

# 基于完形填空的方面级情感四元组预测

彭文忠<sup>1),3)</sup> 夏家莉<sup>2)</sup> 万齐智<sup>1),4)</sup> 刘德喜<sup>1),4)</sup> 万本庭<sup>2)</sup>  
曹重华<sup>2)</sup> 夏池玉<sup>3)</sup>

<sup>1)</sup>(江西财经大学信息管理学院 南昌 330032)

<sup>2)</sup>(江西财经大学软件与物联网工程学院 南昌 330032)

<sup>3)</sup>(江西财经大学财经数据科学重点实验室 南昌 330013)

<sup>4)</sup>(江西财经大学数据与知识工程江西省高校重点实验室 南昌 330013)

**摘要** 方面情感四元组预测(ASQP)任务旨在从给定的评论语句中提取所有方面词以及相应的方面类别、观点表达和情感极性,有助于全面了解用户对产品或服务不同方面的评价情况.现有情感四元组预测方法主要存在以下局限:(1)判别式模型没有利用prompt捕获情感元素之间的语义关系;(2)生成式模型要么简单地将情感元素类型标签组合形成prompt,缺乏理解标签类型涵义语境;要么将离散模板作为解码器的输入,而编码器则无法捕获到模板中情感元素之间的语义关系.为了缓解这些问题,本文首先基于完形填空思想研制离散和连续2类prompt,提供理解4个情感元素类型涵义的语境,帮助模型更好地捕获情感元素之间的语义关系;然后,基于设计的prompt,提出C-ASQP框架,包含判别式模型DC-ASQP和生成式模型GC-ASQP.在DC-ASQP中,采用2阶段策略,先预测4个情感元素中2个较为容易的情感元素,再将预测结果嵌入到设计的prompt中,帮助模型理解情感元素类型的涵义,从而有效预测另外2个情感元素.在GC-ASQP中,将设计的prompt作为编码器的输入,借助预训练模型的学习模式,充分利用预训练模型蕴含的知识提升四元组的生成效果.实验结果显示,DC-ASQP模型在4个常用数据集上的F1值相比同类判别式最优模型分别提高4.70%、6.48%、6.97%和2.60%,GC-ASQP模型的F1值比最优基准模型分别提高0.86%、1.67%、0.15%和1.02%,验证了将ASQP建模为完形填空任务的有效性,所设计的2类prompt以及C-ASQP框架是有效的.

**关键词** 方面情感四元组预测;完形填空;离散和连续prompt;判别式和生成式模型;C-ASQP框架

**中图分类号** TP18 **DOI号** 10.11897/SP.J.1016.2024.01744

## Modeling Aspect Sentiment Quad Prediction as Cloze Task

PENG Wen-Zhong<sup>1),3)</sup> XIA Jia-Li<sup>2)</sup> WAN Qi-Zhi<sup>1),4)</sup> LIU De-Xi<sup>1),4)</sup> WAN Ben-Ting<sup>2)</sup>  
CAO Zhong-Hua<sup>2)</sup> XIA Chi-Yu<sup>3)</sup>

<sup>1)</sup>(School of Information Management, Jiangxi University of Finance and Economics, Nanchang 330032)

<sup>2)</sup>(School of Software and Internet of Things Engineering, Jiangxi University of Finance and Economics, Nanchang 330032)

<sup>3)</sup>(Key Laboratory of Data Science in Finance and Economics, Jiangxi University of Finance and Economics, Nanchang 330013)

<sup>4)</sup>(Jiangxi Key Laboratory of Data and Knowledge Engineering, Jiangxi University of Finance and Economics, Nanchang 330013)

**Abstract** The aspect sentiment quad prediction (ASQP) task aims to extract all aspect terms

收稿日期:2023-05-11;在线发布日期:2024-05-13. 本课题得到国家自然科学基金项目(No. 62272206, No. 62272205, No. 62076112)、江西省主要学科学术和技术带头人培养计划领军人才项目(No. 20213BCJL22041)、江西省自然科学基金项目(No. 20212ACB202002)、江西省教育厅科学技术研究项目(No. GJJ2200560, No. GJJ2200501)、江西省社科基金项目(No. 23TQ02)、教育部人文社会科学研究项目(No. 22YJA880051)资助. 彭文忠, 博士研究生, 助理研究员, 中国计算机学会(CCF)会员, 主要研究领域为信息抽取、情感分析. E-mail: pengwenzhong@jxufe.edu.cn. 夏家莉, 博士, 教授, 博士生导师, 中国计算机学会(CCF)会员, 主要研究领域为文本挖掘和大数据处理. 万齐智, 博士, 讲师, 中国计算机学会(CCF)会员, 主要研究领域为深度学习、信息抽取、自然语言处理、数据挖掘. 刘德喜(通信作者), 博士, 教授, 博士生导师, 中国计算机学会(CCF)会员, 主要研究领域为社会媒体处理、信息检索、自然语言处理. E-mail: dexi.liu@163.com. 万本庭, 博士, 副教授, 主要研究领域为智能决策、智能计算. 曹重华, 博士, 讲师, 中国计算机学会(CCF)会员, 主要研究领域为信息抽取、机器学习. 夏池玉, 博士研究生, 助理研究员, 主要研究领域为数据挖掘和时间序列分析.

along with the corresponding aspect categories, opinion expressions, and sentiment polarities from a given review sentence, providing a holistic understanding through user evaluation of different aspects of a product or service. Existing ASQP methods suffer from the following limitations: (1) Discriminative models fail to capture the semantic relations between sentiment elements using prompts; (2) Generative models either merely combine sentiment element type labels to form prompts, lacking contextual understanding of label semantics, or use a discrete template as input to the decoder, which prevents the encoder from capturing the semantic relations between sentiment elements in the template. To alleviate these issues, this study initially develops two types of prompts, discrete and continuous prompts, based on the cloze-style methodology. These prompts provide a contextual understanding of the meanings of the four sentiment element types and aid in capturing semantic relations between sentiment elements more effectively. Discrete prompts are designed using human prior knowledge, whereas continuous prompts directly add virtual tokens between sentiment elements to represent their semantic relationships and employ prompt-tuning to enable the model to autonomously find the optimal prompt in a continuous semantic space. To enhance the model's capability to autonomously find the optimal prompt in the semantic space, this study designs suitable continuous prompts for all 24 permutations of the four sentiment element types and uses a data augmentation strategy to facilitate cooperative learning among multiple continuous prompts. Subsequently, based on the designed prompts, we propose the C-ASQP framework, which includes the discriminative model DC-ASQP and the generative model GC-ASQP. In DC-ASQP, a two-stage strategy is employed to first extract the aspect category and sentiment polarity from the review sentences, then embed the predicted aspect category and sentiment polarity into the designed prompt, helping the model to extract the corresponding aspect and opinion terms through the label semantics of the aspect category and sentiment polarity and the semantic relationships between all four sentiment elements. In GC-ASQP, the cloze-style designed prompts are concatenated to the review sentences as inputs for the encoder, leveraging the learning patterns of pretrained models to enhance the generation of aspect sentiment quadruples. Moreover, this study explores the effects on the performance of the GC-ASQP model in terms of the order of sentiment elements in the prompts, different decoding strategies, and various multi-prompt data augmentation strategies. Extensive experiments conducted on four widely used datasets show that the DC-ASQP model achieves *F1* scores improvements of 4.70%, 6.48%, 6.97%, and 2.60%, respectively, compared to the best-performing discriminative models. In comparison to the top baseline model utilizing a single prompt (template), the GC-ASQP model based on a single discrete prompt improves *F1* scores by 2.20%, 1.80%, 1.26%, and 0.96%, respectively. Utilizing a data augmentation strategy, the *F1* scores of the GC-ASQP model with 15 continuous prompts outperforms the state-of-the-art by 0.86%, 1.67%, 0.15%, and 1.02%, respectively. These results not only validate the effectiveness of modeling ASQP as cloze tasks but also prove the efficacy of the designed two types of prompts and the C-ASQP framework.

**Keywords** aspect sentiment quad prediction; cloze task; discrete and continuous prompts; discriminative and generative model; C-ASQP framework

## 1 引言

基于方面的情感分析(Asspect-Based Sentiment

Analysis, ABSA),又称为细粒度情感分析<sup>[1-2]</sup>,以预测评论文本中更丰富和细腻的情感信息为目标,具有重要的理论价值和广泛的应用价值<sup>[3-5]</sup>.目前,ABSA已经从只预测方面的情感极性发展到预测方

面情感四元组 (Aspect Sentiment Quad Prediction, ASQP)<sup>[3-4]</sup>, 即预测评论文本中的方面词 (aspect term, 简称  $a$ )、方面类别 (aspect category, 简称  $c$ )、观点词 (opinion term, 简称  $o$ ) 和情感极性 (sentiment polarity, 简称  $s$ )。例如, 评论句 “the seafood taste bad.” 中包含 1 个四元组 (food quality, seafood, bad, negative), 即方面词为 “seafood”, 方面类别为预定义类别中的 “food quality”, 观点词为 “bad”, 表达的情感为 “negative”。本文将这 4 个元素

称为情感元素。

图 1 展示了一个复杂的评论示例, 该评论文本包含 4 个方面级情感四元组。其中,  $(c_i, a_i, o_i, s_i)$  表示第  $i$  个四元组, 用不同颜色表示不同四元组, 红色表示被多个四元组共享的情感元素。由于一个语句可能包含多个情感四元组, 同一个情感元素也可能被多个四元组共享 (如图 1 中方面词 “waterfront setting” 被第 1 和第 2 个四元组共享), 因此准确识别所有情感元素并将其正确组合成四元组十分具有挑战性。

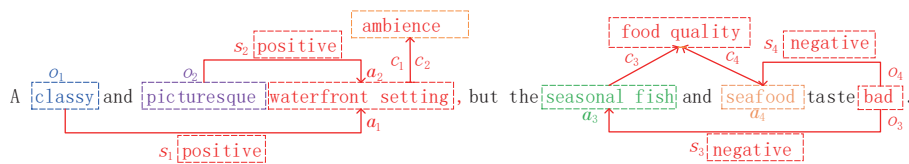


图 1 方面情感四元组预测示例

目前情感四元组预测已取得了较多成果<sup>[6-11]</sup>, 根据模型采用的策略, 主要分为判别式方法和生成式方法 2 类。

对于判别式方法, Cai 等人<sup>[6]</sup>设计了多个判别式模型来解决 ASQP 任务, 其核心是将该任务分解为多个步骤, 通过序列级或 token 级分类模块预测情感元素。然而, 上述工作中的多个判别式模型均未考虑 4 个情感元素之间的关联语义 (指 4 个情感元素在语义空间中的关联或依赖关系)。

生成式方法的现有工作<sup>[7-11]</sup>大多采用模板或 prompt 捕获情感元素之间的关联语义信息。模板和 prompt 区别在于前者作为解码器的输入, 无法采用类似预训练模型 (Pre-trained Models, PTM) 的预训练模式进行学习, 而 prompt 作为编码器的输入, 采用完形填空风格, 与 PTM 的训练方式一致。因此, prompt 作为提示可以更好地发挥 PTM 的作用。Gou 等人<sup>[11]</sup>设计了情感四元组顺序 prompt (如 “ $\langle c \rangle \langle a \rangle \langle o \rangle \langle s \rangle$ ”), 并探讨了 4 个元素之间的各种顺序给预测效果带来的影响。然而, 这种策略缺乏理解情感元素含义的语境, 使得模型不能更好地理解 prompt 中元素的含义。部分学者<sup>[7-9]</sup>将四元组的标签作为文本输入解码器, 生成指定格式的情感四元组。在它们基础上, Zhang 等人<sup>[10]</sup>设计了离散模板 (指利用部分具体的词语连接 4 个情感元素, 使其表达成可理解的语句), 使得模型可以更好地理解情感元素的语义。然而, 这些工作均将模板作为解码器的输入, 导致模型在编码器中无法获取模板中的信息; 同时, 由于下游任务与 PTM 的预训练任务不一致, 难以有效利用 PTM 中丰富的百科全书式知识<sup>[12]</sup>。此

外, 离散模板因给出了具体词语, 限制了理解情感元素之间的丰富情境。

受上述研究启发, 本文首先设计了符合理解情感四元组含义的离散和连续<sup>[13]</sup>两类 prompt, 旨在捕获情感元素之间的关联语义, 其优势体现在三个方面。

