jiang 等研究^[32]的启发,本文深入探讨了将文本简化技术应用于细粒度情感分析的可行性。具体而言,本文利用大型语言模型生成简化子句,以在保持原文情感信息的同时,通过更简洁的表达提升模型的理解能力。此外,本文特别注意确保简化过程中的方面词与观点词之间的语义关联,避免在简化过程中丢失关键信息。

3 模型架构

本章给出本文模型的整体框架,如图 2 所示。模型包括 3 个部分:1)简化子句生成,即利用大型语言模型识别原始文本中的方面词和观点词,并基于识别结果生成简化子句;2)简化子句验证,即设计自我验证机制,利用大型语言模型确保生成的简化子句同时满足情感一致性、相关性和简洁性 3 项关键标准;3)情感元素生成,即结合原始文本和验证后的简化子句共同生成最终的情感元素,以充分发挥两种文本形式的互补优势。

3.1 简化子句生成

简化子句生成的核心目标是确保方面词和意见词与原始文本一致,同时完整保留其情感色彩和语义内涵。为了实现这一目标,首先通过提示工程向大型语言模型提供方面词和观点词的明确定义,从而帮助模型更准确地识别目标词语;然后采用 five-shot 范式进行情感元素预识别,即通过提供 5 个典型样例来引导模型从原始文本 T 中准确识别方面词集合 $A = \{a_1, a_2, \dots, a_M\}$ 和意见词集合 $O = \{o_1, o_2, \dots, o_L\}$,其中 M 和 L 分别表示原始文本 T 中方面词和意见词的数量。这种基于少量样例的方法不仅提高了模型的识别准确率,还增强了其对不同文本风格和上下文环境的适应能力,尤其在处理

多样化和复杂场景时表现出更好的鲁棒性。

为进一步提炼并重构原始文本 T ,在确保保留关键信息的基础上使其易于理解,本文基于识别结果和原始文本 T ,利用大型语言模型生成简化子句 S^* ,设计的提示模板如图 3 所示。

"Given the sentence T and its aspect terms A and opinion terms O , please extract the relevant parts that contain these aspects and opinions from the sentence." (给定句子 T 及其方面词 A 和观点词 O ,请从句子中提取包含这些方面词和观点词的相关部分。)

图 3 生成简化子句提示模板

Fig. 3 Prompt template for generating simplified clauses

通过这一提示,模型能够提取包含方面词和意见词的相关语句片段,并生成更简洁的子句 S^* ,如式(1)所示:

$$S^* = \operatorname{argmax} P(s | T, A, O), s \in S \quad (1)$$

其中, S 为所有可能的简化子句集合, $P(s | T, A, O)$ 表示在给定原始文本 T 及其情感要素 A 和 O 的条件下生成简化子句 s 的概率。

3.2 简化子句验证

提高大型语言模型能力的一种普遍方法是通过高质量的验证反馈进行学习和纠正,其在模型优化的各个方面都起着至关重要的作用。这个过程被称为自我验证(Self-Verification)。此方法旨在提升模型生成准确且与上下文相关的响应的能力,同时最大限度地减少错误并提高整体性能。近年来,自我验证技术在大型语言模型优化中展现出显著优势。Liu 等^[33]的研究表明,成功的自我纠正主要依赖于两个关键因素:将采样温度设置为零,使用中立提示。温度参数为零能确保模型输出的确定性,避免随机性带来的不稳定性;而中立提示则有助于模型保持客观性,不带有预设立场地评估生成内容。

尽管自我验证具有较大的自动化优势,但是其计算开销也是不容忽视的问题。为此,本文模型采用自我验证与基于规则的人工设计验证相结合的双重验证机制。具体而言,模型首先通过自我验证对生成结果进行初步评估,随后使用预先定义的评估公式进行定量验证。这种基于规则的验证方法通过严格的数学指标对简化子句的质量进行量化评估,提供了可解释和可复现的评估标准。该双重验证机制可以确保生成的简化子句严格遵循 3 个特定标准:情感一致性、相关性和简洁性。

