

监督对比学习与提示学习驱动的方面级情感分析

邓思远, 李晓戈, 安晓春

(西安邮电大学 计算机学院, 西安 710061)

(西安邮电大学 陕西省网络数据分析与智能处理重点实验室, 西安 710061)

E-mail: siyuandeng@stu.xupt.edu.cn

摘要: 方面情感三元组抽取是方面级情感分析领域中具有挑战性的信息抽取式任务,旨在从给定评论文本中抽取出所有方面词和观点词,并预测匹配的方面词和观点词对之间的情感极性。方面情感三元组抽取面临着多方面情感的复杂性和隐式情感的识别等挑战,目前已有许多生成方法试图解决这些问题,但现有方法仍面临两个主要挑战:1)难以有效捕捉到情感极性和方面词之间的细粒度关系;2)提示模板单一导致模型泛化能力有限。为了克服这些问题,本文提出了监督对比学习与提示学习的联合方法。首先,该方法设计了基于方面的提示,并对情感信息掩码处理,通过对比学习增强预训练模型对方面词相关情感表示的关注;其次,利用多种元素序列作为提示,指导预训练模型生成多种情感三元组序列,最终通过投票机制选出最合理的三元组。实验结果表明,该方法在公共数据集上显著提升了整体性能,验证了其有效性。

关键词: 方面级情感分析;监督对比学习;提示学习;T5

中图分类号: TP391

文献标识码: A

文章编号: 1000-1220(2026)04-0844-08

Aspect-based Sentiment Analysis Driven by Supervised Contrastive Learning and Prompt Learning

DENG Siyuan, LI Xiaoguo, AN Xiaochun

(Xi'an University of Posts & Telecommunications, Xi'an 710061, China)

(Key Laboratory of Network Data Analysis and Intelligent Processing in Shaanxi Province (Xi'an University of Posts and Telecommunications), Xi'an 710061, China)

Abstract: Aspect-based sentiment triplet extraction is a challenging information extraction task in the field of aspect-based sentiment analysis. It aims to extract all aspect terms and opinion terms from a given review text and predict the sentiment polarity of the matching aspect-opinion pairs. Aspect-based sentiment triplet extraction faces challenges including the complexity of multi-aspect sentiments and implicit emotion recognition. Although numerous generative approaches have been proposed to address these issues, existing methods still encounter two primary limitations: 1) insufficient capability to capture fine-grained relationships between sentiment polarities and aspect terms; 2) restricted model generalization caused by homogeneous prompt templates. To overcome these problems, this paper proposes a joint approach combining supervised contrastive learning and prompt learning. First, the method designs aspect-based prompts and performs sentiment information masking to enhance the pre-trained model's attention to aspect-related sentiment representations through contrastive learning. Then, it uses multiple element sequences as prompts to guide the pre-trained model in generating various sentiment triplet sequences. Finally, a voting mechanism is used to select the most reasonable triplet. Experimental results show that this method significantly improves overall performance on public datasets, demonstrating its effectiveness.

Keywords: aspect-based sentiment analysis; supervised contrastive learning; prompt learning; T5

0 引言

方面级情感分析 (Aspect-Based Sentiment Analysis, AB-SA) 旨在从方面层面分析和理解人们的观点,能够帮助本文深入分析文本中的多个方面及其对应的情感倾向,对产品优化、内容推荐和舆情分析等方面具有重要意义,在过去引起了人们的广泛兴趣。与传统的句子级情感分析不同,方面级情感分析强调情感的解释,明确突出情感的原因,其中方面情感三元组抽取 (Aspect-Sentiment Triplet Extraction, ASTE) 作为方面级情感分析的一个核心子任务提高了方面级情感分析的深

度和广度,使得分析结果更加细致和全面。方面情感三元组提取旨在从句子中提取情感三元组,这些三元组由三个组成部分构成:方面词,表示对讨论目标的明确提及;意见词,表示评论语或短语;情感极性,表示对方面词的情感倾向。这项任务涉及解决情感表达的多样性和语言环境的复杂性等挑战。例如,在餐厅领域的评价“Coffee is a better deal than overpriced Cusi sandwiches”中,可以提取两个具有相反情感极性的情感三元组:(Coffee, better, POS)和(Cusi sandwiches, overpriced, NEG)。在这个例子中,方面词“Coffee”和“Cusi sandwiches”分别为不同的意见词(“better”和“overpriced”)和相反的情感

极性(“better”为积极,“overpriced”为消极),这一任务体现了消费者对某个方面的情感是否积极,还表明了解情感背后的原因,这突出了 ASTE 能够捕捉到文本中的情感目标、情感描述以及情感极性,从而为方面情感分析提供更强的支持。

传统的 ASTE 方法包括管道方法^[1]、机器阅读理解的方法^[2,3]、基于序列标注^[4,5]的方法以及其他混合方法^[6,7]等。这些方法均存在局限性,管道方法需要训练两个模型,模型结构复杂并且可能会导致模型间的错误传播。机器阅读理解方法过于依赖上下文来预测方面-意见对的情感极性导致长距离依存问题,长距离依存问题指的是需要抽取的方面或意见词组相距距离过长,容易出现元素抽取失败或者抽取错误。基于序列标注的方法忽略了多个词组成的情感词语义信息,在提取具有复杂方面词或意见词的语句时效果不佳。近年来,因为可以通过预训练模型的生成过程有效利用丰富的上下文信息,准确地捕捉句子中的情感目标、情感极性以及它们之间的关系,生成方法已被用来统一处理各种 ABSA 任务并取得了良好的性能。Yan 等^[8]提出使用生成模型 BART 同时进行多个方面的情感分析任务,可以避免传统方法中需要手动设计特征的挑战。Zhang 等^[9]基于 T5 模型提出了统一 GAS 框架,定制了注释式建模和抽取式建模两种范式,将原始任务转化为文本生成问题,利用生成模型生成与方面相关的情感输出,展示了生成模型在处理长文本和多方面情感时的优势。Lu 等^[10]提出了一种通用的生成式框架,旨在通过一个通用框架来处理多个信息抽取任务,包括方面情感分析、关系抽取、事件抽取等。