(1) 离散 prompt 能让模型利用人类先验知识来理解、捕获情感元素之间的关联关系。例如, 生成式离散 prompt “The  $\langle c \rangle$  such as  $\langle a \rangle$  is  $\langle o \rangle$ , so I feel  $\langle s \rangle$ ” 蕴含了方面词从属于方面类别、观点词用于修饰方面目标以及语义情感极性判断与情感描述之间的一致性等信息。“such as” 反映了方面类别  $\langle c \rangle$  和方面词  $\langle a \rangle$  的从属关系; “is” 揭示了观点词  $\langle o \rangle$  用于修饰方面词  $\langle a \rangle$  及方面类别  $\langle c \rangle$ ; “so” 表达了前后两个分句语义要一致, 即前面表达的情感要与情感极性  $\langle s \rangle$  保持一致。整个离散 prompt 通过这些具体的词语创建了利于理解情感元素类型含义的语境, 使得既揭示了情感元素的关联语义, 又反映了情感元素类型的语义。

(2) 连续 prompt 使得模型具有在连续语义空间中寻找最优 prompt 的能力。连续 prompt 通过添加虚拟 token 的方式来描述不同情感元素类型之间的关系, 构建了 4 类情感元素的关联语境。虚拟 token 不受具体词语限制, 其语义在模型训练阶段通过反向传播学习<sup>[14-15]</sup>, 具有更好的泛化能力。例如,  $c \rightarrow a \rightarrow o \rightarrow s$  (简称为  $caos$ ) 顺序对应的连续 prompt 为 “ $\langle c \rangle [CA_1] \dots [CA_{n_1}] \langle a \rangle [AO_1] \dots [AO_{n_2}] \langle o \rangle [OS_1] \dots [OS_{n_3}] \langle s \rangle$ ”。其中,  $[\cdot]$  为虚拟

token,  $n_i$  为虚拟 token 个数。

(3) 通过多个连续 prompt 可以强化 prompt 中虚拟 token 学习元素之间关联语义的效果。连续 prompt 中虚拟 token 在具有模糊性的同时也蕴含了更多的灵活性, 即其具有在整个语义空间中寻找最佳 prompt 的潜力, 而不充分的训练会限制虚拟 token 的学习效果。因此, 本文采用数据增强策略, 从不同顺序的连续 prompt 中协同学习元素之间的关联语义, 4 个元素两两组合共有 12 种关系, 即  $\{ca, ao, os, co, cs, as, ac, oa, so, oc, sc, sa\}$ , 描述这 12 种关系的虚拟 token 个数分别设为  $n_1, n_2, \dots, n_{12}$ 。设计连续 prompt 的核心是不同顺序 prompt 中用来描述相同 2 个情感元素语义关系的虚拟 token 应保持一致。例如, 在 *aocs, aosc, caos, saoc, csao* 和 *scao* 这 6 种 prompt 中都使用 “[ $AO_1$ ]  $\dots$  [ $AO_{n_2}$ ]” 来描述  $a \rightarrow o$  顺序下方面词类型  $\langle a \rangle$  和观点词类型  $\langle o \rangle$  的关联关系。这样, 在多 prompt 数据增强时, 如果这 6 种顺序的 prompt 中有多个被选中, 则能够从多视图更好地学习  $a \rightarrow o$  顺序下方面词  $\langle a \rangle$  和观点词  $\langle o \rangle$  之间的语义。

然后, 将设计的 prompt 分别嵌入判别式和生成式模型中, 借鉴完形填空思想提出了 2 种情感四元组预测模型, 即判别式 ASQP 模型 (Discriminative Cloze-style ASQP model, DC-ASQP) 和生成式 ASQP 模型 (Generative Cloze-style ASQP model, GC-ASQP)。

在 DC-ASQP 中, 采用 2 阶段策略。首先, 获取源语句中的  $(c, s)$  对, 再将  $(c, s)$  对作为已知信息指导  $(a, o)$  的抽取, 充分发挥情感各元素的作用 (方面类别  $c$  和情感极性  $s$  的识别相较于方面词  $a$  和观点词  $o$  稍显容易), 降低了同时预测四个元素的难度。此外,  $(c, s)$  对也用于构建第 2 阶段 4 个元素语义相关的 prompt, 这样的预测策略能够让模型借助  $(c, s)$  标签语义及元素之间关联语义来抽取对应的  $(a, o)$ , 进而获取完整的四元组。

在 GC-ASQP 中, 将设计的 prompt 拼接于输入语句后, 作为生成模型编码器的输入, 这样使得模型在编码器中能了解情感元素的格式和理解标签语义的语境, 从而有效地编码输入语句, 并生成符合格式的情感四元组。

本文的主要贡献包括以下 3 点:

(1) 为了捕获情感元素之间的关联语义, 设计了离散和连续 2 类 prompt。离散 prompt 使模型能够借

助人类先验知识来理解、捕获情感元素之间的关联语义; 连续 prompt 使得模型可以在连续语义空间中寻找最优 prompt。

(2) 基于设计的 2 类 prompt, 提出了判别式模型 DC-ASQP 和生成式模型 GC-ASQP。在 DC-ASQP 中, 先获取  $(c, s)$  对并将  $(c, s)$  对作为已知信息构建 4 个元素语义相关的 prompt, 使模型能够借助  $(c, s)$  标签语义及元素之间关联关系来抽取对应的  $(a, o)$ 。在 GC-ASQP 中, 将 prompt 拼接在输入语句后作为编码器输入, 使模型在编码器中能了解情感元素的格式和理解标签语义的语境。同时, 探讨了所设计的不同 prompt、解码器中不同解码策略和不同的多 prompt 数据增强策略在情感四元组预测上的效果。

(3) 在多个数据集上构造了丰富的实验, 实验结果验证了本文设计的离散和连续 prompt 的价值。在 4 个常用的公开数据集上, 判别式模型 DC-ASQP 的  $F1$  值比判别式最好基线分别提升 4.70%、6.48%、6.97% 和 2.60%。对比单 prompt (模板) 的最优基准模型, 基于单离散 prompt 的生成式模型 GC-ASQP 的  $F1$  值分别提升 2.20%、1.80%、1.26% 和 0.96%; 在多 prompt 数据增强场景中, 15 连续 prompt 数据增强的 GC-ASQP 与最优基准 MvP 相比,  $F1$  值分别提升 0.86%、1.67%、0.15% 和 1.02%。

## 2 相关工作

### 2.1 方面情感四元组

早期, 情感分析主要处理方面级情感分类<sup>[16-37]</sup>, 旨在分析已知方面词对应的情感极性  $s$ 。由于很多场景无法事先获知方面词, 文献<sup>[38-45]</sup>考虑抽取评论语句中所有的方面词  $a$ 。随后, 包含方面词及其情感极性的方面情感二元组预测任务<sup>[46-50]</sup>被关注, 紧接着面向方面的观点词抽取任务比较流行<sup>[51-54]</sup>, 用于提取给定方面词对应的观点词  $o$ 。早几年, 方面情感三元组 (方面词、观点词、情感极性) 抽取 (Aspect Sentiment Triplet Extraction, ASTE) 任务<sup>[7-8, 55-65]</sup>吸引了较多学者。由于情感四元组 (包括方面类别、方面词、观点词、情感极性) 预测任务无需给定方面类别或方面词且预测结果更全面, 具有更广泛的应用场景。因此, 近些年, 少量研究者开始关注四元组 ASQP 预测任务。

ASQP 是 ABSA 最新子任务, 现有研究<sup>[6-11]</sup>根据

模型采用的策略,主要分为判别式方法和生成式方法2类.前者以流水线框架将ASQP建模为序列级分类或token级分类(即序列标注)任务,后者将ASQP建模为序列生成任务.

判别式ASQP模型通常基于BERT<sup>[66]</sup>获取上下文相关的文本特征表示,再与精心设计的面向特定任务的神经网络结合.文献[6]是该类模型的代表工作.Cai等人<sup>[6]</sup>基于流水线架构设计了多个基准模型,包括Double-Propagation-ACOS、JET-ACOS、TAS-BERT-ACOS和Extract-Classify-ACOS.

Double-Propagation-ACOS模型基于句法规则<sup>[67]</sup>提取语句中的情感三元组 $(a, o, s)$ ,然后基于规则为每个 $(a, o, s)$ 三元组分配方面类别.由于基于规则的方法难以理解语句的复杂语义和上下文,所以Cai等人<sup>[6]</sup>进一步设计了基于PTM的基准模型.例如,JET-ACOS模型首先基于具有位置感知的拓展序列标注方案提取 $(a, o, s)$ 三元组<sup>[59]</sup>,再为每个三元组预测其方面类别以得到四元组,但该模型无法识别隐式情感元素.

TAS-BERT-ACOS模型采用输入转换策略<sup>[68]</sup>识别语句中的有效 $(c, s)$ ,并在此条件下以序列标注的方式提取对应的方面词和观点词.Extract-Classify-ACOS模型使用拓展序列标注方案提取方面词和观点词<sup>[69]</sup>,以笛卡尔积方式得到候选 $(a, o)$ ,再通过分类任务将 $(a, o)$ 与方面类别 $c$ 和情感极性 $s$ 进行关联并过滤掉无效的四元组.为了识别隐式情感元素,它们通过额外的二分类来判断语句中是否存在隐式方面词或隐式观点词.

上述模型中Extract-Classify-ACOS性能最优<sup>[6]</sup>,但它将情感元素标签视为分类或序列标注的目标,没有显式地利用情感元素标签的语义信息.同时,以笛卡尔积作为情感元素关联的核心方法,忽略了4个情感元素之间的关联语义.

生成式ASQP模型<sup>[7-11,70]</sup>采用从输入序列生成输出序列(sequence to sequence, seq2seq)的方式来解决ASQP任务.他们大多数以情感元素所在的源语句作为模型输入,通过设计合理的目标序列(即作为解码器输入来指导模型生成的序列)来指导模型生成情感四元组.具体地,Yan等人<sup>[55]</sup>将对应情感元素的指针索引和分类索引组成的序列作为目标序列.然而,该方法将标签视为数字索引,忽略了标签的语义信息.为了利用标签语义信息,Zhang等人<sup>[7]</sup>将四元组标签的文本序列作为模型的目标序列.在此基础上,Mao等人<sup>[8]</sup>在解码器中采用束搜索替代

了传统的贪心解码策略;Hu等人<sup>[9]</sup>探讨了模板中4个元素不同排列顺序给预测效果带来的影响,并通过基于不同顺序模板的数据增强提升模型预测效果;Zhang等人<sup>[10]</sup>设计了离散模板将四元组标签转述为自然语句,使得模型可以更好地理解情感元素的语义.Bao等人<sup>[70]</sup>则将方面情感四元组转换为观点树结构,以线性化观点树作为目标序列,并让模型在大量相近领域数据集上学习从原始语句生成其线性化选区解析树及线性化抽象意义表示解析树的能力,整个模型的训练成本巨大.然而,这些工作均将模板作为解码器的输入,导致模型在编码器中无法获取模板中的信息;同时,由于下游任务与PTM的训练过程不一致,难以有效利用PTM中丰富的百科全书式知识<sup>[12]</sup>.

鉴于提示调优(prompt tuning)方法在NLP领域的许多应用<sup>[12-13,46,71-84]</sup>中均获得显著成功.Gao等人<sup>[85]</sup>将prompt引入ABSA领域,并研究ABSA领域中从简单任务prompt到复杂任务prompt的迁移.Gou等人<sup>[11]</sup>将4个情感元素按不同顺序排列组合得到24种顺序prompt,再基于多个不同的顺序prompt进行数据增强,最后通过投票策略对不同顺序prompt的预测结果进行聚合,取得了显著成功.然而,他们设计的prompt缺乏理解情感元素含义的语境,因此本文旨在设计能够理解情感元素语境的prompt,帮助模型更好地捕获情感元素之间的关联关系及语义,进而提升情感四元组预测效果.

## 2.2 在ABSA中显式建模情感元素之间关联关系

如何对方面情感元素之间的关联关系进行建模是ABSA各类子任务的研究热点<sup>[3-4]</sup>.本文将对情感元素关联语义建模的方式归纳为以下几种:

### (1) 基于依存句法树的方法

该类方法通常借助图卷积网络编码依存句法树信息,旨在增强语句中方面词和观点词的依赖关系<sup>[23,50,56-57]</sup>,丰富PTM的token嵌入表示.刘欣逸等人<sup>[57]</sup>提出了句法增强的ASTE模型,将ASTE分解为方面词抽取、观点词抽取和情感极性分析3个子任务.为了缓解误差传播,Chen等人<sup>[56]</sup>提出了基于网格标注方案<sup>[60]</sup>的ASTE模型,使用多通道图卷积网络编码依存句法信息等多种语言学特征.然而,这类方法主要对语句中的方面词和观点词之间的关系进行建模,未显式考虑方面类别和情感极性之间的相关性.