3.2.1 情感一致性

情感一致性是评估简化子句质量的首要标准,也是确保生成结果能准确传达原始文本语义的核心指标。在简化过程中,简化子句必须完整保留原始文本中的方面词和观点词,同时准确传达原文的情感倾向。为实现这一目标,本文采用大型语言模型进行自我验证,在验证过程中将温度参数设置为零,并设计了针对性的中立提示模板对情感要素进行评估。如图 4 所示。

"Given the sentence X , please extract the aspect and opinion term(s) and determine the sentiment polarity." (给定句子 X ,请提取方面词和观点词并确定它们的情绪极性。)

图 4 评估情感一致性提示模板

Fig. 4 Prompt template for evaluating sentiment consistency

该提示模板分别应用于原始文本 T 和简化子句 S^* , 验证过程要求两次的响应结果在情感要素识别和情感极性判断上保持一致。这种基于自我验证的方法能够有效捕捉文本中的细微情感差异, 确保简化后的文本在情感传递上与原始文本保持一致, 为后续的规则验证提供可靠基础。

3.2.2 相关性

为了进一步评估简化子句与原始文本之间的相关性, 采用多维相似度计算方法, 将词汇相似度和语义相似度整合为一个统一的综合得分。该方法不仅考虑了词汇层面的匹配, 还充分融入了语义层面的深层次关系。每个相似度指标都根据其在整体评估中的重要性分配权重, 从而构建更加精确且符合实际需求的衡量标准, 确保在不同层面上都能够真实反映简化子句与原始文本的相关性。

词汇相似度用来评估原始文本与简化子句之间的词汇重叠度, 本文使用 Jaccard 相似度作为计算指标。Jaccard 相似度可以忽略词频的影响, 重点关注词汇存在与否, 从而有效消除冗余词汇的干扰, 能够精确地评估简化过程中词汇的保留情况。具体而言, V_T 和 V_{S^*} 分别表示原始文本和简化子句的词汇集合, 计算方法如式(2)所示:

$$le = \frac{|V_T \cap V_{S^*}|}{|V_T \cup V_{S^*}|} \quad (2)$$

语义相似度用于评估原始文本与简化子句之间的语义接近程度。本文使用预训练语言模型分别生成原始文本与简化子句的上下文嵌入表示 E_T 和 E_{S^*} , 这两个嵌入向量能够有效地捕捉句子中词语的语义及其上下文关系, 并通过计算嵌入向量间的余弦相似度来量化原始文本与简化子句在语义空间中的接近程度。计算方法如式(3)所示:

$$se = \frac{E_T \cdot E_{S^*}}{\|E_T\| \cdot \|E_{S^*}\|} \quad (3)$$

最终的综合相关性得分 Score 的计算如式(4)所示:

$$Score = \alpha \cdot le + \lambda \cdot se \quad (4)$$

其中, α 和 λ 分别表示词汇相似度和语义相似度的权重。本文根据大量实验, 将 α 设为 0.4, λ 设为 0.6, 并且将综合相关性得分的阈值设为 0.7, 即综合相关性得分超过 0.7 的简化子句被认为满足相关性标准。

3.2.3 简洁性

简洁性用于评估简化子句的简单程度, 以确保其仅包含与关键情感成分相关的内容。为此, 采用 Flesch 可读性分数^[34]作为衡量指标。Flesch 可读性分数是一种广泛应用于文本分析的可读性测试方法, 其核心思想是通过量化单词的平均音节数和句子的平均单词数来评估文本的阅读和理解难度。具体计算如式(5)所示:

$$F = \kappa - \varphi \times \frac{N_w}{N_s} - \mu \times \frac{N_y}{N_w} \quad (5)$$

其中, κ , φ 和 μ 是常数, 其值分别为 206.835, 1.015 和 84.6; N_w , N_s 和 N_y 分别表示文本中的单词总数、句子总数和音节总数。Flesch 分数越高, 表明文本的可读性越强, 更容易被理解, 因此满足简洁性标准的简化子句应该比原始文本获得更高的 Flesch 分数。