然而,生成式方法仍然面临两个重大挑战。首先,现有的基于 Transformer 架构的预训练模型,通常缺乏足够的监督来挖掘与细粒度情感分析任务相关的情感信息,虽然通过监督对比学习(Supervised Contrastive Learning, SCL)^[11]增强了情感判别能力,却难以在细粒度层面准确识别和关联方面词与其对应的情感极性^[12]。因为现有技术将监督对比学习应用于句子级情感表达,相比之下,在 ASTE 中,情感是在方面层面定义的,这可能导致情感极性的错误预测。

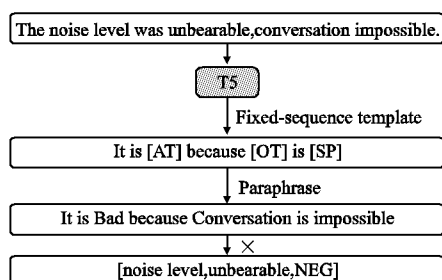


图1 PARAPHRASE 中的固定序列方法
Fig. 1 Fixed sequence in PARAPHRASE

其次,以前的生成方法通过提示学习引导模型在固定的从左到右的序列中生成情感元素序列,忽视了情感三元组中元素的相互依赖性和无序性质^[13]。例如图1所示,在 PARAPHRASE^[14]中的固定“s→a→o”序列会导致不确定性,因为情感元素之间的关系不是有序的,而是相互依赖的。Hu 等^[15]的研究表明,基于不同目标的提示模板序列的性能存在显著差异,并且以前的预测误差可能会累积并影响后续预测。

为了解决上述挑战,本文提出了一种的多序列监督对比提示方法。本文的方法结合了基于方面的监督对比学习和提示学习以增强 ASTE 性能。具体来说,本文将一个句子输入到 T5 模型^[16]的编码器中,再将带有 [MASK] 标记的基于方面的提示传递给解码器,这些 [MASK] 标记的词嵌入作为对应情感标签的方面感知表示,本文对这些嵌入进行监督对比学习,表1显示了几个示例句子以及从中派生的用于预训练的基于方面的提示。

表1 句子示例以及从中派生的基于方面的提示来预训练编译器-解码器框架

Table 1 Sentences examples and aspect-based prompts derived from them fro pre-training enc-dec frameworks

句子	预训练提示
Fish was overdone.	< aspect > Fish < sentiment > [MASK]
Waiting for the i7 was, well worth it, great value for the price.	< aspect > price < sentiment > [MASK] < aspect > i7 < sentiment > [MASK]
Service was slow , but the people were friendly.	< aspect > people < sentiment > [MASK] < aspect > Service < sentiment > [MASK]

接着,在预训练模型后,受提示链^[17,18]的启发,本文引入了控制情感元素生成顺序的标记序列提示。与前人研究的固定序列生成相比,这些标记可以自由排列并聚合不同序列生成的情感元素以减少不稳定性,从而提取出正确的方面情感三元组。

总之,本文的贡献有两方面:

1) 本文引入方面级的监督对比学习,以增强模型对方面情感表示的感知能力。具体来说,给定一个句子,引入基于方面的提示,并掩码相应的情感极性。通过在解码器的 [MASK] 标记所代表的方面情感向量上执行监督对比学习来对 T5 模型进行预训练。通过捕捉更专注于方面词的信息来提高 ASTE 任务的性能。

2) 本文采用基于标记序列的提示学习,使用不同的提示序列引导模型生成多样化的情感三元组,通过允许情感元素自由排列,本文的方法解决了固定序列生成方法的局限性,例如错误传播和处理复杂情感表达时的不灵活性,并通过投票策略提高三元组预测稳定性。

1 相关工作

1.1 监督对比学习

监督对比学习是一种通过在标签约束下学习数据表示的方法,其目标是使得具有相同标签的样本在嵌入空间中更加接近,而具有不同标签的样本则相互远离。与传统的分类任务不同,监督对比学习侧重于学习样本之间的相对距离关系,而不仅仅是单一标签的分类。Wang 等^[19]提出了一种面向文本聚类的自监督学习框架 CEIL,其目的是旨在通过引入分类目标来迭代改进特征表示,从而在整体上提高聚类算法的聚类性能。Li 等^[20]使用提出了一种结合对比学习的有监督预训练方法,以对齐隐式情感与显式情感,从而提升模型在低资源情况下的方面级情感分析能力。Liang 等^[21]在 BERT 的基础上,结合 SCL 方法进行训练,优化方面和情感的联合表示。虽然

监督对比学习在基于方面的情感分析领域进行过探索,但是他们的目标与本文的不同,因为他们将对对比学习应用于句子级情感表示,而在方面情感三元组提取中情感本质上是在方面级别定义的,本文在处理掩盖了情感对应的观点词的同时,保持了模型对方面词的感知.之前的研究多应用 SCL 于只具有编码器架构的 BERT,本文使用 SCL 应用于具有编码器-解码器架构的 T5 模型,可以同时优化输入的和输出的生成,增强模型的情感区分能力.

1.2 提示学习

随着 GPT-3 的出现,提示学习^[22] (prompt learning) 作为一种新的基于 PLMs 的一种新的学习范式受到了越来越多的关注,相较于传统的微调方法,它更加充分地利用了模型的知识.具体来讲,提示学习是通过在输入文本中添加自然语言提示或特殊的掩码标记,来重构输入文本. Huang 等^[23] 以 RoBERTa 模型为基础,使用不同于传统微调方法的集成提示方法,通过提示帮助模型进一步理解下游任务,改善模型对文本情感特征的提取能力. Gao 等^[24] 提出一种改进的 GPT-3 模型,对不同任务的不同标签构建多个模板再筛选,在少量训练样本时就达到较好的效果. Schick 等^[25] 提出人工构建模板然后利用大规模未标注的外部语料库进行知识蒸馏的方法,应用于文档级情感分类任务. Hu 等^[26] 提出利用外部知识库来扩展分类任务的标签词映射器. Shin 等^[27] 提出了一种基于梯度指导的方法为下游任务自动构建提示模板.但是这些模型都没有关注到提示模板的不同顺序对于任务的影响.