### (2) 基于阅读理解的方法

部分研究者<sup>[61-62,86-87]</sup>将方面情感多元组抽取建

模为阅读理解任务,以多轮问答方式逐步获取情感元素;通过设计问题反映情感元素之间的关联语义,使得后面轮次可以将已获取的情感元素标签作为先验知识来设计约束性问题.该方法的缺点是随着元组中元素个数增多,需要问答的轮次也越多,存在误差传播风险.

### (3) 基于模板的方法

部分生成式模型<sup>[10]</sup>通过设计离散模板将元组转述为自然语句后作为解码器输入,通过微调模型来学习情感元素之间的关联关系.然而,由于微调任务与PTM预训练方式不一致,这使得仅通过微调模型难以有效捕捉情感元素之间的关联语义信息<sup>[70,88]</sup>.

### (4) 构建观点树的方法

部分研究者<sup>[70]</sup>通过直接构建结构化观点树对情感元素之间的关联关系进行建模. Bao等人<sup>[70]</sup>将方面情感四元组转换为观点树结构,但该策略的缺点是直接将源语句映射为复杂的目标观点树结构是困难的,而构建额外的辅助自监督任务又极大地增加了模型的训练成本.

这几类显式对情感元素之间关联关系建模的方法在ABSA领域取得了不同程度的成功,但均未采用完形填空风格的prompt来关联情感元素,未能充分利用PTM中丰富的百科全书式知识.

## 2.3 完形填空建模

完形填空建模,即将下游任务设计为与PTM预训练过程相似的完形填空任务,可以减少预训练与任务微调之间的差距.通过合适的prompt,能够帮助PTM从广博的预训练知识中引导出与下游任务相关的知识<sup>[89]</sup>,提升下游任务的性能. Petroni等人<sup>[82]</sup>将问答中的问题改写为完形填空任务形式(答案为待填空项)来探索PTM中的知识.在不进行微调的情况下,BERT在开放域问答任务上达到了与利用监督知识训练出来的传统模型相当的性能. Tom等人<sup>[83]</sup>提出GPT-3模型,将超大规模模型与prompt相结合,在零样本和少样本场景下,GPT-3在多个NLP任务上达到有竞争力的性能,在部分任务上超越了最先进的有监督模型.

提示调优是受GPT-3模型<sup>[83]</sup>启发的一种新的微调范式<sup>[89]</sup>,也是完形填空建模思想最成功的应用方式之一.例如,基于小规模PTM(如BERT), Schick等人<sup>[73]</sup>提出PET模型(一种半监督的训练方法),使用自然语言模式将所有分类任务转换为完形填空任务. PET模型在少样本场景下具有显著优

势,标准场景下在多个数据集上的性能相比传统方法也略有提升.受上述研究启发,较多研究者开始关注提示调优,并在多个任务上取得了显著成功,包括文本分类<sup>[13,71-74]</sup>、命名实体识别<sup>[75-76]</sup>、事件论元抽取<sup>[77]</sup>、实体类型标记<sup>[78]</sup>、实体关系抽取<sup>[79-81]</sup>、方面级情感分析<sup>[11,46,85]</sup>等.

由于人工无法穷举所有可能的离散prompt,部分学者提出使用连续prompt<sup>[14-15]</sup>和混合prompt<sup>[14]</sup>,引入可学习的虚拟token,期望模型在训练阶段通过微调及反向传播自动在全局空间搜索最优prompt.

## 3 ASQP任务定义及建模思想

### 3.1 ASQP任务定义

ASQP任务旨在预测源语句 $X$ 中所有方面情感四元组: $Q_1, Q_2, \dots, Q_i, \dots$ . 其中, $Q_i = (c_i, a_i, o_i, s_i)$ , 方面类别 $c_i$ 属于预定义的类别集合 $V_c$ ,即 $c_i \in V_c$ ;方面词 $a_i$ 和观点词 $o_i$ 是 $X$ 中的文本跨度,允许为空,即 $a_i \in V_X^{\text{span}} \cup \{\emptyset\}$ ,  $o_i \in V_X^{\text{span}} \cup \{\emptyset\}$ ,  $V_X^{\text{span}}$ 为 $X$ 中所有文本跨度的集合.情感极性 $s_i \in V_s$ ,  $V_s$ 为预定义情感极性标签集合,包括“positive”(正面)、“negative”(负面)和“neutral”(中性).

### 3.2 ASQP任务建模策略

本文借鉴完形填空思想设计了2种情感四元组预测模型,即判别式DC-ASQP模型和生成式GC-ASQP模型.本节通过图2中一个具体案例来展示这2种不同模型预测四元组的具体过程.

同时预测多个方面四元组的所有待空白极具挑战,因此,DC-ASQP采用2阶段策略.如图2(左)所示.第1阶段,通过“方面类别-情感极性”分类模块获取源语句中的 $(c, s)$ 对,如源语句“A classy and picturesque waterfront setting, but the seasonal fish and seafood taste bad.”中包含2个 $(c, s)$ 对,分别为 $(ambience, positive)$ 和 $(food\ quality, negative)$ .第2阶段,将获取的 $(c, s)$ 对作为已知信息构建能反映4个元素关联关系的prompt.例如,基于 $(food\ quality, negative)$ 对构建的离散prompt为“The food quality such as [MASK] is [MASK], so I feel negative”.其中,2个[MASK]分别为方面词和观点词的预留空白.这样的预测策略能够让第2阶段“方面词-观点词”完形填空模块借助prompt中 $(c, s)$ 标签文本语义及4个元素之间关联信息来抽取对应的 $(a, o)$ .该案例中 $(food\ quality, negative)$ 对应的

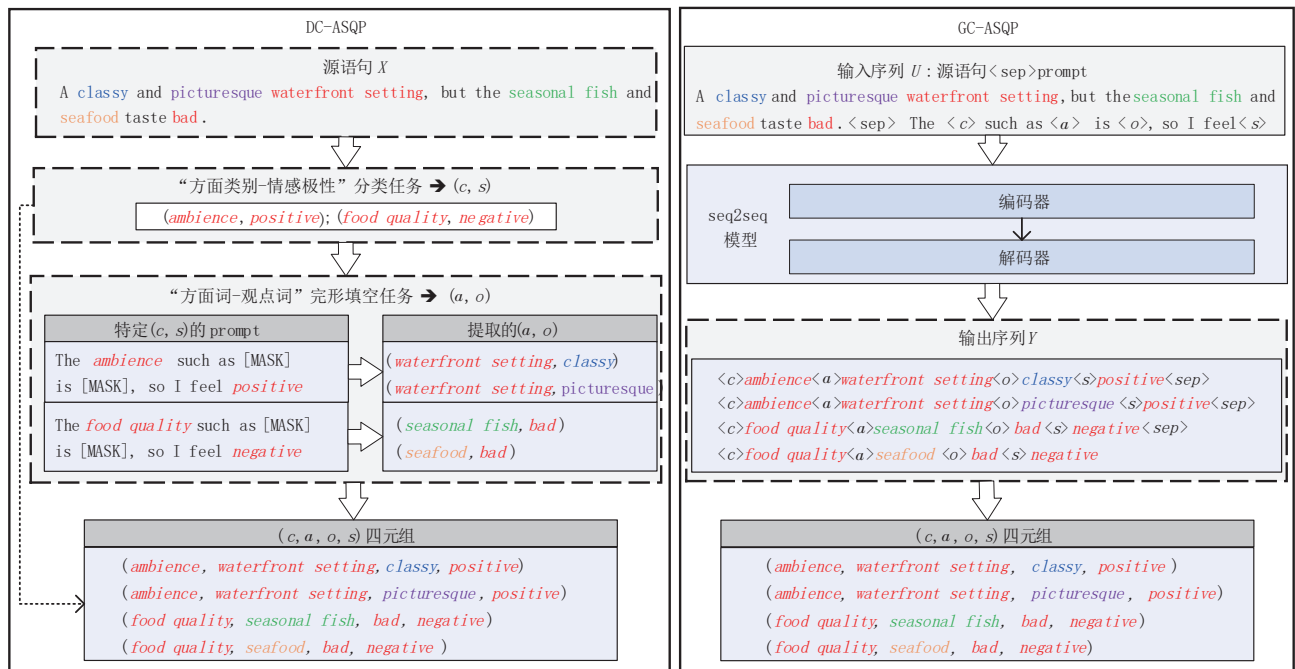


图2 判别式和生成式ASQP模型完形填空任务建模方案示例

$(a, o)$ 有2个,即 $(seasonal\ fish, bad)$ 和 $(seafood, bad)$ ;类似地, $(ambience, positive)$ 对应的 $(a, o)$ 也有2个,即 $(waterfront\ setting, classy)$ 和 $(waterfront\ setting, picturesque)$ 。因此,该评论句中中共包含4个四元组。

GC-ASQP采用4个不同的哨兵token<sup>[90]</sup> $\langle c \rangle$ 、 $\langle a \rangle$ 、 $\langle o \rangle$ 和 $\langle s \rangle$ 分别表示4种情感元素类型,通过离散prompt来反映4类情感元素之间的关联语义,以特殊token $\langle sep \rangle$ 将语句和prompt拼接后作为编码器的输入。GC-ASQP以自回归方式逐token生成情感元素,生成每个元素时能看到完整的prompt和已生成的元素信息,模型根据prompt语义约束来指导后续情感元素的生成,从而完成同一四元组中四个情感元素的组合以及多个四元组的生成任务。

## 4 基于完形填空的C-ASQP框架

### 4.1 判别式完形填空模型DC-ASQP

判别式完形填空模型DC-ASQP主要包括3个模块,如图3所示。

(1) 编码模块。基于BERT的跨度级编码模块旨在编码源语句中的候选跨度,得到各候选跨度的嵌入表示。

(2) 分类模块。“方面类型-情感极性”分类模块采用并发方式预测给定语句中所有可能的“方面类

型-情感极性” $(c, s)$ 。

(3) 填空模块。“方面词-观点词”完形填空模块首先构建特定 $(c, s)$ 的prompt,然后以完形填空的方式预测prompt中为方面词和观点词预留的空白(即[MASK])。

#### 4.1.1 编码模块

采用BERT<sup>[66]</sup>对源语句 $X = (x_1, x_2, \dots, x_n)$ 进行编码,得到所有tokens的表示 $H^w = \{h_1^w, h_2^w, \dots, h_n^w\} \in \mathbb{R}^{n \times d}$ 和特殊token[CLS]的表示 $h_{cls}^w$ 。为了处理隐式方面词及隐式观点词,设置1个特殊token[NULL],表示为 $h_{NULL}^w$ 。穷举 $X$ 中所有候选跨度,得到 $V_X^{span} = (span_1, span_2, \dots, span_{|V_X^{span}|})$ ,其中 $span_i \in V_X^{span}$ 的表示 $h_{span_i}^w$ 为

$$h_{span_i}^w = FNN([h_{START(span_i)}^w; h_{END(span_i)}^w; f_{width}(span_i)]) \quad (1)$$

其中, $h_{START(span_i)}^w$ ;  $h_{END(span_i)}^w$ 分别为跨度首、尾tokens的表示, $f_{width}(span_i)$ 为跨度长度的嵌入表示<sup>[63]</sup>,”表示向量拼接。注意,采用跨度级编码得到的跨度表示是相互独立的,所以能够表示具有重叠tokens的不同跨度。

所有候选跨度表示记为 $H_{span} = \mathbb{R}^{|V_X^{span}| \times d}$ , $|V_X^{span}|$ 表示候选跨度数量,通过前馈神经网络(Feed-forward Neural Network, FNN)使跨度表示的维度与token表示的维度一致,均为 $d$ 。

再经过2个独立的FNN对 $H_{span}$ 作如下处理:

$$\begin{aligned} H_{span}^{cs} &= H_{span} W_{span}^{cs} + b_{span}^{cs} \\ H_{span}^{ao} &= H_{span} W_{span}^{ao} + b_{span}^{ao} \end{aligned} \quad (2)$$