3.3 情感元素生成

本文在 Paraphrase^[10]的基础上, 将 ABSA 任务转化为序

列到序列的生成任务, 将生成目标 $Q = (a, c, o, s)$ 转化为线性化的输出格式 Y , 同时将原始评论文本 T 和简化子句 S^* 组合为输入文本 C , 不仅保留了原始评论的关键信息, 还融入了简化后的语义特征, 从而增强了模型对输入的理解能力。组合方式如式(6)所示:

$$C = \langle \text{original text} \rangle T + \langle \text{simplified text} \rangle S^* \quad (6)$$

对于序列到序列的学习, 可以利用编码器-解码器架构, 例如 Transformer^[35]。假设模型 M 由编码器 Enc 和解码器 Dec 组成, 其中 Enc 的作用是将输入序列 C 映射到一个上下文向量 H , 而 Dec 则从 H 中生成目标输出序列 Y 。模型 M 的参数可以表示为 θ , M 的目标是估计给定输入句子的目标句子的条件概率, 即 $p(Y|C)$ 。以 $H = Enc(C)$ 为编码器对 C 的上下文隐藏表示, 条件概率可以重写为:

$$p(Y|C) = \prod_{i=1}^N p(y_i | Y_{<i}, H) \quad (7)$$

其中, y_i 是目标序列中第 i 个元素; $Y_{<i}$ 表示前 $i-1$ 个元素的序列。 $p(y_i | Y_{<i}, H)$ 是指预训练语言模型词汇表上的 softmax 归一化概率, 定义为:

$$p(y_i | Y_{<i}, H) = \begin{cases} p(y_1 | H), & \text{if } i=1 \\ p(y_i | y_1, \dots, y_{i-1}, H), & \text{if } i>1 \end{cases} \quad (8)$$

最后, 给定大小为 $|D|$ 的训练集 D , 可以通过最大化正确目标句子的对数似然来训练序列到序列模型 M , 如式(9)所示:

$$p(Y|C) = \prod_{i=1}^N p(y_i | Y_{<i}, H) \quad (9)$$

训练后的模型 M 以 C 为输入, 返回条件概率 $p(Y|C)$, 然后应用贪婪解码将条件概率解码为目标句子 Y , 最后还原得到最终结果 Q 。

4 实验与结果分析

4.1 实验数据集

为了全面验证本文模型的有效性与适用范围, 选择了具有不同特征的 3 个公开数据集进行实验测试: Restaurant, Laptop 和 Phone 数据集。其中, Restaurant 和 Laptop 数据集是由 Cai 等^[36]构建的领域特定数据集, 分别包含餐厅和笔记本电脑领域的英文评论文本; Phone 数据集则收集了来自多个电商平台的手机评论, 涵盖了 12 个主流品牌, 此数据集不仅语言复杂度更高, 而且情感表达更为密集, 相较于前两个数据集有着更高的识别要求^[37]。为确保实验的公平性和可重复性, 本文严格遵循原始数据集的划分方式来进行训练集、验证集和测试集的分配。表 1 列出了 3 个数据集的详细统计信息。其中, #s 表示句子总数, #w/s 表示平均句长(单词数/句子数), #q 表示情感四元组数量, #EA 和 #IA 分别表示显式方面词和隐式方面词的数量。

表 1 数据集的基本信息

Table 1 Basic information of datasets

Datasets	#s	#w/s	#q	#EA	#IA
Laptop	2286	15.11	3661	2781	880
Restaurant	4076	15.73	5773	4519	1254
Phone	7115	25.78	15884	15884	—

4.2 实验设置

采用 GPT-4o-mini 作为 3.1 和 3.2 节中描述的大型语言

模型, T5-Large 作为 3.3 节中描述的预训练语言模型 M 。在验证集上通过网格搜索来调整模型的参数, 模型训练采用学习率为 0.0003 的基于随机梯度下降算法的 Adam 优化器, batch 大小设置为 16, max length 设置为 256。每个数据集上的实验均在 Nvidia RTX 4090 GPU 上独立运行 20 轮, 以确保结果的稳定性与可靠性, 并提升模型的泛化能力和实验可重复性。模型参数设置如表 2 所列。