1.3 基于预训练模型的方面级情感分析

最近,生成方法^[1,14,28] 通过减轻管道方法中潜在的错误传播并充分利用丰富的标签语义信息在 ABSA 中获得了良好的性能.他们选择不同的预训练模型并将情感元素序列^[15]、自然语言^[29,30,31] 和结构化提取模式^[10] 等方式作为生成目标. Zhang 等^[32] 提出将大语言模型与本地预训练模型相结合,使用多级推理的方式逐级得出目标的方面、潜在观点,使模型能够更轻松地推理出最终的情感极性. LEGO-ABSA^[33] 和 UnifiedABSA^[34] 通过利用任务提示或指令设计进行多任务处理. DLO^[35] 采用数据增强的方法提高方面情感四元组预测任务的效果,但这种增强方法会使模型与同一输入的多个目标混淆,从而导致推理和训练之间的差异.本文提出的方法通过将训练和推理与多序列提示学习消除了这种差异.同时,本文采用广泛应用于方面级情感分析任务的 T5^[16] 模型,该模型基于 Transformer 架构,相较于只采用编码器架构的 BERT 模型和解码器架构的 GPT 模型, T5 模型结合了两者的优势,通过编码器-解码器架构统一理解双向信息和文本生成任务.

2 方法

本文提出的模型整体架构图 2 所示.本文使用基于方面的提示在编码器-解码器模型上进行对比学习,然后基于标记为三元组采用 6 种不同的提示序列,依据多样化的提示序列,

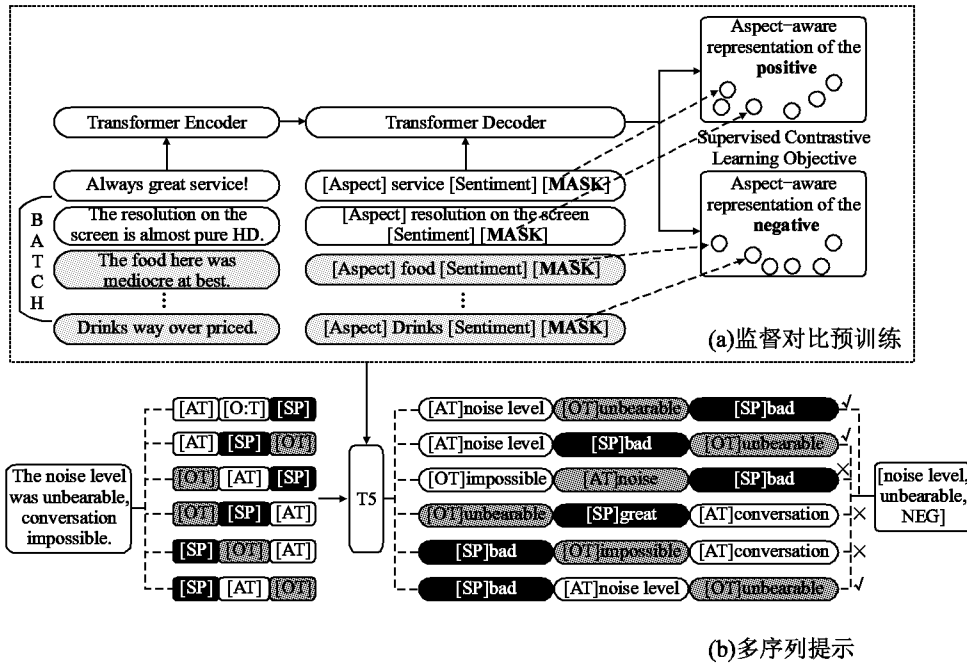


图 2 多序列监督对比提示的方面情感三元组提取

Fig. 2 Multi-sequence supervised contrastive prompting for aspect sentiment triplet extraction

模型能够生成不同的方面情感三元组.本研究采用聚合投票机制来选择出现频率最高的三元组作为最终结果,这种方式可以纠正随机误差错误,从而确保模型的鲁棒性和结果的可靠性.

2.1 任务定义

在这一节中,本文正式定义了 ASTE 任务如下:给定一个

输入句子 $X = \{x_1, x_2, \dots, x_n\}$, 其中每一个 x_i 代表一个单词, n 是句子的最大长度. ASTE 任务是抽取输入句子 X 中所有的情感三元组 $T = \{(a, o, s)\}_{i=1}^N$, 其中 N 表示三元组数量. 每个三元组由一个方面词 a , 一个意见词 o , 以及相应的情感极性 s . 在预测情感元素时,为了利用标签的语义,本文根据先前的工作^[30] 将这些元素解释为自然语言. 例如,情感极

性标签 NEG 被映射为“bad”,而标签 POS 被映射为“great”。

2.2 监督对比学习

本文将 ASTE 任务建模为一个结构化预测问题,并使用 T5 编码器-解码器框架作为模型架构的主干。为实现对方面级情感信息的精细捕捉,本文使用基于方面的提示持续预训练整个模型,解码器从编码器生成的输入句子的上下文表示中学习,同时接收一个基于方面的提示,如“[AT] service [SP][MASK]”,然后解码[MASK]得到特定方面词对应的情感极性向量表示,即预测方面词“service”对应的情感极性表示。本文对在[MASK]位置生成的情感向量表示进行监督对比学习,使得相同情感的方面词的情感极性嵌入表示更加接近,而不同情感的方面词的情感极性嵌入彼此推远。这样可以提升模型在多方面词情感冲突场景下的判别能力,能更精准地捕捉方面词的独立情感。

为了预训练模型,本文首先使用基于方面的提示 p_i 针对每个方面术语 a_i 从给定的句子 sen_i 中获得方面感知的情感表示。尽管一个句子可以有多个方面,但组合 (sen_i, p_{ij}) 只有一个单一的情感标签 S_{ij} 。本文将数据点 (sen_i, p_{ij}) 表示为 x_i , 并将相应的情感标签表示为 y_i 。如图 2(a) 所示,本文首先将输入句子传递给编码器,并向解码器提供基于方面的提示。这个提示包括一个[MASK]标记,该标记掩盖了与该方面词相关的情感。解码器对[MASK]标记的输出形成了对应于数据点 x_i 的以方面为中心的情感表示 z_i 。第 I 批次的监督对比损失定义如下:

$$L_I^{Sup} = \sum_{i \in I} - \frac{1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(z_i \cdot z_p / \tau)}{\sum_{a \in A(i)} \exp(z_i \cdot z_p / \tau)} \quad (1)$$