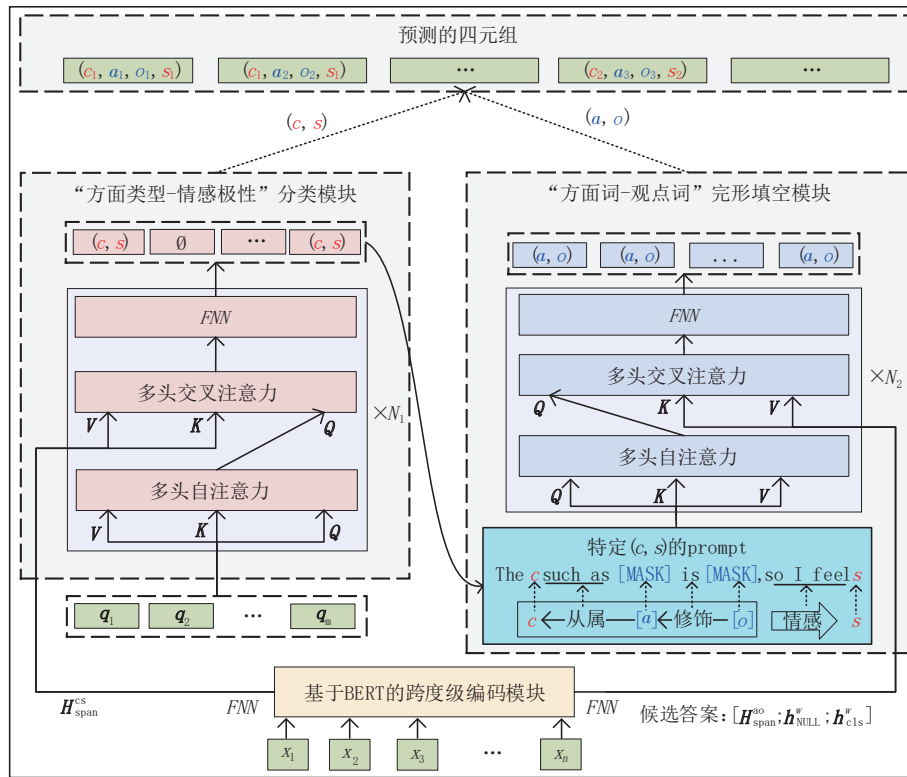


图3 DC-ASQP模型架构图

其中,  $W_{span}^{cs}, W_{span}^{ao} \in \mathbb{R}^{d \times d}$  为权重,  $b_{span}^{cs}, b_{span}^{ao} \in \mathbb{R}^d$  为偏置项, 均为可训练参数.  $H_{span}^{cs}, H_{span}^{ao} \in \mathbb{R}^{|V_{span}^{cs}| \times d}$  将分别用于“方面类型-情感极性”分类模块和“方面词-观点词”完形填空模块.

#### 4.1.2 “方面类型-情感极性”分类模块

“方面类型-情感极性”提取被表述为集合预测任务, 即预测  $X$  中存在的  $(c, s)$  集合, 记作  $V_X^{cs}$ ,  $V_X^{cs} \subset V_{cs}$ ,  $V_{cs} = (V_c \times V_s) \cup \{\emptyset\}$ . 与 Transformer<sup>[91]</sup> 非自回归解码器类似, 分类模块堆叠  $N_1$  个解码块, 每个解码块由多头自注意力层、多头交叉注意力层和输出层组成.

**多头自注意力层.** “多头自注意力层”的输入为  $m$  个  $(c, s)$ , 将随机初始化的向量作为其初始嵌入表示, 记为  $Q^{cs} \in \mathbb{R}^{m \times d}$ . 其中,  $m$  通常设置为略大于所有源语句  $X$  中不同  $(c, s)$  数量的最大值.

多头缩放点积注意力<sup>[91]</sup>计算公式为  $MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W^O$   
 $head_i = Attention(QW_i^Q, KW_i^K, VW_i^V)$  (3)

其中,  $Q, K, V$  均为  $Q^{cs}$ ;  $W_i^Q, W_i^K, W_i^V \in \mathbb{R}^{d \times d}$ ,  $W^O \in \mathbb{R}^{hd \times d}$ ,  $h$  为并发头数. “多头自注意力层”输出为  $\tilde{H}^{cs} = MultiHead(Q^{cs}, Q^{cs}, Q^{cs}) \in \mathbb{R}^{m \times d}$ .

**多头交叉注意力层.** “多头交叉注意力层”基于公式(3)计算注意力, 其中  $Q = \tilde{H}^{cs}$ ,  $K = V = H_{span}^{cs}$ , 该层让  $H_{span}^{cs}$  中的候选跨度以不同的权重融入入到  $(c, s)$  的嵌入表示  $\tilde{H}^{cs}$  中. “多头交叉注意力层”的输出为  $H^{cs} = MultiHead(\tilde{H}^{cs}, H_{span}^{cs}, H_{span}^{cs}) \in \mathbb{R}^{m \times d}$ .

**输出层.** 通过 FNN,  $H^{cs}$  中第  $i \in [1, m]$  个嵌入表示  $h_i^{cs} \in \mathbb{R}^d$  在所有  $(c, s)$  上的概率分布为

$$P_i^{cs} = Softmax(W^{cs} h_i^{cs} + b^{cs}) \quad (4)$$

其中,  $W^{cs} \in \mathbb{R}^{|V_{cs}| \times d}$ ,  $b^{cs} \in \mathbb{R}^{|V_{cs}|}$  分别为权重和偏置项,  $|V_{cs}|$  为  $V_{cs} = (V_c \times V_s) \cup \{\emptyset\}$  中元素总数量. 如果  $P_i^{cs}$  中概率最高的  $(c, s)$  不为  $\emptyset$  类别, 则将其加入到源语句  $X$  的  $(c, s)$  集合  $V_X^{cs}$  中.

**损失函数.** “方面类别-情感极性”分类模块以并发的方式预测  $X$  对应的  $(c, s)$  集合  $V_X^{cs}$ , 对  $V_X^{cs}$  中元素排序不作要求. 因此, 损失函数采用对顺序不敏感的二部图匹配损失<sup>[92]</sup>. 预测结果和标准答案之间的最佳匹配即为具有最小损失的排列策略  $\pi^*$ :

$$\pi^* = argmin_{\pi \in \Pi(m)} \left( - \sum_{i=1}^m I(y_i^{cs}) \cdot P_{(\pi(i))}^{cs}(y_i^{cs}) \right) \quad (5)$$

其中,  $\Pi(m)$  为  $(c, s)$  集合  $V_X^{cs}$  中全部元素所有可能的排列策略,  $\pi$  为其中一种排列策略.  $y_i^{cs}$  表示标准答案中第  $i$  个元素, 记为  $(c_i, s_i)$ .  $I(\cdot)$  为指示函数, 如果



$y_i^{cs} \neq \emptyset$ , 则  $I(y_i^{cs})=1$ , 否则  $I(y_i^{cs})=0$ .  $P_{(\pi(i))}^{cs}$  表示在第  $\pi$  种排列策略下第  $i$  个位置上预测结果为  $(c_i, s_i)$  的概率. 通过匈牙利算法<sup>[92]</sup>可以在多项式 ( $O(m^3)$ ) 时间内计算出最优排列策略  $\pi^*$ , 进而得到二部图匹配损失  $\mathcal{L}^{cs}$  为

$$\mathcal{L}^{cs} = - \sum_{i=1}^m \log P_{(\pi^*(i))}^{cs}(y_i^{cs}) \quad (6)$$

#### 4.1.3 “方面词-观点词”完形填空模块

**Prompt 构建.** 为  $V_X^{cs}$  中的每个  $(c, s)$  构建一个 prompt. Prompt 构建规则如下:

- (1) Prompt 包含方面类别  $c$  和情感极性  $s$  的标签文本;
- (2) Prompt 中预留两个 [MASK] (即待填空项), 对应为方面词待空白和观点词待空白;
- (3) Prompt 应能反映 4 个情感元素之间的关联语义.

如图 2 左部分所示, 源语句  $X$  对应的  $(c, s)$  集合  $V_X^{cs} = \{(ambience, positive), (food\ quality, negative)\}$ . 以 prompt 格式 “The  $c$  such as [MASK] is [MASK], so I feel  $s$ ” 为例, 特定  $(food\ quality, negative)$  的 prompt 为: “The  $food\ quality$  such as [MASK] is [MASK], so I feel  $negative$ ”. Prompt 中 2 个 [MASK] 需要分别填入合适的方面词和观点词. 示例中  $(food\ quality, negative)$  有 2 个  $(a, o)$ , 分别为:  $(seasonal\ fish, bad)$ 、 $(seafood, bad)$ . 为了以并发方式一次性抽取与  $(c, s)$  关联的所有  $(a, o)$ , 我们将 prompt 复制  $k$  次并以 [SEP] 拼接, 记为  $T^{ao} = (t_1, t_2, \dots, t_{|T^{ao}|})$ . 将  $T^{ao}$  输入填空模块后以非自回归的方式同时对  $T^{ao}$  中的  $2k$  个 [MASK] 空白进行填空.  $k$  一般设置为大于样本中一个  $(c, s)$  对应  $(a, o)$  的典型数量.

与分类模块类似, 填空模块堆叠  $N_2$  个相同解码块, 每个解码块由多头自注意力层、多头交叉注意力层和输出层组成.

**多头自注意力层.** 填空模块通过自注意力机制让 [MASK] 理解其所在 prompt 的上下文语境, 使 [MASK] 嵌入表示中蕴涵 prompt 中方面类别和情感极性标签语义及 4 个情感元素之间的关联语义等信息. 查询 BERT 词库可获取序列  $T^{ao}$  的初始嵌入表示, 记为  $Q^{ao}$ . 将  $Q^{ao}$  作为该层输入, 由公式 (3) 得到输出为  $\tilde{H}^{ao} = MultiHead(Q^{ao}, Q^{ao}, Q^{ao}) \in \mathbb{R}^{|T^{ao}| \times d}$ .

**多头交叉注意力层.** 基于交叉注意力机制,  $\bar{H} = [H_{span}^{ao}; h_{NULL}^w; h_{cls}^w] \in \mathbb{R}^{(|V_X^{span}|+2) \times d}$  中  $|V_X^{span}|+2$  个

候选跨度会以不同权重融入到  $\tilde{H}^{ao}$  的每个 token 表示中. 由公式 (3) 得到该层的输出为  $H^{ao} = MultiHead(\tilde{H}^{ao}, \bar{H}, \bar{H}), H^{ao} \in \mathbb{R}^{|T^{ao}| \times d}$ . 将  $H^{ao}$  中  $2k$  个 [MASK] 的嵌入表示记为  $H_{blk}^{ao} \in \mathbb{R}^{2k \times d}$ .

**输出层.** 填空模块的目标是分别从候选跨度中选择最合适的跨度作为答案填入  $T^{ao}$  的  $2k$  个 [MASK] 中.  $T^{ao}$  中第  $i$  个 [MASK] 对应的候选跨度 (即候选答案) 嵌入表示如下:

$$\tilde{h}_{blk_i}^{ao} = \tanh(\bar{H}W_{blk}^{ao,1} + W_{blk}^{ao,2}h_{blk_i}^{ao} + b_{blk}^{ao}) \quad (7)$$

其中,  $\bar{H}$  为初始候选跨度表示;  $h_{blk_i}^{ao} \in \mathbb{R}^d$  是  $H_{blk}^{ao}$  中第  $i$  个 [MASK] 的嵌入表示;  $W_{blk}^{ao,1}, W_{blk}^{ao,2} \in \mathbb{R}^{d \times d}$  是可训练的权重,  $b_{blk}^{ao} \in \mathbb{R}^d$  是可训练的偏置项.

第  $i$  个 [MASK] 填入候选跨度的概率分布  $P_{blk_i}^{ao}$  如公式 (8) 所示.

$$P_{blk_i}^{ao} = Softmax(\tilde{h}_{blk_i}^{ao}, \mathbf{u}_{blk}^{ao}) \quad (8)$$

其中,  $\mathbf{u}_{blk}^{ao} \in \mathbb{R}^d$  是可训练权重. 取  $P_{blk_i}^{ao} \in \mathbb{R}^{(|V_X^{span}|+2)}$  中概率最高的跨度作为填入第  $i$  个 [MASK] 的答案. 如果答案为 [NULL] 则填入 “NULL”, 表明对应情感元素 (即方面词或观点词) 是隐式的; 如果答案为 [CLS], 表明未能找到适合填入该 [MASK] 的情感元素, 对应四元组不作为有效的预测结果.

采用交叉熵损失作为填空模块的损失  $\mathcal{L}^{ao}$ :

$$\mathcal{L}^{ao} = - \sum_{i=1}^m \sum_{j=1}^{2k} \log P_{i, blk_j}^{ao}(y_{i, blk_j}^{ao}) \quad (9)$$

其中,  $y_{i, blk_j}^{ao}$  是第  $i$  个  $(c, s)$  对应  $T^{ao}$  中第  $j$  个 [MASK] 的答案, 即正确的方面词或观点词.