表 2 模型参数
Table 2 Model parameters

参数	取值
optimizer	Adam
loss	CrossEntropy
batch size	0.0003
max length	256
epochs	20

4.3 评估方法

本文采用严格的评估标准: 当且仅当预测四元组的 4 个元素及其组合与黄金标准的四元组完全一致时, 才判定为正确预测。计算准确率(P)、召回率(R)和 F1 分数(F1), 并使用 F1 分数作为最终评估指标。

$$P = \frac{TP}{TP + FP} \quad (10)$$

$$R = \frac{TP}{TP + FN} \quad (11)$$

$$F1 = \frac{2PR}{P + R} \quad (12)$$

其中, TP 是真阳的数目, FP 是假阳的数目, FN 是假阴的数目。

4.4 基线模型

为了全面研究模型的有效性, 将其与有关 ABSA 四元组提取的基准模型进行了比较。基线模型分为 3 组。

1) 基于管道的模型

(1) TAS-BERT^[5]: 将方面类别和情感极性附加到评论句

子中并作为 BERT 的输入, 在一个统一框架内集成了基于方面类别的情感分类与方面词提取任务。

(2) Extract-Classify^[38]: 先对方面词和意见词进行联合抽取, 再基于提取的方面词-意见词对预测对应的方面类别与情感极性。

2) 生成式模型

(1) GAS^[8]: 将 ABSA 任务重构为情感元素序列生成问题, 提出统一的生成框架来处理多个 ABSA 子任务。

(2) Paraphrase^[10]: 提出释义建模范式, 将 ABSA 任务转化为释义生成过程, 以实现情感元素的端到端联合抽取。

(3) GEN-SCL-NAT^[39]: 融合监督对比学习 (SCL) 与自然注释树 (NAT), 提出的新型目标文本构建方法。

(4) Opinion Tree^[12]: 基于树状结构设计, 实现评论句子中情感元素的层次化联合检测。通过构建树形结构捕捉情感元素之间的依赖关系, 提升了情感分析的准确性和细致度。

(5) ADA^[40]: 提出一种自适应数据增强框架, 通过自动生成多样化的训练样本, 有效缓解四元组和方面类别分布不均的问题, 从而提升模型在不平衡数据集上的泛化能力。

(6) TOG^[41]: 设计专门的过渡系统生成观点树, 并结合结构化任务的特点, 优化预训练语言模型在微调过程中的性能。

3) 大型语言模型

(1) ChatGPT-4o-mini^[42]: 基于 OpenAI 开发的大规模生成式预训练语言模型, 通过少样本学习方式来实现情感要素的识别和抽取。

(2) LLaMA-8B^[43]: Meta 提出的轻量级大规模语言模型, 采用纯解码器架构设计, 通过任务特定的指令微调实现四元组的高效抽取。

4.5 实验结果分析

本节在 3 个数据集上比较了所提模型和基准模型的 ABSA 任务性能。实验结果如表 3 所列, 其中加粗字体表示当前指标中的最优结果, 下划线表示次优结果。

表 3 与基线模型比较结果

Table 3 Comparison results with baseline models

Models	Restaurant			Laptop			Phone		
	P	R	F1	P	R	F1	P	R	F1
TAS-BERT	26.29	46.29	33.53	47.15	19.22	27.31	34.53	22.07	26.93
Extract-Classify	38.54	52.96	44.61	45.56	29.48	35.80	31.28	33.23	32.23
GAS	60.69	58.52	59.59	41.60	42.57	42.17	50.72	48.15	49.40
Paraphrase	58.98	59.11	59.04	41.77	<u>45.04</u>	43.34	50.62	53.20	51.87
GEN-SCL-NAT	—	—	62.62	—	—	45.16	45.16	51.56	48.15
Opinion Tree	63.96	61.74	62.83	<u>46.11</u>	44.79	<u>45.44</u>	53.02	56.96	54.74
ADA	60.15	61.95	61.04	45.03	44.53	44.78	—	—	—
TOG	64.32	<u>62.48</u>	<u>63.38</u>	45.32	44.57	44.94	<u>54.41</u>	<u>56.07</u>	<u>55.23</u>
ChatGPT-4o-mini	39.32	42.03	40.63	22.60	25.40	23.92	22.87	29.52	25.77
LLaMA-8B	62.55	61.41	61.97	42.51	43.67	43.08	52.40	52.59	52.49
Ours	<u>64.02</u>	63.67	63.84	46.86	45.56	46.20	55.17	55.87	55.52