其中索引 i 表示锚点, $A(i) \equiv I \setminus i$ 代表除锚点之外的所有索引的集合, $P(i) = \{p \in A(i) : y_p = y_i\}$ 是与 i 不同的所有正例(相同情感标签)的索引集合,并且 $|P(i)|$ 是与 i 属于同一类的其他索引数量。 \cdot 符号表示两个嵌入之间的内点积, $\tau \in R^+$ 是标量温度参数。

2.3 基于标记序列的提示学习

为了控制情感元素的预测序列,如图 2(b) 所示,本文引入了基于标记序列的提示机制。具体来说,本文使用有序目标模式,并将相应的标记序列作为提示输入。为了表示不同的情感元素,本文遵循 DLO 方法^[15],并设计特殊标记来表示信息的结构^[1]。方面词、观点词和情感极性的标记分别是[AT]、[OT]和[SP]。通过连接这些标记,本文设计了指示情感元素生成顺序的标记序列提示(例如,“[OT][AT][SP]”表示按照“o→a→s”的预测序列)。这些标记的排列称为 p_i , 然后这些标记序列提示被追加到每个输入句子的后缀,以创建最终输入 x_{p_i} 。接着,本文将相应的标记作为前缀添加到每个元素,并按照给定的排列 p_i 将它们连接形成目标序列。如果输入句子包含多个情感元素,本文将使用特殊标记[SSEP]来连接它们各自的目标序列,从而得到最终的目标序列 y_{p_i} 。因此,获得了用于训练的输入-输出对,例如输入: Quick and friendly service[OT][AT][SP] 输出:[OT] quick[AT] service[SP] great[SSEP][OT] friendly[AT] service[SP] great。

本文发现,设计标记序列提示有效地指导了情感元素的生成顺序,从而通过标记的排列灵活完成任务。由于不同序列的性能各异,选择适当的标记顺序至关重要。在研究提示排

序^[15]后,本文基于训练集上候选排列的平均熵选择潜在表现更好的序列。步骤如下:1)使用情感元素的所有可能标记排列 p_i 作为候选;2)给定一个输入句子 x 及其目标三元组,本文构建排列 p_i 的有序目标模式 y_{p_i} , 将其中的标记替换为空格以避免噪声,并查询预训练的模型以获得条件生成分数 $P(y_{p_i} | x)$;3)计算标记排列 p_i 在训练集 D 上的平均分数:

$$S_{p_i} = \frac{\sum_D P(y_{p_i} | x)}{|D|} \quad (2)$$

其中, $|D|$ 表示训练集 D 中样本的数量。通过计算每个排列的平均分数,可以评估哪些排列在训练集上表现更好,并选择前 n 个平均分数较高的排列作为生成情感元素的顺序训练。这种方法有助于优化模型的性能,并通过选择信息熵较低的排列来减少预测过程中的不确定性和错误。

在训练时,本文选择标记序列的组合来构建输入-目标对。通过选择 n 个排列,本文为每个句子构建 n 个不同的有序提示和目标。给定一个输入目标对 (x, y) , 预训练序列到序列的语言模型如 T5, 可以通过最小化以下负对数似然损失来进行微调:

$$LNLL = - \sum_{t=1}^T \log P(y_t | y_{<t}, x) \quad (3)$$

其中, T 是目标序列 y 的长度, $y_{<t}$ 表示先前生成的标记。

2.4 多序列推理

推理时,本文提示训练好的生成模型根据先前选定的序列生成多个情感三元组。由于每个序列可能会预测多个三元组,本文首先聚合所有序列的结果,然后使用在多个三元组预测中出现超过一半次数的三元组。具体来说,对于输入句子 x , 本文提示模型从 n 个选定的排列中生成情感三元组,排列 p_i 的预测情感三元组集合是 T'_{p_i} , 其中可能包含一个或多个情感三元组。然后通过以下等式获得最终聚合结果 T'_{Model} :

$$T'_{Model} = \left\{ t \mid t \in \bigcup_{i=1}^n T'_{p_i} \ \& \ \left(\sum_{i=1}^n I_{T'_{p_i}}(t) \geq \frac{n}{2} \right) \right\} \quad (4)$$

3 实验

3.1 数据集

在 ASTE-Data-V2 数据集上验证所提模型的有效性。这些数据集包括 14Res、14Lap^[36]、15Res^[37] 和 16Res^[38], 这些数据集标注了方面词、意见词和情感极性的三元组信息。如表 2

表 2 ASTE-V2 数据集的统计信息

Table 2 Statistics of the ASTE-V2 datasets

Datasets	#S	#T	POS	NEU	NEG	
Lap14	Train	906	1460	817	126	517
	Dev	219	345	169	36	141
	Test	328	541	364	63	116
Res14	Train	1266	2337	1692	166	480
	Dev	310	577	404	54	119
	Test	492	994	779	66	155
Res15	Train	605	1013	783	25	205
	Dev	148	249	185	11	53
	Test	322	485	317	25	143
Res16	Train	857	1394	1015	50	329
	Dev	210	339	252	11	76
	Test	326	514	407	29	78

所示为 4 个数据集的训练集、验证集以及测试集的统计结果, 其中“S”为句子的数量,“T”为对应三元组的数量,“POS”、

“NEU”和“NEG”分别表示积极、中性和消极三元组的数量。

3.2 参数设置

在实验中基于深度学习框架 Pytorch 来搭建模型。模型采用 T5 的基础版本^[39], T5 模型是一个经典的编码器解码器架构^[40]。采用 AdamW 优化器来优化模型,学习速率为 2e-5,最小批量大小为 16,丢弃率为 0.1,温度参数设置为 0.07。在解码过程中默认使用贪婪搜索,排列的数量设置为 5。在 NVIDIA RTX 4090 GPU 上训练的迭代轮次为 20。

3.3 评估指标

采用 ASTE 任务性能中常用的 3 种评价指标:准确率 P,召回率 R 和 F1 值来评价模型的综合性能,计算分别为公式(5)~公式(7)所示:

$$P = \frac{TP}{TP + FP} \quad (5)$$

$$R = \frac{TP}{TP + FN} \quad (6)$$

$$F1 = \frac{2 \times P \times R}{P + R} \quad (7)$$

其中, TP 为模型正确预测的正样本数量, FP 为模型错误预测的正样本数量, FN 为模型未能正确预测的正样本数量。

3.4 对比实验

本文将所提模型和其他预训练模型的基准模型进行性能比较。通常可以分为以下几类:

基于管道的方法: CMLA 和 RINAME^[41] 是联合提取方面词和观点词的管道方法。CMLA + 、RINANTE + 是 Peng 等^[42] 提出的改进版本,这些方法联合提取方面、观点以及相应的情感。

表 3 在 ASTE 任务的 4 个数据集上的评估结果

Table 3 Evaluation results on 4 datasets of the ASTE task

Methods	14Lap			14Res			15Res			16Res		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
CMLA +	30.06	36.94	33.21	39.21	47.13	42.81	36.64	39.81	37.03	41.26	42.13	41.69
RINANTE +	21.73	18.75	20.13	31.41	39.44	35.03	29.89	30.08	30.02	25.73	22.28	23.98
BMRC	70.55	48.98	57.82	75.61	61.77	67.99	68.51	53.40	60.02	71.20	61.08	65.75
Dual-MRC	57.39	53.88	55.58	71.55	69.14	70.32	63.78	51.87	57.2	68.60	66.64	67.40
JET-BERT	55.39	47.33	51.04	70.56	55.94	62.40	64.45	51.96	57.53	70.42	58.37	53.83
GTS-BERT	57.82	51.32	54.36	67.76	67.29	67.50	62.59	57.94	60.15	66.08	69.91	67.93
BDTF	68.94	55.97	61.74	75.53	53.24	74.35	68.76	63.71	66.12	71.44	73.13	72.27
MWWPT	61.73	60.58	61.15	72.92	72.18	72.53	61.12	61.16	61.64	69.25	75.90	72.42
GAS *	61.65	58.19	59.87	71.08	71.67	71.37	60.01	63.67	61.78	67.76	71.67	69.66
PARAPHRASE *	62.99	58.30	60.55	70.87	70.90	70.89	60.80	64.98	62.82	70.35	74.04	72.15
DLO *	-	-	61.46	-	-	72.39	-	-	64.26	-	-	73.03
Seq2Path *	64.57	60.04	62.22	73.28	74.23	73.75	62.62	65.48	64.02	71.59	75.41	73.40
Ours *	64.89	62.85	63.85	74.48	75.45	74.96	66.67	69.69	67.15	71.32	76.46	73.80

注: * 数据代表基于生成的方法,最佳和次佳的 F1 分数分别以粗体和下划线标注。型的最佳结果相比,本文方法在 14Lap、14RES、15Res 和 16Res 数据集上的值分别提高了 1.63%、0.61%、1.03% 和 0.40%,证明了本文方法的有效性。

本文观察到,管道模型由于精度较低,他们的整体性能要差得多。基于机器阅读理解的方法和基于标注的方法比管道方法更具竞争力,因为它们减轻了错误传播并建立了相关子任务之间的相关性。基于生成的方法因为更充分利用丰富的标签语义,所以在 ASTE 任务中获得了较好的表现。

本文设计的基于方面的提示,通过对[MASK]令牌的解码器生成的方面感知情感嵌入执行监督对比学习来预训练本文的编码器-解码器框架。基于这一预训练方法,通过聚合多个序列的结果,本文的方法超越了以前的多数单序列方法。与

基于机器阅读理解的方法:BMRC^[2]提出了一种新的方法可以将 ASTE 任务处理为双向机器阅读理解任务,这种方法通过考虑文本的双向信息流来提取方面和情感信息。Dual-MRC^[3]构建了两个机器阅读理解问题,并通过联合训练两个参数共享的 BERT-MRC 模型来解决所有子任务。

基于标注的方法:JET-BERT^[4]是一个端到端的方法,它引入了一种新颖的位置感知标注方案,考虑词语在句子中的位置信息来提高方面词和意见词的提取精度。GTS-BERT^[5]将 ASTE 任务建模为一个网格标注任务,将文本中的每个词语视为网格中的节点并考虑节点之间的关系。BDTF^[43]将每个三元组表示为一个关系区域中的二维表,并转换成检测和分类的关系区域的 ASTE 任务。MWWPT^[44]基于 BERT 模型构建多窗口特征学习机制,通过动态注意力机制融合局部情感特征与深层语义表征,并采用优化标注策略实现文本三元组的端到端解析。

基于生成的方法:GAS^[9]提出了一个基于 T5 的解决方案,使用线性化的模板目标。这种方法通过将生成任务转化为一个模板化的线性问题来简化生成过程,使得模型能够更直接地生成所需的情感三元组。PARAPHRASE^[14]建议将三元组解码为模板化的自然语言释义。Seq2Path^[45]生成三元组作为树的路径,然后选择有效的三元组。DLO^[15]通过模板顺序排列设计了数据集级别的数据增强。

3.5 实验结果

本文模型和基线模型的实验结果如表 3 所示,总体上本文模型在 4 个数据集上的测试结果优于基线模型。与基线模

以固定序列应用标记模板的方法 Paraphrase 相比,本文平均提高了 3.33%,验证了聚合多序列的有效性。基于标记序列的提示学习通过训练和推理来有效指导三元组的生成。与目标端的 DLO 增强数据相比,本文通过设计基于标记序列的提示来可控地生成目标三元组,获得了 2.16% 的改进。

3.6 实验分析

3.6.1 消融实验

为了研究本文模型中不同模块的有效性,在监督设置下对 4 个数据集进行了消融研究。其中 T5-SCL 表示模型使用通过对[MASK]令牌的解码器生成的方面感知情感嵌入执行监督对比学习来预训练本文的编码器-解码器框架;T5-PROMPT 表示模型不通过执行监督对比学习来预训练编码

器-解码器框架,仅依赖于基于标记序列的提示学习训练和微调模型。

表 4 消融实验结果(%)

Table 4 Result of ablation experiment(%)

Model	Lap14	Res14	Res15	Res16
T5-SCL	61.43	72.81	64.78	72.99
T5-PROMPT	62.95	74.14	65.25	73.55
Full Model	63.85	74.96	67.15	73.80