#### 4.1.4 联合训练

为了减少误差传播, 本文以多任务联合学习方式同时训练分类模块和填空模块, 并共享编码层的参数. 联合训练总损失  $\mathcal{L}^{total}$  为

$$\mathcal{L}^{total} = \lambda \mathcal{L}^{cs} + (1 - \lambda) \mathcal{L}^{ao} \quad (10)$$

其中, 参数  $\lambda$  用于调节 2 个子任务损失的比重.

## 4.2 生成式完形填空模型 GC-ASQP

GC-ASQP 将 ASQP 建模为 seq2seq 序列生成任务, 在输入序列中增加含有 4 个空白占位符 (即 4 个哨兵 token<sup>[90]</sup>, 对应 4 个情感元素类别) 的 prompt, 并将其作为提示, 让生成式 PTM 通过 prompt 理解 4 个情感元素之间的关联关系后以自回归方式逐 token 生成源语句  $X$  中所有方面情感四元组. 下面从完形填空任务建模和模型训练等方面对 GC-ASQP 模型进行介绍.

#### 4.2.1 完形填空任务建模

GC-ASQP将源语句和prompt(即完形填空的题目)拼接后作为模型输入,完形填空答案由4类空白占位符及其对应答案组成.

**Prompt构建.**将输入序列中prompt记为 $T$ ,其构建应遵循以下原则:

(1)Prompt中应包含4类情感元素对应空白占位符,且4类空白占位符在prompt中的顺序可调整;

(2)Prompt应尽可能反映4类情感元素之间的语义关系.

将prompt $T$ 拼接在源语句 $X$ 后作为模型输入序列,记为 $U$ .

**目标序列构建.**源语句 $X$ 中可能存在多个四元组,第 $i$ 个四元组 $(c_i, a_i, o_i, s_i)$ 对应输出序列为“ $\langle c \rangle c_i \langle a \rangle a_i \langle o \rangle o_i \langle s \rangle s_i$ ”,不同四元组以 $\langle \text{sep} \rangle$ 进行拼接后作为目标序列 $Y$ .注意,目标序列中4类空白占位符的顺序与prompt中出现这4类空白占位符的顺序应一致.

部分学者<sup>[10,93]</sup>将情感极性的答案空间映射为[“great”“ok”“bad”],考虑到“ok”在评论语句中很多时候会用来表达负面情绪(如“The dinner was ok, nothing I would have again.”),本文直接使用原始标签[“positive”“neutral”“negative”].

#### 4.2.2 模型训练

**编码器.**输入序列 $U=(u_1, u_2, \dots, u_N)$ 中包含源语句和prompt,输出序列 $Y=(y_1, y_2, \dots, y_M)$ 为对应完形填空答案,基于生成式PTM完成从输入序列 $U$ 到输出序列 $Y$ 的映射.生成式seq2seq PTM具有编码器-解码器架构<sup>[91]</sup>.

基于PTM编码器得到输入序列嵌入表示 $H^{\text{en}}$ :

$$H^{\text{en}} = \text{Encoder}(u_1, u_2, \dots, u_N) \quad (11)$$

**解码器.**解码器根据 $H^{\text{en}}$ 逐token地生成序列 $Y$ ,预测第 $i$ 个token $y_i$ 以及对应解码器状态 $h_i^{\text{de}}$ 为:

$$y_i, h_i^{\text{de}} = \text{Decoder}([H^{\text{en}}; h_1^{\text{de}}, \dots, h_{i-1}^{\text{de}}], y_{i-1}) \quad (12)$$

**损失函数.**输出序列 $Y$ 的条件概率 $P(Y|U)$ 由生成每一个token的概率 $P(y_i|y_{<i}, U)$ 逐步相乘:

$$P(Y|U) = \prod_{i=1}^M P(y_i|y_{<i}, U) \quad (13)$$

其中, $y_{<i} = y_1 \dots y_{i-1}$ , $P(y_i|y_{<i}, U)$ 是 $\text{softmax}$ 归一化后在当前候选token列表(见表1)上的概率分布.

给定输入序列 $U$ ,模型训练的目标是使生成目标序列 $Y$ 的概率最大化.故采用负对数似然损失函数计算损失 $\mathcal{L}$ 为

$$\mathcal{L} = -\frac{1}{|D|} \sum_{(U, Y) \in D} \log P(Y|U; \theta) \quad (14)$$

其中, $\theta$ 为模型参数, $(U, Y)$ 为数据集 $D$ 中(输入序列,输出序列)对, $\log P(Y|U; \theta)$ 为

$$\log P(Y|U; \theta) = \sum_{i=1}^M \log P(y_i|y_{<i}, U; \theta) \quad (15)$$

#### 4.2.3 基于模式的约束解码

Prompt可以引导模型生成目标序列,然而,模型可能会因为在下游任务上微调不充分导致生成不符合目标模式的结果,特别是当数据集规模较小时.

启发于约束解码策略在生成式模型上取得的成功<sup>[11,70,94]</sup>,为了将目标模式知识注入解码过程,确保模型生成的元素位于相应的词汇集中,本文在实验中应用了动态约束解码算法,其思想是根据生成状态来动态调整候选token列表,而不是直接将整个词汇表作为候选token.本文设计的动态候选token列表具体见表1.

表1 动态候选token列表

当前token	候选token
$\langle c \rangle$	预定义类别集 $V_c, \langle \text{sep} \rangle$
$\langle a \rangle$	源语句 $X, \text{NULL}, \langle \text{sep} \rangle$
$\langle o \rangle$	源语句 $X, \text{NULL}, \langle \text{sep} \rangle$
$\langle s \rangle$	positive, negative, neutral, $\langle \text{sep} \rangle$
$\langle \text{sep} \rangle$	$\langle c \rangle, \langle a \rangle, \langle o \rangle, \langle s \rangle$

模型推理时,采用贪心策略在每个时间步中选择候选token列表中概率最大的token,得到生成的目标序列 $Y$ ,再通过格式解析得到预测的情感四元组;即先通过预定义的特殊分隔标记 $\langle \text{sep} \rangle$ 进行四元组分割,再通过 $\langle c \rangle, \langle a \rangle, \langle o \rangle$ 和 $\langle s \rangle$ 这4个哨兵token来提取对应的 $c, a, o$ 和 $s$ 元素,得到四元组 $(c, a, o, s)$ .

## 5 实验结果

### 5.1 数据集

本文在Rest15、Rest16、Restaurant和Laptop四个公开且广泛使用的ASQP数据集上进行了实验. Rest15和Rest16均为餐厅领域数据集,分别起源于SemEval 2015<sup>[95]</sup>和SemEval 2016<sup>[96]</sup>任务.部分学者<sup>[58,68]</sup>先后标注了三元组 $(a, o, s)$ 和 $(a, c, s)$ 数据集;在此基础上,Zhang等人<sup>[10]</sup>进一步补全元素,并将元素关联为四元组. Restaurant和Laptop分别为

餐厅和笔记本领域数据集,由Cai等人<sup>[6]</sup>构建,其中Restaurant起源于SemEval 2016<sup>[96]</sup>任务,Laptop为新标注的数据集.相比Rest15和Rest16,Restaurant和Laptop数据集不仅考虑了隐式方面词,还补全了对隐式观点词的标注,并且数据规模更大.为了公平,本文使用与已有工作相同的数据划分方式,具体如下表2所示.

表2 方面情感四元组数据集划分情况

数据集名称	训练集	验证集	测试集	方面类别数 $V_C$	情感类别数 $V_S$
Rest15	834	209	537	13	3
Rest16	1,264	316	544	13	3
Restaurant	1,531	170	585	13	3
Laptop	2,934	326	816	121	3

## 5.2 实验设置

### (1) DC-ASQP模型的超参数设置

编码模块中初始化词嵌入的PTM为BERT-base<sup>[66]</sup>,模型训练采用优化器AdamW,训练批量大小(batch size)设为8,训练回合(epoch)设为120.预训练的参数学习率和交叉注意力的参数学习率分别为 $1e-5$ 和 $5e-5$ ,其他层学习率设为 $7e-5$ .根据数据集中方面词和观点词文本跨度长度,在Rest15、Rest16、Restaurant和Laptop四个数据集上分别将候选跨度最多包含的token数max\_tokens\_in\_span设置为13、13、13和9.方面类别-情感极性分类模块中层数 $N_1$ 设为2,并发注意力层的头数为8,隐藏层维度 $d$ 为768.方面类别-情感极性查询向量 $m$ (略大于评论语句中包含的“方面类别-情感极性”最大个数)在4个数据集上均设为8.“方面词-观点词”完形填空模块的层数 $N_2$ 为12,该模块中对prompt的重复次数 $k$ (大于语句中每个“方面类别-情感极性”对应“方面词-观点词”的典型数量)在Rest15、Rest16、Restaurant和Laptop数据集上分别设为8、8、9和7.总损失函数中超参数 $\lambda$ 在Rest15、Rest16、Restaurant和Laptop数据集分别设为0.7、0.7、0.5和0.2.

### (2) GC-ASQP模型的超参数设置

seq2seq生成模型选择预训练模型T5-base<sup>[90]</sup>,prompt中 $\langle c \rangle$ 、 $\langle a \rangle$ 、 $\langle o \rangle$ 和 $\langle s \rangle$ 分别对应于T5模型中备用的特殊token $\langle extra\_id\_0 \rangle$ 、 $\langle extra\_id\_1 \rangle$ 、 $\langle extra\_id\_2 \rangle$ 和 $\langle extra\_id\_3 \rangle$ .本文输入和输出序列最大长度设为128,模型训练批量大小、学习率和训练回合分别为16、 $3e-4$ 、20.

ASQP任务常用的评测指标为精确率

(Precision, P)、召回率(Recall, R)以及 $F1$ 值.与现有工作<sup>[6-8,10,55,70,85]</sup>一样,本文采用 $F1$ 值作为主要评价指标.本文实验环境为Nvidia A40 GPU服务器,单卡显存为48G;实验中每个模型以不同随机种子重复运行5次,取5次实验结果平均值作为评估结果.

## 5.3 对比模型

根据模型采用的策略,已有工作可分为判别式模型和生成式模型2类.其中,生成式模型按是否采用数据增强策略可细分为基于单prompt(模板)模型和基于多prompt数据增强模型2类.因此,本文将基准模型共分为以下3类:

### (1) 判别式基准模型

① JET-ACOS(Cai等人,2021)<sup>[6]</sup>:采用二阶段流水线架构,首先利用JET模型<sup>[59]</sup>提取 $(a, o, s)$ 三元组,然后为每个三元组预测其方面类别以支持四元组预测.

② TAS-BERT-ACOS(Cai等人,2021)<sup>[6]</sup>:采用多阶段流水线架构,首先识别评论语句中的有效 $(c, s)$ ,并在此条件下以序列标注的方式提取对应的方面词和观点词,然后以笛卡尔积方式得到候选四元组,并通过分类模块过滤掉无效的四元组.

③ Extract-Classify-ACOS(Cai等人,2021)<sup>[6]</sup>:采用多阶段流水线架构,以序列标注方式提取方面词和观点词<sup>[65]</sup>,通过二分类任务判断评论语句中是否含有隐式方面词或隐式观点词,然后以笛卡尔积方式得到候选 $(a, o)$ ,再通过分类任务将 $(a, o)$ 与方面类别 $c$ 和情感极性 $s$ 进行关联并过滤掉无效四元组.

### (2) 基于单prompt(模板)的生成式基准模型

① ABSA-BART(Yan等人,2021)<sup>[55]</sup>:将源语句作为输入序列,将指针索引和分类索引线性化后作为模型的目标序列.由于原论文未考虑ASQP任务,我们修改该模型,在其支持的ASTE任务目标序列中增加方面类别分类索引,使模型支持处理ASQP任务.

② GAS(Zhang等人,2021a)<sup>[7]</sup>:该工作设计了注解式和抽取式两种建模方案.其中抽取式建模方案将源语句作为输入序列,语句中多个元组对应的文本直接拼接后作为目标序列.由于原论文未考虑ASQP任务,本文基于其抽取式建模方案直接处理ASQP任务.

③ ParaPhrase(Zhang等人,2021b)<sup>[10]</sup>:输入序列为源语句,目标序列为将元组映射为含有对应4个情感元素的自然语句,多个自然语句之间用分

隔符“[SSEP]”进行连接。

④ Seq2Path(Mao 等人, 2022)<sup>[8]</sup>: 输入序列为源语句, 目标序列为元组对应的文本. 该模型通过束搜索解码以树路径的方式生成目标序列, 并通过可学习的鉴别 token 来辅助剪枝以保留有效候选四元组.