本文模型在 3 个数据集上的 F1 指标均优于其他对比模型, 在 Laptop 数据集上的表现尤为突出。进一步分析发现, 与 Para-phrase 模型相比, 本文模型的 F1 指标平均提升了 3.77 个百分点, 最大提升了 4.8 个百分点。这一结果充分说明了结合原始文本与简化子句的方法能够有效去除评论中的

无关信息和噪声干扰, 在减轻文本复杂性的同时保留了情感相关的核心信息, 从而大幅提高了模型细粒度情感分析的能力。从表 3 中还可以看出, 相较于基于管道的模型, 生成式模型表现出更为优越的性能。这一现象反映了生成式方法在处理复杂情感元素时的优势, 尤其是能够直接生成包含所有情

感元素的四元组,避免了管道方法中多步骤处理导致的信息损失和错误积累。管道方法需要逐步完成情感元素的提取和分类,每一阶段的误差都会对最终结果造成影响,导致模型的累积误差,进而影响整体性能。相比之下,生成式方法通过端到端的学习框架,能够在统一的体系下直接生成包含所有情感元素的四元组,从根本上减少了中间步骤中的错误传播和信息丢失。值得注意的是,两个大型语言模型在3个数据集上的表现均不突出,这表明细粒度情感分析仍然是大语言模型面临的一项挑战,尤其是在复杂语境和多维度情感表达的情况下,模型难以准确捕捉细微的情感变化,导致其对情感细节的捕捉和理解能力有限。

本文模型优化了生成式方法的设计,通过结合原始文本和简化子句,利用统一生成体系结构将自然语言标签编码到目标输出中,有效缓解了管道方法的错误传播问题,还充分利用了标签的语义信息。这种设计提升了模型在情感元素提取任务中的泛化能力,尤其在面对隐含情感表达和复杂文本噪声时表现得更加稳定和高效。通过简化原始文本的复杂性,同时保留与情感相关的核心信息,模型能够更加精准地捕捉情感极性变化,显著提高了情感分析的准确性和可靠性,展现了更强的适应性和鲁棒性。

4.6 消融实验

为了验证本文方法中生成简化子句和验证简化子句的有效性,设计消融实验并对结果进行了分析。

Random Fragment:从原始文本中随机选取一个片段作为简化子句。

ChatGPT Extract:利用 GPT-4o-mini 模型直接从原始文本中提取一个片段作为简化子句。

ChatGPT Generate₁:去掉模型中的情感元素预识别模块,给定方面词和观点词的定义,利用 GPT-4o-mini,根据原始文本直接生成简化子句。

ChatGPT Generate₂:给定方面词和观点词的定义,利用 GPT-4o-mini,根据原始文本同时生成方面词、观点词和简化子句。

Ours-w/o-sv:去除模型的自我验证模块。

消融实验结果如表4所列。可以看出,Random Fragment和ChatGPT Extract分别通过随机选择片段和直接提取片段来生成简化子句,其F1分数分别为60.83%和61.10%,略高于原始文本,但提升较为有限。这表明,这些方法尽管能够生成简化子句,但未能有效捕捉情感元素的核心信息,导致情感分类性能较低。相比之下,ChatGPT Generate₁和ChatGPT Generate₂通过GPT模型生成简化子句,其中ChatGPT Generate₂在生成简化子句的同时考虑了方面词和观点词,表现较为优越,F1分数达到62.36%。这表明在生成简化子句的过程中,结合方面词和观点词的信息能够显著提升模型的性能,增强情感分析的准确性。然而,相比使用本文设计的生成简化子句提示模板的Ours-w/o-sv,这些方法在Restaurant数据集上均表现出不同程度的性能下降,这证明了基于预识别情感元素生成简化子句方法的有效性,其能够更好地保留与情感分析相关的核心信息。