上述模型的消融实验结果如表 4 所示. 平均 F1 分数数据显示,没有进行监督对比学习的 T5-PROMPT 比于完整模型在 14Lap 数据集上降低 0.9%, 在 14Res 数据集上降低 0.25%, 在 15Res 数据集上降低了 1.90%, 在 16Res 数据集上降低了 0.81%, 说明不同数据集对更细粒度的方面词敏感度不同;没有序列标记提示策略的 T5-SCL 相比于完整模型在 14Lap 数据集上降低 2.42%, 在 14Res 数据集上降低 2.1%, 在 15Res 数据集上降低了 2.37%, 在 16Res 数据集上降低了 0.81%, 说明序列标记提示能够更好的学习句中的语义信息, 从而提升模型性能。

3.6.2 监督对比预训练方法的比较

如 2.2 节所讨论的,本文提出了一种基于方面的提示方法,本文提出的预训练策略借鉴了掩码语言模型的掩码预测机制,但将其扩展至方面级情感分析任务中. 为了进一步分析 [MASK] 在本任务中的作用,本文从 Lap14 数据集的验证集中抽取数据,本文对解码器在 [MASK] 位置生成的情感向量进行了监督对比预训练前后的 t-SNE 可视化分析。

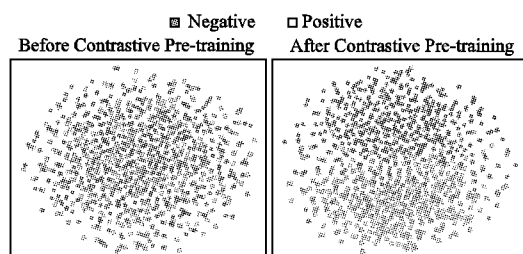


图 3 t-SNE 可视化

Fig. 3 t-SNE visualization

结果如图 3 所示,经过监督对比预训练后,相同情感类别的 [MASK] 表示在二维空间中更加紧密地聚集,而不同情感极性的表示则明显分离. 这些现象表明, [MASK] 的设计有效引导解码器专注于特定方面词的情感极性生成,验证了 [MASK] 在引导模型学习方面情感表达中的关键作用,从而提升了模型在多方面词情感冲突场景下的判别能力。

同时本文将提出的预训练策略与现有基于方面的情感分析工作中采用的策略进行了比较. 例如, Li 等^[20] 使用监督对比学习进行预训练, 而 Liang 等^[21] 则使用它来微调他们各自的模型. 与本文不同的是, 他们都将监督对比学习应用于句子级别的情感表示. 为了在实验中与他们的方法进行比较, 本文通过将监督对比学习应用于编码器最后一层的句子的均值池化表示来预训练本文的编码器-解码器框架, 并将其称为 T5-SCL-SENT. 模型权重使用 T5-base, 框架使用 AdamW 优化器预训练 20 个周期. 本文使用批量大小为 16, 学习率为 1e-5. 本

文从这个预训练检查点为每个数据集单独微调模型,并在表 5 中报告了实验的结果。

表 5 监督对比预训练方法对比结果

Table 5 Supervised contrastive pre-training method comparison results

Model	Lap14	Res14	Res15	Res16
T5-SCL-SENT	63.45	74.05	64.33	73.42
Full Model	63.85	74.96	67.15	73.80

本文观察到 F1 分数总体平均提高了 1.1%. 因为在 ASTE 任务中,情感极性是依赖于具体方面的,而非整个句子. 在本文的编码器-解码器架构中, [MASK] 标记使解码器生成特定方面对应的情感极性向量,通过对该向量进行监督对比学习可以提升模型在多方面词情感冲突场景下的判别能力. 这证明相比监督对比学习应用于句子级别的情感表示,对以特定方面为中心的情感表示执行监督对比学习更加适合。

3.6.3 多序列提示策略效果

本文首先研究了不同序列聚合策略的有效性,如表 6 所示. Model-single 是一种单序列策略,它在推理过程中根据生成的多序列的预测分数选择排名最高的序列. 本文观察到,用推理过程中预测得分最高的单序列替换多序列投票会导致 F1 分数的下降,表明多序列投票是处理复杂 ASTE 任务时更加稳定。

表 6 根据推理中预测分数最高选择单序列和多序列投票策略对比

Table 6 Comparison of single-sequence selection based on inference scores and multi-sequence voting strategy

Model	Lap14	Res14	Res15	Res16
Model-single	61.36	71.98	66.26	72.46
Full Model	63.85	74.96	67.15	73.80

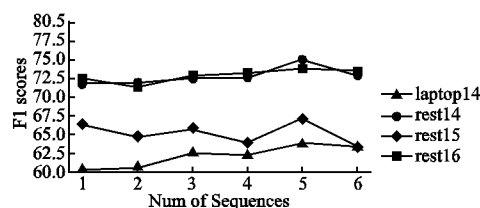


图 4 基于不同数量的基于标记序列提示的 F1 分数

Fig. 4 F1 scores of different numbers of token-based sequence prompts

同时本文的方法提出一个问题,即在标记序列提示策略中训练和推理时候选择了足够数量的序列. 本文通过改变序列数量来进一步探索这个问题,结果如图 4 所示。

可以观察到随着序列数量的增加,曲线最初呈现出上升趋势,但是在合并了全部 6 个序列的情感三元组结果之后, F1 分数出现了轻微的下降. 本文认为,一些排名较低的序列可能效果较差,并引入了噪声. 基于这些发现,本文确定在提出的模型中使用 5 个序列提示并投票获得结果,以确保在性能和计算效率之间取得了最佳平衡。

3.6.4 不同规模 T5 模型对任务性能的影响

在前述实验中,本文采用了 T5-base 作为主要实验模型。

然而,大语言模型的规模对任务性能的影响仍然是一个值得探讨的问题.通常,较大规模的模型具有更强的特征表达能力,但同时也伴随着更高的计算成本.因此,本文在相同的数据集和训练超参数和训练策略一致下,进一步研究 T5 系列模型中 T5-small、T5-base 以及 T5-large 这 3 个不同参数规模的模型对方面级情感分析任务的影响,以评估模型大小与任务效果之间的关系.其中 T5-small 参数规模约为 60M、T5-base 为 220M 以及 T5-large 为 770M.实验结果如表 7 所示.