⑤ LEGO-ABSA(Gao 等人, 2022)<sup>[85]</sup>: 该工作将多个元素 prompt 拼接作为任务 prompt, 借鉴搭建乐高积木的思想将多个简单任务 prompt 组装为复杂任务的 prompt, 实现由简单二元组预测任务到复杂四元组预测任务的迁移.

(3) 基于多 prompt(模板)增强的生成式基准模型

① DLO(Hu 等人, 2022)<sup>[9]</sup>: 将 4 个情感元素 24 种

不同排列的顺序模板作为候选目标序列. 按顺序模板在数据集上的平均熵对其升序排序, 基于前  $Top_p$  (默认为 3) 种顺序模板做数据增强后通过投票策略对不同顺序模板的预测结果进行聚合.

② MvP(Gou 等人, 2023)<sup>[11]</sup>: 按顺序 prompt 在数据集上的平均熵对 24 种顺序 prompt 做升序排序, 基于前  $Top_p$  (默认为 15) 种顺序 prompt 做数据增强后通过投票策略对不同顺序 prompt 的预测结果进行聚合.

#### 5.4 主要实验结果

各模型的实验对比结果如表 3 所示, 其中\*号表示实验结果来自于原论文, †号表示实验结果来自于 MvP<sup>[11]</sup>作者的复现, 其余为我们基于文献提供的开源代码在目标数据集上的评估结果.

表 3 主要实验对比结果

模型	Rest15			Rest16			Restaurant			Laptop		
	<i>P</i>	<i>R</i>	<i>F1</i>	<i>P</i>	<i>R</i>	<i>F1</i>	<i>P</i>	<i>R</i>	<i>F1</i>	<i>P</i>	<i>R</i>	<i>F1</i>
JET-ACOS	—	—	—	—	—	—	59.81*	28.94*	39.01*	44.52*	16.25*	23.81*
TAS-BERT-ACOS	—	—	—	—	—	—	26.29*	46.29*	33.53*	47.15*	19.22*	27.31*
Extract-Classify-ACOS	33.41	39.15	36.05	37.08	50.60	42.80	38.54*	52.96*	44.61*	45.56*	29.48*	35.80*
<b>DC-ASQP</b>	43.86	38.05	<b>40.75</b>	51.94	46.88	<b>49.28</b>	56.43	47.50	<b>51.58</b>	42.62	34.94	<b>38.40</b>
ABSA-BART	45.84	44.10	44.95	55.68	54.91	55.29	56.83	55.65	56.23	41.87	40.89	41.37
GAS	44.77	47.10	45.91	55.23	57.83	56.50	57.08	57.46	57.27	42.83	42.10	42.46
ParaPhrase	46.16*	47.72*	46.93*	56.63*	59.30*	57.93*	59.80	59.67	59.73	44.01	43.24	43.62
Seq2Path	47.72	45.41	46.54	56.54	56.62	56.58	—	—	58.41*	—	—	42.97*
LEGO-ABSA	—	—	46.10*	—	—	57.60*	59.63	59.20	59.41	43.81	43.34	43.57
<b>GC-ASQP</b>	48.44	49.84	<b>49.13</b>	58.48	61.04	<b>59.73</b>	61.44	60.54	<b>60.99</b>	44.75	44.41	<b>44.58</b>
DLO	47.08*	49.33*	48.18*	57.92*	61.80*	59.79*	—	—	59.99 <sup>†</sup>	—	—	43.64 <sup>†</sup>
MvP	—	—	51.04*	—	—	60.39*	—	—	61.54*	—	—	43.92*
<b>GC-ASQP(15连续prompt)</b>	51.55	52.25	<b>51.90</b>	60.95	63.21	<b>62.06</b>	62.54	60.87	<b>61.69</b>	45.24	44.64	<b>44.94</b>

判别式完形填空模型 DC-ASQP 与已有的最优判别式基准模型 Extract-Classify-ACOS 相比, 在 Rest15、Rest16、Restaurant 和 Laptop 四个数据集上 *F1* 值分别提升 4.70%、6.48%、6.97% 和 2.60%. DC-ASQP 是基于离散 prompt 的实验结果, 性能显著提升的原因主要包括以下几个方面:

(1) 相比 Extract-Classify-ACOS 的多阶段流水线架构, DC-ASQP 采用了两阶段联合学习架构, 缓解了误差传播问题;

(2) 利用了方面类别和情感极性的标签语义;

(3) 利用了离散 prompt 中蕴含的情感元素之间的关联关系;

(4) 相比序列标注, DC-ASQP 采用了基于跨度的编码能识别具有重叠 token 的方面词或观点词.

基于单离散 prompt 的生成式模型 GC-ASQP 与单 prompt(模板)场景下的最优基准 ParaPhrase 相比, 在 Rest15、Rest16、Restaurant 和 Laptop 四个数据集上的 *F1* 值分别提升了 2.20%、1.80%、1.26% 和 0.96%, 主要原因是: (1) 模型编码器输入完形填空风格的 prompt, 能够更好地利用 PTM 预训练时学到的知识; (2) 本文设计的离散 prompt 更能反映情感元素之间的关联语义. 与 LEGO-ABSA 模型相比, GC-ASQP 在 4 个数据集上 *F1* 值分别提升了 3.03%、2.13%、1.58% 和 1.01%, 主要原因是 LEGO-ABSA 中使用的离散 prompt “opinion: <*o*>, aspect: <*a*>, sentiment: <*s*>, category: <*c*>” 未考虑情感元素之间的语义关系. 此外, 在 Rest15 和 Rest16 上提升的效果显著高于 Restaurant

和 Laptop 的原因是 Paraphrase 和 LEGO-ABSA 均未使用约束解码策略,该策略对小规模数据集更有效<sup>[11]</sup>.

在多 prompt(模板)数据增强场景下,本文基于 15 个连续 prompt 数据增强的 GC-ASQP 模型(实验细节请参考 6.2 节拓展实验(6))比该场景下最优的 MvP 模型在 4 个数据集上  $F1$  值分别提升 0.86%、1.67%、0.15% 和 1.02%,表明本文设计的连续 prompt 及数据增强方案中虚拟 token 在多 prompt 中协同学习元素关联语义是有效的.此外,DLO 与 MvP 本质上都是基于不同元素排列顺序进行数据增强的模型.区别在于 DLO 基于不同顺序模板为一个输入构建多种顺序标签,而 MvP 基于多种顺序 prompt 做数据增强,其将不同顺序 prompt 分别与输入拼接并将标签中 4 个元素按照 prompt 中给定的顺序排列后作为目标序列.MvP 效果更好的原因主要包括以下 3 点:

(1)DLO 模型需要在输出序列中同时输出不同顺序的四元组.相比之下,MvP 分组处理不同 4 元组排列顺序,模型每次只需要输出 prompt 指定顺序的四元组即可;

(2)完形填空风格的 prompt 更有利于引导 PTM 中百科全书式知识;

(3)MvP 模型通过约束解码策略降低了生成不符合格式序列的概率.

此外,我们注意到基于单离散 prompt 的 GC-ASQP 与基于 3 个顺序模板数据增强的 DLO 相比,在 4 个数据集上平均  $F1$  值提升 0.71%,除了 GC-ASQP 采用完形填空任务形式和应用约束解码策略有利于四元组预测外,以下原因不可忽视:

(1)DLO 的模板设计时只考虑了元素顺序,忽略了元素的关联关系;

(2)DLO 需要在输出序列中输出 3 种元素排列

顺序不同的四元组,这会导致输出的元组数量增加 3 倍,生成任务变得更加复杂(如更多的元组导致生成元组的顺序更加复杂);

(3)DLO 采用投票策略保留出现次数多的四元组,这也可能导致仅在单顺序模板中正确预测的元组被错误地剪枝.

基于单离散 prompt 的 GC-ASQP 在 Laptop 数据集上比 MvP 的  $F1$  值高 0.66%,说明对于复杂的数据集(Laptop 数据集的方面类别数是其他数据集的 9.3 倍),4 个元素正确关联为四元组变得更困难,模型通过离散 prompt 借助人先验知识来理解、捕获情感元素之间的关联关系效果更显著.

最后,所有生成式模型都比判别式模型性能好.纵向对比本文提出的两个完形填空模型,相比 DC-ASQP,生成式模型 GC-ASQP 在 4 个数据集上  $F1$  值分别高出了 8.38%、10.45%、9.41% 和 6.18%,提升幅度很大.主要原因是:

(1)使用的预训练模型不同(DC-ASQP 采用 BERT,GC-ASQP 采用 T5);相比 BERT,生成式预训练模型 T5 具有更多的参数且预训练时使用的语料库更大,其能力更强<sup>[8]</sup>;

(2)虽然 DC-ASQP 模型联合学习两个子任务,但在推理阶段仍存在一定的误差传播问题;

(3)GC-ASQP 在生成四元组序列的过程中,既能更好地利用情感元素标签的语义信息,也能有效地区分各个四元组.

## 6 实验分析

### 6.1 消融实验

为了验证本文设计 prompt 的有效性,本文在 DC-ASQP 和 GC-ASQP 两个模型上分别进行 prompt 消融实验,实验结果如表 4 所示.

表 4 消融实验结果

模型	Rest15			Rest16			Restaurant			Laptop		
	$P$	$R$	$F1$	$P$	$R$	$F1$	$P$	$R$	$F1$	$P$	$R$	$F1$
DC-ASQP	43.86	38.05	40.75	51.94	46.88	49.28	56.43	47.50	51.58	42.62	34.94	38.40
DC-ASQP (w/o template)	41.85	37.95	39.80	48.95	47.43	48.18	52.35	48.20	50.19	39.20	35.58	37.30
GC-ASQP	48.44	49.84	49.13	58.48	61.04	59.73	61.44	60.54	60.99	44.75	44.41	44.58
GC-ASQP (w/o template)	46.78	48.66	47.70	56.53	59.90	58.17	59.46	58.37	58.91	43.14	42.89	43.01

(1)DC-ASQP 模型消融实验

DC-ASQP (w/o template)表示模型中直接用可学习向量替换特定于( $c, s$ )的 prompt.由表 3 可

知,替换之后,模型在 Rest15、Rest16、Restaurant 和 Laptop 数据集上的  $F1$  分别下降了 0.95%、1.10%、1.39% 和 1.10%.这主要是因为如果( $c, s$ )直接由

可学习向量表示,那么方面类别 $c$ 和情感极性 $s$ 对应的标签语义信息无法被利用,并且无法通过prompt来利用4个情感元素之间语义关系,从而导致 $(a, o)$ 完形填空任务性能下降.

### (2)GC-ASQP模型消融实验

GC-ASQP (w/o template) 表示 GC-ASQP 模型中输入为源语句,即去掉了输入序列中的prompt,输出保持不变.消融实验表明,去掉输入序列中的prompt,模型的F1在Rest15、Rest16、Restaurant和Laptop数据集上分别下降了1.43%、1.56%、2.08%和1.57%.由于移除了prompt,情感元素之间关联语义无法被显式利用,导致性能下降.此外,完形填空的题目(即prompt)缺少导致下游任务与PTM预训练任务相似性降低,不利于PTM中通用知识向下游任务迁移.

与表3中的基线GAS相比,GC-ASQP (w/o template)与GAS的输入均为原始评论语句,但GC-ASQP (w/o template)在输出序列中使用不同的哨兵token  $\langle c \rangle$ 、 $\langle a \rangle$ 、 $\langle o \rangle$ 和 $\langle s \rangle$ 来分隔情感元素,替代了GAS中的逗号.结果显示,使用哨兵token使得模型在4个数据集上的F1值分别得到1.79%、1.67%、1.64%和0.55%的提升.主要原因:

① 通过哨兵token,同一四元组中的不同哨兵token可以对不同情感元素加以区分,而不同四元组的相同哨兵则能学习到相应情感元素的相似特性,这种相似特性在不同样本之间以及同样本的不同四元组间进行迁移和更新;

② 模式约束解码策略带来的性能提升.

## 6.2 拓展实验

### (1)模型推理时间对比

模型的推理速度是衡量模型优劣的标准之一,在工业界尤为重要.我们对以下问题感兴趣:①自回归与非自回归完形填空模型在推理速度上的差异;②与判别式基准相比,DC-ASQP在推理速度上是否具有优势;③生成式模型中,贪心解码策略和束搜索解码策略对推理速度的影响.因此,本文选择对DC-ASQP、GC-ASQP、Extract\_Classify\_ACOS和Seq2Path四个模型的推理时间进行对比.为了公平,比较时将4个模型在推理阶段的批量大小(batch size)统一设为16,同时模型在测试集上推理时单独占用1张Nvidia A40 GPU卡,以确保GPU算力不会被其他任务分享.本文分别记录了4个模型在Rest15、Rest16、Restaurant和Laptop测

试数据集上运行3次的平均用时,对比结果如图4所示.