表4 消融实验结果

Table 4 Ablation experiment results

Methods	P	R	F1
Original Text	60.94	60.33	60.64
Random Fragment	61.91	59.78	60.83
ChatGPT Extract	61.65	60.56	61.10
ChatGPT Generate ₁	61.92	60.89	61.40
ChatGPT Generate ₂	62.36	62.22	62.36
Ours-w/o-sv	63.52	62.11	62.81
Ours	64.02	63.67	63.84

从表4还可以看出,去除自我验证模块的Ours-w/o-sv模型在所有评估指标上均表现出显著的性能下降,其F1分数为62.81%,明显低于完整模型Ours的63.84%。这一结果进一步验证了自我验证模块在提升模型性能方面的关键作用。自我验证模块通过反馈机制对生成的简化子句进行有效的校正和优化,显著提高了模型对情感极性的判定能力和鲁棒性。总的来说,完整的Ours模型实现了最佳的情感分类性能,充分证明了本文方法中各个模块在提升情感分析性能方面的有效性。

4.7 评估简化子句质量

为了评估简化子句的质量,采用人工评分方法,从4.6节所述的每种简化方法中随机抽取100个简化子句,共计500个样本。邀请3名自然语言处理领域的研究生作为评分者,从情感一致性S_S、相关性S_R和简洁性S_C3个维度对样本进行独立评分。为确保评价的客观性,评分者无法获知简化子句的来源信息。评分结果如表5所列。

表5 人工评分结果

Table 5 Manual scoring results

Methods	S _S	S _R	S _C
Random Fragment	0.55	0.41	0.64
ChatGPT Extract	0.67	0.63	0.61
ChatGPT Generate ₁	0.63	0.68	0.61
ChatGPT Generate ₂	0.60	0.69	0.59
Ours	0.72	0.79	0.68

可以看出,提出的简化方法在3个评价维度上均显著优于对比方法,在相关性维度S_R上的优势最为明显。这一结果表明,相较于其他简化方法,本文方法能够更好地保持简化子句与原始评论文本在词汇分布和语义空间上的一致性,从而有效提升了简化子句的质量。

4.8 简化子句质量的影响

简化子句的质量对模型的整体性能有显著影响。为此,本文在Restaurant数据集上采用多轮迭代并结合人工检查的方式生成简化子句,旨在通过这一方法逼近模型的理论边界(Theoretical Boundaries, TB)。首先,在大型语言模型中定义方面词和观点词的概念,并分别生成仅涉及方面词或仅涉及观点词的简化子句。随后,基于仅涉及方面词的简化子句提取方面词,基于仅涉及观点词的简化子句提取观点词。接着,将提取的方面词、观点词、相关简化子句及原始文本共同输入大型语言模型,以生成最终的简化子句。最后,为确保简化子句的质量,对其进行人工检查,若子句不满足情感一致性、相

关性或简洁性中的任一标准,则人工生成符合要求的简化子句。实验结果如表 6 所列。

表 6 模型 ER 和 TB 的比较结果

Table 6 Comparison results of model ER and TB (%)

Methods	P	R	F1
Ours-ER	64.02	63.67	63.84
Ours-TB	64.83	63.80	64.31

可以看出,模型理论边界高于实证结果(Empirical Results,ER)。多轮迭代结合人工检查能够更精准地保留情感信息,减少信息冗余与情感偏移,并通过分阶段优化信息提取过程,有效降低错误传播对最终结果的影响,从而提升简化子句的相关性与简洁性。这一结果进一步验证了简化子句质量在情感分类任务中的关键作用。

结束语 本文提出了一种用于 ABSA 任务的文本简化模型。该模型首先通过大型语言模型生成原始文本的简化子句;接着采用自我验证方法对生成的简化子句进行验证,确保其满足情感一致性、相关性和简洁性,以保证简化过程不会损失重要信息或改变原文的情感色彩;最后将简化子句与原始文本结合,针对 3 个公开数据集进行了一系列实验,验证了所提模型在性能上的优势以及其组成部分的可解释性。