表 7 不同规模 T5 模型在 ASTE 任务上的 F1 分数

Table 7 F1 scores of T5 models of different scales on ASTE tasks

Model	Lap14	Res14	Res15	Res16
T5-small	59.18	71.26	61.05	70.15
T5-base	63.85	74.96	67.15	73.80
T5-large	60.44	72.87	66.47	73.76

从表 7 可以观察到,T5-base 在所有数据集上的性能均优于 T5-small 和 T5-large,说明模型规模的增加并不总是带来性能提升.从 T5-small 到 T5-base,模型参数规模增加,性能在大多数数据集上都有显著提升.在 Lap14 数据集上,F1 分数从 59.18% 提升到 63.85%,Res15 数据集上从 61.05% 提升到 67.15%.这表明随着模型规模的增加,能够更好地捕捉方面、观点和情感之间的关系,从而在 ASTE 任务上表现更好.而从 T5-base 到 T5-large,模型参数规模进一步增加,但性能提升并不明显,甚至在某些数据集上出现了轻微的下降.在 Lap14 数据集上,F1 分数从 63.85% 下降到 60.44%.

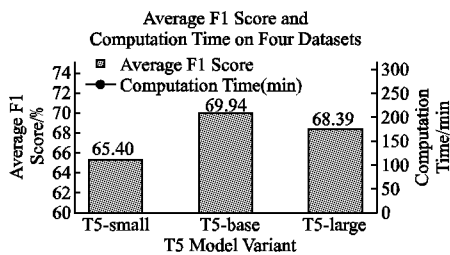


图 5 不同规模 T5 模型在 4 个数据集上的平均 F1 值与计算时间对比

Fig. 5 Comparison of F1 values and computation time of T5 models of different scales on four datasets

如图 5 所示,T5-base 在 F1 值和计算时间之间取得了最佳平衡.图 5 展示了 T5-small、T5-base 和 T5-large 模型在 4 个数据集上的平均 F1 值和计算时间.可以看出,T5-base 的 F1 值最高,达到了 69.94%,优于 T5-small 和 T5-large,T5-large 模型规模最大且计算时间最长,说明 T5-base 在模型复杂度和性能之间取得了较好的平衡.这表明模型规模的增加并不总是带来性能的提升,T5-large 虽然参数更多,但并未带来显著的性能提升,反而增加了训练难度和计算成本.

实验结果表明,模型大小与 ASTE 任务性能之间存在一定的关系,但这种关系并不是线性的.T5-base 在大多数数据集上表现最佳,而 T5-large 的性能提升并不显著并且需要更多的计算资源和训练时间,甚至在某些数据集上出现了下降.因此在资源有限的情况下,使用 T5-base 是更优的选择.

4 结语

方面级情感分析(ABSA)主要关注情感的解释和情感极性的识别,方面情感三元组(ASTE)进一步提升了情感分析的细粒度能力,因为该任务不仅要求识别方面的情感极性,还需要深入挖掘情感与特定方面之间的关联,提供更为细致的情感分析.现有方法在细粒度情感分析能力和固定提示序列的泛化性方面仍然存在挑战.针对这些问题,本文提出了一种名为基于监督对比学习和提示学习的方面情感分析方法,它利用对比学习和基于标记序列的提示来持续预训练 T5 模型以处理 ASTE 任务.为了实现这一点,给定一个句子,本文首先获取基于方面的提示并掩盖相应的情感.然后,本文对解码器中掩盖标记的方面级情感嵌入应用监督对比学习,显著提高了模型捕捉细粒度情感信息的能力.与在句子级情感嵌入上执行监督对比学习相比,本文展示了在词级情感嵌入上执行监督对比学习可以使 ASTE 任务的下游性能上取得了更好的结果.随后,训练好的模型通过灵活的多序列提示和聚合结果改进了三元组情感元素的生成,解决了固定序列方法的局限性.本文的方法在大多数 ASTE 基准数据集上取得了最先进的结果.在这项工作中,本文仅限于现有 ASTE 的训练集.然而,使用更多数据进行预训练的影响仍有待研究.在后续研究中,本文希望可以确认有多少数据足以有效地预训练模型.

References:

- [1] Paolini G, Athiwaratkun B, Krone J, et al. Structured prediction as translation between augmented natural languages[C]//International Conference on Learning Representations, 2021:1-26.
- [2] Chen S, Wang Y, Liu J, et al. Bidirectional machine reading comprehension for aspect sentiment triplet extraction[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2021:12666-12674.
- [3] Mao Y, Shen Y, Yu C, et al. A joint training dual-mrc framework for aspect based sentiment analysis[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2021:13543-13551.
- [4] Xu L, Li H, Lu W, et al. Position-aware tagging for aspect sentiment triplet extraction[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2020:2339-2349.
- [5] Wu Z, Ying C, Zhao F, et al. Grid tagging scheme for aspect-oriented fine-grained opinion extraction[C]//Findings of the Association for Computational Linguistics, 2020:2576-2585.
- [6] Xu L, Chia Y K, Bing L. Learning span-level interactions for aspect sentiment triplet extraction[C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2021:4755-4766.
- [7] Chen H, Zhai Z, Feng F, et al. Enhanced multi-channel graph convolutional network for aspect sentiment triplet extraction[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2022:2974-2985.
- [8] Yan H, Dai J, Ji T, et al. A unified generative framework for aspect-based sentiment analysis[C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2021:2416-2429.
- [9] Zhang W, Li X, Deng Y, et al. Towards generative aspect-based sentiment analysis[C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), 2021:504-510.
- [10] Lu Y, Liu Q, Dai D, et al. Unified structure generation for universal information extraction[C]//Proceedings of the 60th Annual Meet-