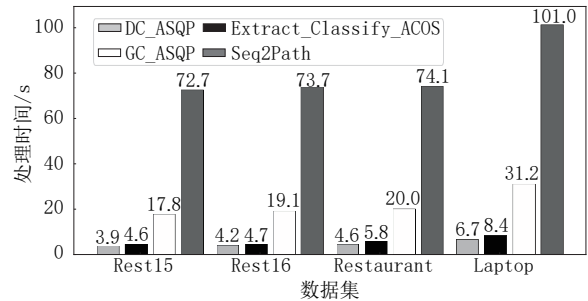


图4 模型推理时间对比

实验结果显示:GC-ASQP模型在Rest15、Rest16、Restaurant和Laptop测试集上的推理平均耗时分别是DC-ASQP模型的4.6、4.5、4.3和4.7倍,证明了非自回归模型在推理速度上显著优于自回归模型.其次,与已有的最优判别式基准模型Extract\_Classify\_ACOS相比,本文提出的DC-ASQP模型在性能具有显著优势的前提下并未花费更多时间,相反,其在推理速度上也具有一定优势,主要原因是模型采用了并发预测框架,有助于提升模型的推理速度.最后,对比Seq2Path和GC-ASQP推理时间可知,采用贪心解码策略在解码效率上要显著优于基于束搜索的解码策略.

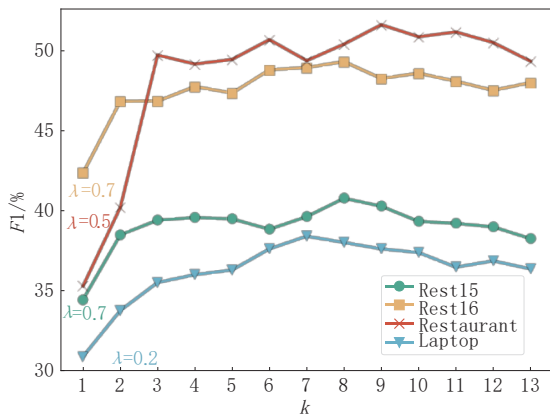
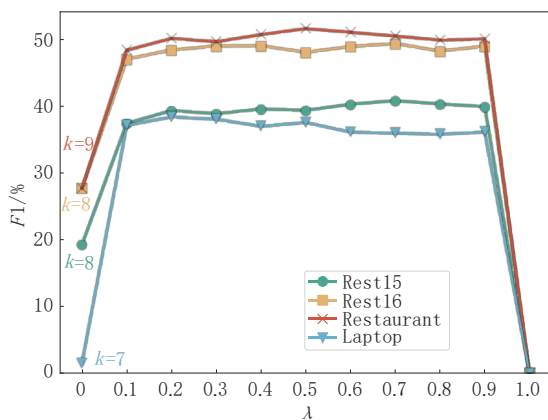
### (2)DC-ASQP参数研究

探究DC-ASQP模型中与填空模块相关的两个参数:对prompt的重复次数 $k$ 和总损失函数中参数 $\lambda$ .本文采用网格搜索(grid search)寻找最优参数,两个参数的实验结果如图5和图6所示.参数 $k$ 在4个数据集上变化趋势相似,即先上升到峰值后逐渐下降,在Rest15、Rest16、Restaurant和Laptop四个数据集上峰值分别为8,8,9和7.由于一个 $(c, s)$ 可能对应多个 $(a, o)$ , $k$ 值过小可能会遗漏四元组, $k$ 值过大会增加错误预测 $(a, o)$ 的风险且需要更多的计算量.

如图6所示,随着参数 $\lambda$ 增大,模型性能先快速上升,然后波动上升到峰值再波动下降,最后在 $\lambda$ 为1时趋于0.该实验结果反映出以下结论:

①  $\lambda$ 为1时,仅基于 $(c, s)$ 分类模块得到的总损失无法对 $(a, o)$ 完形填空模块进行有效训练,填空模块性能趋于0,故最终性能趋于0;

②  $\lambda$ 为0时,在4个数据集上的F1值分别为19.23%、27.66%、27.69%和1.52%,说明仅基于 $(a, o)$ 完形填空得到的总损失对微调 $(c, s)$ 分类模

图5 参数 $k$ 的影响图6 参数 $\lambda$ 的影响

块同样有一定作用,可能的原因是,填空模块基于理想 $(c, s)$ 构建的prompt来预测对应的 $(a, o)$ ,即该模块输入中蕴含了正确的 $(c, s)$ 信息;

③ Laptop数据集上性能峰值对应的参数 $\lambda$ 为0.2,说明对于更复杂的数据集(例如Laptop数据集有121个方面类别),模型应该更关注有助于情感元素关联的填空任务.

④ 联合训练时参数 $\lambda$ 会同时影响 $(c, s)$ 分类模块、 $(a, o)$ 完形填空模块以及这两个子任务正确关联的性能.如图6所示,在Rest15、Rest16、Restaurant和Laptop四个数据集上网格搜索的最优参数 $\lambda$ 分别为0.7、0.7、0.5和0.2.

### (3) DC-ASQP中子任务的误差传播分析

DC-ASQP模型的表现( $F1$ 值)在Rest15、Rest16和Restaurant上比在Laptop上高出2.35%、10.88%和13.18%,如表5所示.该差异主要来自于各自 $(c, s)$ 分类子任务和 $(a, o)$ 完形填空子任务带来的误差传播.在 $(c, s)$ 分类子任务中,Laptop上的 $F1$ 值比Rest15、Rest16和Restaurant上 $F1$ 值分别低17.48%、24.7%和21.63%.Laptop上 $(c, s)$ 分类

子任务表现较差的原因在于类别总数量多,Laptop数据集中方面类型 $c$ 的数量为121个,而Rest15、Rest16和Restaurant中方面类型 $c$ 的数量为13个,情感类别 $s$ 总数量一样,故前者 $(c, s)$ 的类别总数约为后者的9.3倍.对于 $(c, s)$ 分类子任务,Rest16上的 $F1$ 值比Rest15上高7.22%,主要原因是Rest16数据集规模更大,模型训练得更充分;Rest16上的 $F1$ 值比Restaurant上高3.07%,可能原因是,Restaurant中增加了含有隐式观点词的数据,对 $(c, s)$ 分类任务有更大的挑战,因为缺少显式观点词不利于对情感极性的判断.

表5 DC-ASQP子任务实验分析

数据集	子任务	$P$	$R$	$F1$
Rest15	$(c, s)$ 分类任务	73.90	64.13	68.67
	$(a, o)$ 填空任务	56.00	48.62	52.05
	ASQP任务	43.87	38.05	40.75
Rest16	$(c, s)$ 分类任务	80.03	72.15	75.89
	$(a, o)$ 填空任务	62.82	56.70	59.60
	ASQP任务	51.94	46.88	49.28
Restaurant	$(c, s)$ 分类任务	79.74	67.00	72.82
	$(a, o)$ 填空任务	69.06	58.14	63.13
	ASQP任务	56.43	47.50	51.58
Laptop	$(c, s)$ 分类任务	56.81	46.58	51.19
	$(a, o)$ 填空任务	72.47	59.39	65.28
	ASQP任务	42.63	34.94	38.40

对 $(a, o)$ 完形填空子任务评测时,直接将理想 $(c, s)$ 对应的prompt作为输入.在Rest15、Rest16、Restaurant和Laptop四个数据集上 $(a, o)$ 填空子任务对应的 $F1$ 值是递增的,与数据集规模正相关.Restaurant数据集上 $(a, o)$ 填空子任务对应的 $F1$ 值比Rest16数据集高3.53%,表明即使增加含有隐式观点词的数据,如果方面类别和情感极性已知,DC-ASQP依然很好地完成了 $(a, o)$ 填空子任务,说明DC-ASQP模型具有处理隐式方面情感的能力.

### (4) GC-ASQP模型中不同离散prompt对比

通过对比不同离散prompt对GC-ASQP模型最终性能的影响,分析prompt构建时的关键因素.我们选择在餐厅领域的Restaurant数据集和笔记本领域的Laptop数据集上进行实验探究.设计prompt时,主要考虑情感元素的顺序<sup>[9,11]</sup>,以及prompt是否能反映4个情感元素之间的语义关系.实验结果如表6所示.

表6 不同prompt对比实验

Prompt 元素 编号 顺序	Prompt 内容	Restaurant			Laptop			Both Avg(F1)
		P	R	F1	P	R	F1	
T1	$\langle c \rangle$ is $\langle s \rangle$ because $\langle a \rangle$ is $\langle o \rangle$	60.11	60.18	60.14	44.48	43.96	44.22	52.18
T2	Category is $\langle c \rangle$ , sentiment is $\langle s \rangle$ , aspect is $\langle a \rangle$ and opinion is $\langle o \rangle$	59.60	59.42	59.51	43.76	43.38	43.57	51.54
T3	$\langle a \rangle$ is $\langle o \rangle$ so $\langle c \rangle$ is $\langle s \rangle$	60.68	60.58	60.63	43.97	43.20	43.58	52.11
T4	Aspect is $\langle a \rangle$ , opinion is $\langle o \rangle$ , category is $\langle c \rangle$ and sentiment is $\langle s \rangle$	59.50	59.25	59.37	43.36	43.07	43.21	51.29
T5	The category of $\langle a \rangle$ is $\langle c \rangle$ and the sentiment of $\langle o \rangle$ is $\langle s \rangle$	60.89	60.12	60.50	43.71	43.34	43.52	52.01
T6	Aspect is $\langle a \rangle$ , category is $\langle c \rangle$ , opinion is $\langle o \rangle$ and sentiment is $\langle s \rangle$	59.79	59.39	59.59	43.25	42.78	43.01	51.30
T7	The $\langle c \rangle$ such as $\langle a \rangle$ is $\langle o \rangle$ , so I feel $\langle s \rangle$	61.44	60.54	60.99	44.75	44.41	44.58	52.79
T8	Category is $\langle c \rangle$ , aspect is $\langle a \rangle$ , opinion is $\langle o \rangle$ and sentiment is $\langle s \rangle$	60.21	59.48	59.84	44.22	43.78	44.00	51.92

Prompt T2、T4、T6和T8未充分考虑情感元素之间语义关系,仅按不同顺序排列4个情感元素待填空项(即哨兵token)。对比prompt T2、T4、T6和T8,实验结果显示仅考虑元素顺序的情况下,排列顺序优先级为:caos>csao>acos>aocs,表明情感元素生成顺序一定程度上影响了模型最终性能,与已有工作结论类似<sup>[9,11]</sup>。原因是,模型以自回归方式逐个生成token时,已生成token的正确率会影响后续token的生成质量。方面类别c排在方面词a前面效果略好一些,可能是受隐式方面词的影响。

为了验证构建prompt时情感元素之间语义关系的表述是否有效,针对不同的元素顺序分别设计了prompt T1、T3、T5和T7。相比prompt T2、T4、T6和T8,建立情感元素语义联系后的prompt在两个数据集上的平均F1值分别提升了0.64%、0.82%、0.71%和0.87%。在以上prompt中,prompt T7(即“The  $\langle c \rangle$  such as  $\langle a \rangle$  is  $\langle o \rangle$ , so I feel  $\langle s \rangle$ ”)效果最好,它是符合人类直观理解的,因为该prompt反映了方面词a和方面类别c的从属关系,观点词o是对方面类别和方面词的情感表达<sup>[3]</sup>,而将情感极性s排在最后是因为情感极性与方面类别、方面词以及观点词三个情感元素均相关<sup>[68]</sup>。

#### (5)GC-ASQP模型中连续prompt效果

连续prompt在简单文本任务<sup>[13]</sup>中一般表现比人工设计的离散prompt效果略好,因此,我们好奇在复杂的ASQP任务上该规律是否依旧成立?