当前模型尽管已经取得了较好的性能,但仍有许多方面值得进一步研究。未来的工作将主要集中在以下几个方向:1)提高模型元素预识别的质量,从而使模型的最终性能逼近理论上限;2)探索该模型在其他 ABSA 任务中的应用效果,例如方面三元组抽取等任务。

参考文献

- [1] LIU P, JOTY S, MENG H. Fine-grained opinion mining with recurrent neural networks and word embeddings[C]// Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015:1433-1443.
- [2] XU H, LIU B, SHU L, et al. Double embeddings and CNN-based sequence labeling for aspect extraction[J]. arXiv:1805.04601, 2018.
- [3] ZHOU X, WAN X, XIAO J. Representation learning for aspect category detection in online reviews[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2015.
- [4] PENG H, XU L, BING L, et al. Knowing what, how and why: A near complete solution for aspect-based sentiment analysis[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2020:8600-8607.
- [5] WAN H, YANG Y, DU J, et al. Target-aspect-sentiment joint detection for aspect-based sentiment analysis[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2020:9122-9129.
- [6] ZHANG W, LI X, DENG Y, et al. A survey on aspect - based sentiment analysis: Tasks, methods, and challenges[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 35(11):11019-11038.
- [7] YAN H, DAI J, QIU X, et al. A unified generative framework for aspect-based sentiment analysis[J]. arXiv:2106.04300, 2021.
- [8] ZHANG W, LI X, DENG Y, et al. Towards generative aspect-based sentiment analysis[C]// ACL 2021. 2021.
- [9] LIU J, TENG Z, CUILI, et al. Solving aspect category sentiment analysis as a text generation task[J]. arXiv:2110.07310, 2021.
- [10] ZHANG W, DENG Y, LI X, et al. Aspect sentiment quad prediction as paraphrase generation[J]. arXiv:2110.00796, 2021.
- [11] LU Y, LIU Q, DAI D, et al. Unified structure generation for universal information extraction[J]. arXiv:2203.12277, 2022
- [12] BAO X, WANG Z, JIANG X, et al. Aspect-based Sentiment Analysis with Opinion Tree Generation[C]// IJCAI. 2022:4044-4050.
- [13] TANG D, QIN B, FENG X, et al. Effective LSTMs for target-dependent sentiment classification[J]. arXiv:1512.01100, 2015.
- [14] BU J, REN L, ZHENG S, et al. ASAP: A Chinese review dataset towards aspect category sentiment analysis and rating prediction [J]. arXiv:2103.06605, 2021.
- [15] CHEN Z, QIAN T. Relation-aware collaborative learning for unified aspect-based sentiment analysis[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020:3685-3694.
- [16] ZHANG M, QIAN T. Convolution over hierarchical syntactic and lexical graphs for aspect level sentiment analysis[C]// Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing(EMNLP). 2020:3540-3549.
- [17] GOU Z, GUO Q, YANG Y. Mvp: Multi-view prompting improves aspect sentiment tuple prediction[J]. arXiv:2305.12627, 2023.
- [18] YU J, GUO Y, RUAN Q M. Aspect-level sentiment triple extraction based on set prediction[J]. Journal of Chinese Information Processing, 2024, 38(8):147-157.
- [19] WANG Q, DING K, LUO X, et al. Improving in-context learning via sequentially selection and preference alignment for few-shot aspect-based sentiment analysis[C]// Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2024:2462-2466.
- [20] VACAREANU R, VARIA S, HALDER K, et al. A Weak Supervision Approach for Few-Shot Aspect Based Sentiment[J]. arXiv:2305.11979, 2023.
- [21] GOODING S. On the ethical considerations of text simplification [J]. arXiv:2204.09565, 2022.
- [22] BIRAN O, BRODY S, ELHADAD N. Putting it simply: a context-aware approach to lexical simplification[C]// Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics; Human Language Technologies. 2011:496-501.
- [23] GOODING S, KOCHMAR E. Recursive context-aware lexical simplification[C]// Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019:4853-4863.
- [24] SIDDHARTHAN A. Syntactic simplification and text cohesion