- ing of the Association for Computational Linguistics (Volume 1: Long Papers),2022 :5755-5772.
- [11] Khosla P, Teterwak P, Wang C, et al. Supervised contrastive learning[J]. Advances in Neural Information Processing Systems,2020, 33 :18661-18673, doi:10.48550/arXiv.2004.11362.
- [12] Su Y, Lan T, Wang Y, et al. A contrastive framework for neural text generation[J]. Advances in Neural Information Processing Systems, 2022,35 :21548-21561, doi:10.48550/arXiv.2202.06417.
- [13] 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.
- [14] Zhang W, Deng Y, Li X, et al. Aspect sentiment quad prediction as paraphrase generation[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing,2021 :9209-9219.
- [15] Hu M, Wu Y, Gao H, et al. Improving aspect sentiment quad prediction via template-order data augmentation[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing,2022 :7889-7900.
- [16] Raffel C, Shazeer N, Roberts A, et al. Exploring the limits of transfer learning with a unified text-to-text transformer[J]. Journal of Machine Learning Research,2020,21(140) :1-67.
- [17] Wei J, Wang X, Schuurmans D, et al. Chain-of-thought prompting elicits reasoning in large language models[J]. Advances in Neural Information Processing Systems, 2022, 35 :24824-24837, doi:10.48550/arXiv.2201.11903.
- [18] Wang X, Wei J, Schuurmans D, et al. Self-consistency improves chain of thought reasoning in language models[J]. arXiv preprint arXiv:2203.11171,2022.
- [19] WANG M Z. Research on text clustering based on self-supervised contrastive learning[D]. Beijing:North China Electric Power University,2023.
- [20] Li Z, Zou Y, Zhang C, et al. Learning implicit sentiment in aspect-based sentiment analysis with supervised contrastive pre-training [C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing,2021 :246-256.
- [21] Liang B, Luo W, Li X, et al. Enhancing aspect-based sentiment analysis with supervised contrastive learning [C]//Proceedings of the 30th ACM International Conference on Information & Knowledge Management,2021 :3242-3247.
- [22] Liu P, Yuan W, Fu H N G. Pre-train, prompt, and predict: a systematic survey of prompting methods in natural language processing [J]. ACM Computing Surveys,2023,55(9) :1-35.
- [23] HUANG T F, MA J. Text sentiment classification algorithm based on prompt learning enhancement[J]. Data Analysis and Knowledge Discovery,2024,8(3) :77-84.
- [24] Gao T, Fisch A, Chen D. Making pre-trained language models better few-shot learners[J]. arXiv preprint arXiv:2012.15723,2020.
- [25] Schick T, Schütze H. Exploiting cloze-questions for few-shot text classification and natural language inference[C]//Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume,2021 :255-269.
- [26] Hu S, Ding N, Wang H, et al. Knowledgeable prompt-tuning: incorporating knowledge into prompt verbalizer for text classification [C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2022 :2225-2240.
- [27] Shin T, Razeghi Y, Logan IV R L, et al. AutoPrompt: eliciting knowledge from language models with automatically generated prompts[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing,2020 :4222-4235.
- [28] Yu C, Wu T, Li J, et al. Syngen: a syntactic plug-and-play module for generative aspect-based sentiment analysis[C]//IEEE International Conference on Acoustics, Speech and Signal Processing, 2023 :1-5.
- [29] Liu J, Teng Z, Cui L, et al. Solving aspect category sentiment analysis as a text generation task[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing,2021 :4406-4416.
- [30] Zhang W, Deng Y, Li X, et al. Aspect-based sentiment analysis in question answering forums [C]//Findings of the Association for Computational Linguistics,2021 :4582-4591.
- [31] Bao X, Wang Z, Jiang X, et al. Aspect-based sentiment analysis with opinion tree generation[C]//Proceedings of the 31st International Joint Conference on Artificial Intelligence,2022 :4044-4050.
- [32] ZHANG X Y, YAN Z. Three-level joint prompt-tuning implicit sentiment analysis method incorporating LLMs[J]. Application Research of Computers,2024,41(10) :2900-2905.
- [33] Gao T, Fang J, Liu H, et al. LEGO-ABSA: a prompt-based task-assemblable unified generative framework for multi-task aspect-based sentiment analysis[C]//Proceedings of the 29th International Conference on Computational Linguistics,2022 :7002-7012.
- [34] Wang Z, Xia R, Yu J. Unifiedabsa: a unified absa framework based on multi-task instruction tuning [J]. arXiv preprint arXiv:2211.10986,2022.
- [35] Wang A, Jiang J, Ma Y, et al. Generative data augmentation for aspect sentiment quad prediction[C]//Proceedings of the 12th Joint Conference on Lexical and Computational Semantics, 2023 :128-140.
- [36] Pontiki M, Galanis D, Pavlopoulos J, et al. SemEval-2014 task 4: aspect based sentiment analysis [J]. SemEval, 2014 :27, doi:10.3115/v1/s14-2004.
- [37] Pontiki M, Galanis D, Papageorgiou H, et al. SemEval-2015 task 12: aspect based sentiment analysis [C]//North American Chapter of the Association for Computational Linguistics, Association for Computational Linguistics,2016, doi:10.18653/V1/S16-1002.
- [38] Pontiki M, Galanis D, Papageorgiou H, et al. Semeval-2016 task 5: Aspect based sentiment analysis [C]//International Workshop on Semantic Evaluation,2016 :19-30.
- [39] Wolf T, Debut L, Sanh V, et al. Transformers: state-of-the-art natural language processing [C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing: System Demonstrations,2020 :38-45.
- [40] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [J]. Advances in Neural Information Processing Systems, 2017,30 :5998-6008, doi:10.48550/arXiv.1706.03762.
- [41] Dai H, Song Y. Neural aspect and opinion term extraction with mined rules as weak supervision[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019 :5268-5277.
- [42] 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.
- [43] Zhang Y, Yang Y, Li Y, et al. Boundary-driven table-filling for aspect sentiment triplet extraction [C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing,2022 :6485-6498.
- [44] LIN J, LIU J H, CHEN L Y, et al. Word-pair tagging sentiment triplet extraction of fusing multi-window features [J]. Computer Engineering and Applications,2024,60(16) :159-167.
- [45] Mao Y, Shen Y, Yang J, et al. Seq2path: generating sentiment tuples as paths of a tree[C]//Findings of the Association for Computational Linguistics,2022 :2215-2225.
- 附中文参考文献:**
- [19] 王梦真. 基于自监督对比学习的文本聚类研究[D]. 北京:华北电力大学,2023.
- [23] 黄泰峰, 马静. 基于提示学习增强的文本情感分类模型[J]. 数据分析与知识发现,2024,8(3) :77-84.
- [32] 张小艳, 闫壮. 融合大语言模型的三级联合提示隐式情感分析方法[J]. 计算机应用研究,2024,41(10) :2900-2905.
- [44] 林杰, 刘建华, 陈林颖, 等. 融合多窗口特征的词对标记情感三元组抽取[J]. 计算机工程与应用,2024,60(16) :159-167.