- [J]. *Research on Language and Computation*, 2006, 4(1): 77-109.
- [25] ALUÍSIO S, GASPERIN C. Fostering digital inclusion and accessibility; the PorSimples project for simplification of Portuguese texts[C]// *Proceedings of the NAACL HLT 2010 Young Investigators Workshop on Computational Approaches to Languages of the Americas*. 2010:46-53.
- [26] BURSTEIN J, HORBACH A, KOCHMAR E, et al. *Proceedings of the 16th Workshop on Innovative Use of NLP for Building Educational Applications*[C]// *Proceedings of the 16th Workshop on Innovative Use of NLP for Building Educational Applications*. 2021.
- [27] DEVARAJ A, SHEFFIELD W, WALLACE B C, et al. Evaluating factuality in text simplification[C]// *Proceedings of the conference. Association for Computational Linguistics, Meeting*. 2022, 2022:7331.
- [28] XIAO Z H, CHENG M M, GONG J F, et al. Research on Chinese vocabulary simplification based on hint fine-tuning[J]. *Journal of Chinese Information Processing*, 2024, 38(8):34-43.
- [29] SUN R, XU W, WAN X. Teaching the pre-trained model to generate simple texts for text simplification[J]. *arXiv*:2305.12463, 2023.
- [30] ANSCHÜTZ M, OEHMS J, WIMMER T, et al. Language models for German text simplification: Overcoming parallel data scarcity through style-specific pre-training [J]. *arXiv*: 2305.12908, 2023.
- [31] TRIENES J, JOSEPH S, SCHLÖTTERER J, et al. Info-LossQA: Characterizing and recovering information loss in text simplification[J]. *arXiv*:2401.16475, 2024.
- [32] JIANG X, YOU P, CHEN C, et al. Exploring scope detection for aspect-based sentiment analysis[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2023, 32:83-94.
- [33] LIU D, NASSERELDINE A, YANG Z, et al. Large language models have intrinsic self-correction ability [J]. *arXiv*: 2406.15673, 2024.
- [34] FLESCHE R. A new readability yardstick[J]. *Journal of applied psychology*, 1948, 32(3):221.
- [35] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// *Proceedings of the 31st International Conference on Neural Information Processing Systems*. 2017: 6000-6010.
- [36] CAI H, XIA R, YU J. Aspect-category-opinion-sentiment quadruple extraction with implicit aspects and opinions[C]// *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*. 2021:340-350.
- [37] ZHOU J, YANG H, HE Y, et al. A unified one-step solution for aspect sentiment quad prediction[J]. *arXiv*:2306.04152, 2023.
- [38] XU L, LI H, LU W, et al. Position-aware tagging for aspect sentiment triplet extraction[J]. *arXiv*:2010.02609, 2020.
- [39] PEPER J J, WANG L. Generative aspect-based sentiment analysis with contrastive learning and expressive structure[J]. *arXiv*: 2211.07743, 2022.
- [40] ZHANG W, ZHANG X, CUI S, et al. Adaptive data augmentation for aspect sentiment quad prediction[C]// *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024:11176-11180.
- [41] MA T, WANG Z, ZHOU G. Transition-based Opinion Generation for Aspect-based Sentiment Analysis[C]// *Findings of the Association for Computational Linguistics ACL 2024*. 2024: 3078-3087.
- [42] BROWN T, MANN B, RYDER N, et al. Language models are few-shot learners[J]. *Advances in Neural Information Processing Systems*, 2020, 33:1877-1901.
- [43] TOUVRON H, MARTIN L, STONE K, et al. LLaMA 2: Open foundation and fine-tuned chat models[J]. *arXiv*:2307.09288, 2023.



WANG Ye, born in 2002, postgraduate. Her main research interests include natural language processing and sentiment analysis.



WANG Zhongqing, born in 1987, Ph.D., associate professor. His main research interests include natural language processing and sentiment analysis.

(责任编辑:柯颖)