本文选择性能最好的离散prompt T7作为基准,按照相同的顺序(即caos)设计对应的连续prompt,通过添加虚拟token的方式来描述不同情感元素类型之间的关系,同时构建了4类情感元素的关联语境。因此,caos顺序对应的连续prompt被设计为“ $\langle c \rangle$  [ $CA_1$ ]  $\dots$  [ $CA_{n_1}$ ]  $\langle a \rangle$  [ $AO_1$ ]  $\dots$

[ $AO_{n_2}$ ]  $\langle o \rangle$  [ $OS_1$ ]  $\dots$  [ $OS_{n_3}$ ]  $\langle s \rangle$ ”,其他顺序的连续prompt设计方法类似。其中[·]为虚拟token, $n_1$ 、 $n_2$ 和 $n_3$ 为虚拟token个数。

此外,需要进一步探索超参数 $n_1$ 、 $n_2$ 和 $n_3$ 的最优取值。然而,即使 $n_1$ 、 $n_2$ 和 $n_3$ 只在一个小范围变化,其组合的数量 $n_1 * n_2 * n_3$ 也是非常多的。因此,我们采用粗略超参数调优策略<sup>[13]</sup>,即让 $n_1$ 、 $n_2$ 和 $n_3$ 取相同的值进行调优,记作 $n = n_1 = n_2 = n_3$ 。Prompt中虚拟token个数n分别取{0, 1, 2, 3, 4, 5}对预测性能的影响如图7所示。

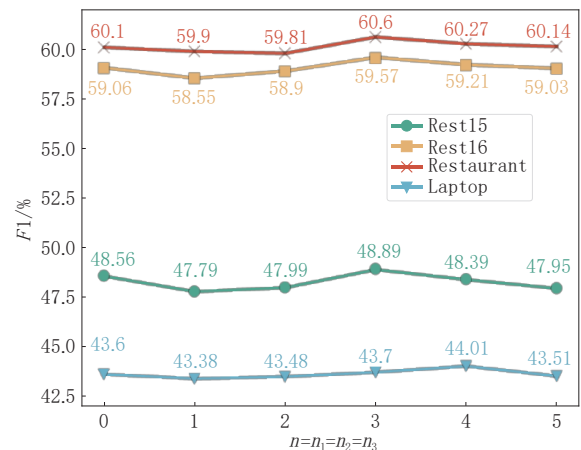


图7 单连续prompt中参数n的影响

图7显示了prompt中不同n值对ASQP任务的影响。当n=0时表示prompt中不包含任何虚拟token,即为MvP<sup>[11]</sup>模型使用的顺序prompt。

根据图7,可以得出以下结论:

① caos离散prompt(即表6中T7)在4个数据集上的性能均略高于表现最优的连续prompt。说明对于复杂的ASQP任务,让模型通过虚拟token自动学习prompt中4元素之间的关联语义是相对困难的。此外,离散prompt与最优连续prompt在Laptop数据集上的性能相差最多,说明连续prompt在复杂



任务(Laptop的方面类别数较多)中找到最优prompt的难度更大。

② 最优的连续prompt比顺序prompt效果略好,说明选择合适数量的虚拟token能够学到了一些有助于ASQP任务的知识。

③ 连续prompt(即 $n$ 取值1到5)的性能是先上升后下降的,因此, $n$ 是关键超参数之一。实验结果表明使用少量虚拟token足以学习元素之间的关联语义,但需要选择合适的 $n$ 值才有效果。随着 $n$ 增大,prompt具有更多的灵活性,同时也具有更多的模糊性<sup>[13]</sup>,我们认为 $n$ 值的选择与数据集规模和任务难度有关。

#### (6)GC-ASQP中多prompt数据增强效果

启发于MvP模型<sup>[11]</sup>中采用多顺序prompt数据增强策略的显著效果,本文有兴趣进一步探索基于不同顺序情感元素关联prompt进行数据增强的效果。虽然在单一 $caos$ 顺序场景下,本文设计的连续prompt效果要略差于人工设计的离散prompt。但由于连续prompt使用虚拟token来表示元素之间的关系,这些token不会与源语句中的token共享,在具有模糊性的同时也蕴含了更多的灵活性,即其具有在整个语义空间中寻找最佳prompt的潜力,而不充分的训练会限制虚拟token的学习效果。因此,理论上采用数据增强策略,从不同顺序的连续prompt中协同学习元素之间的语义关联关系是可行的,4个元素两两组合共有12种关系,即 $\{ca, ao, os, co, cs, as, ac, oa, so, oc, sc, sa\}$ ,描述这12种关系的虚拟token个数分别设为 $n_1, n_2, \dots, n_{12}$ 。设计连续prompt的核心是不同顺序prompt中用来描述相同两个情感元素语义关系的虚拟token应保持一致。

例如,我们在 $aocs, aosc, caos, saoc, csao$ 和 $scao$ 这6种prompt中都使用“ $[AO_1] \dots [AO_n]$ ”来描述 $a \rightarrow o$ 顺序下方面词类型 $\langle a \rangle$ 和观点词类型 $\langle o \rangle$ 的关联关系。这样,在多prompt数据增强时,如果这6种顺序的prompt中有多个被选中,则能够从多视图更好地学习 $a \rightarrow o$ 顺序下方面词 $\langle a \rangle$ 和观点词 $\langle o \rangle$ 之间的语义。

具体地,我们考虑了下面3组数据增强方案:

① 3离散prompt:取表6中效果最好的前3个离散prompt,即T1、T3和T7;

② 多连续prompt:为了消除元素顺序的影响,本文直接使用MvP<sup>[11]</sup>基于最小熵<sup>[9]</sup>对24种顺序prompt的排序,超参数 $Top_p$ 也与MvP保持一致,即 $Top_p$ 为15。然后,将每个数据集中排序前15的顺序prompt替换为对应顺序的连续prompt进行数据增强;

③ 15跨类prompt:保持 $Top_p$ 为15不变,用于数据增强的prompt为T1、T3和T7这3个离散prompt与12个连续prompt,为了保证3个离散prompt一定被选中,我们将其排序序号设置为前3,再从其余21个prompt中选出排序前12的连续prompt。

为了公平比较,我们与MvP一样,基于投票策略<sup>[11]</sup>对不同顺序下的预测结果进行聚合,保留在大多数顺序下都能被预测出来的四元组。我们采用粗略调优策略来确定虚拟token数 $n_1, n_2, \dots, n_{12}$ ,即实验时让 $n=n_1=n_2=\dots=n_{12}, n \in \{1, 2, 3, 4, 5\}$ 。多prompt数据增强实验结果如表7所示,对于包含连续prompt的数据增强方案,每个数据集记录的是不同 $n$ 下最优的实验结果。表中MvP、单离散prompt和单连续prompt是用于对比的3个基准。

表7 多prompt数据增强实验结果

模型	Rest15			Rest16			Restaurant			Laptop		
	$P$	$R$	$F1$	$P$	$R$	$F1$	$P$	$R$	$F1$	$P$	$R$	$F1$
MvP	—	—	51.04*	—	—	60.39*	—	—	61.54*	—	—	43.92*
GC-ASQP(单离散prompt)	48.44	49.84	49.13	58.48	61.04	59.73	61.44	60.54	60.99	44.75	44.41	44.58
GC-ASQP(单连续prompt)	48.03	49.79	48.89	58.30	60.90	59.57	61.07	60.13	60.60	44.13	43.90	44.01
GC-ASQP(3离散prompt)	51.01	50.38	50.69	59.33	60.45	59.88	61.39	58.84	60.09	44.54	43.49	44.01
GC-ASQP(15连续prompt)	51.55	52.25	51.90	60.95	63.21	62.06	62.54	60.87	61.69	45.24	44.64	44.94
GC-ASQP(15跨类prompt)	50.80	51.54	51.17	60.74	62.16	61.44	62.15	60.48	61.30	45.19	44.36	44.77

从表7实验结果,可得出以下结论:

① 相比 $caos$ 顺序的单离散prompt(即表6中T7),3离散prompt数据增强方案在Rest15、Rest16上 $F1$ 分别提升1.56%和0.15%,但在Restaurant和

Laptop上 $F1$ 分值反而下降了0.90%和0.57%。在前两个数据集上性能增强的主要原因是这两个数据集规模较小,数据增强策略使模型训练更充分,性能提升更显著;在后两个数据集上性能反而下降,可能

原因是3个离散prompt预测正确的四元组中重叠的比例相对较小,导致模型采用投票策略<sup>[11]</sup>聚合结果时使仅在单个prompt中正确预测的四元组被错误地剪枝.这也是导致15跨类prompt数据增强方案比15连续promot数据增强方案效果略差的原因.

② 相比 *caos* 的单连续 prompt, 15 连续 prompt 的数据增强方案在 4 个数据集上 F1 值分别提升 3.01%、2.49%、1.09% 和 0.93%, 这说明我们在不同顺序 prompt 中使用一定数量的相同虚拟 token 来描述给定顺序下两类情感元素之间语义关系的策略是有效的. 基于此设计, 多个连续 prompt 不再像离散 prompt 一样单独发挥作用, 而是协同发挥作用, 使得不同连续 prompt 生成的正确四元组重叠比例增大, 基于投票策略聚合不同 prompt 预测结果时效果更好. 此外, 数据增强策略使得模型在前 2 个小规模数据集上提升效果更显著.

③ 相比 MvP 模型, 15 连续 prompt 的数据增强方案只将 MvP 中 15 个顺序 prompt 替换为对应顺序的连续 promot, 其在 4 个数据集上 F1 值分别提升 0.86%、1.67%、0.15% 和 1.02%, 说明本文设计的连续 prompt 在有充足的训练数据后, 其虚拟 token 能够学到元素之间的关联关系, 进而提升了模型的预测效果.

④ 相比最优的单离散 prompt T7, 15 连续 prompt 数据增强方案的 F1 值在 4 个数据集上全部提升, 提升幅度分别为 2.77%、2.33%、0.70% 和 0.36%. 这表明通过 15 连续 prompt 数据增强方案, 连续 prompt 中虚拟 token 学到的 4 元素关联语义比离散 prompt 中人类给定的 4 元素关联语义泛化性更强, 在未知数据集上推理效果更好.

多 promot 数据增强方案中, 模型关键超参数之一一是用于描述 2 类元素语义关系的虚拟 token 个数  $n$ . 考虑到论文篇幅, 这里只给出在 15 连续 prompt 数据增强方案下对参数  $n$  的粗略调参结果(15 跨类 prompt 结果类似), 即设  $n=n_1=n_2=\dots=n_{12}$ , 研究  $n$  分别取 {1, 2, 3, 4, 5} 对预测性能的影响, 实验结果如图 8 所示.

由图 8 可知, 15 连续 prompt 的数据增强方案中, 随着  $n$  增加, 四元组预测性能先上升后下降, 这与图 7“单连续 prompt”的整体趋势是类似的. 我们注意到  $n$  取不同值在 4 个数据集上差异可达 1.53%、1.70%、2.16% 和 0.96%. 因此, 包含多个连续 prompt 的数据增强方案中超参数  $n$  的选择非常关键. 由于 15 连续 prompt 数据增强方案在 4 个数据

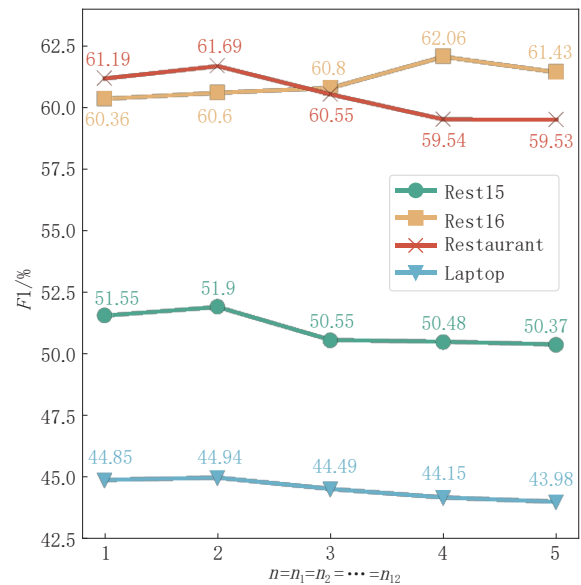


图 8 15 连续 prompt 数据增强方案中参数  $n$  的影响

集上的最优  $n$  值分别为 2、4、2 和 2, 与图 7“单连续 prompt”中的 3、3、3 和 4 不一样. 因此, 包含多个连续 prompt 的数据增强方案中, 影响最优  $n$  的因素可能包括但不限于: 任务难度、数据集规模和  $\text{Top}_p$  (即数据增强方案中选择的 prompt 个数) 等.

#### (7) GC-ASQP 模型中 prompt 输入位置的影响

探索 prompt 放在输入序列的不同位置对 GC-ASQP 模型性能的影响, 主要考虑两种位置: 将 prompt 放在源语句前面或后面, 分别记为  $T-X$  或  $X-T$ . 本节选用上节中性能较好的离散 prompt T1、T3、T5 和 T7 在 Restaurant 数据集上进行实验, 实验结果如表 8 所示.

实验结果显示, 将 prompt 放在源语句后面 ( $X-T$ ) 时模型的表现更好. 因此, 本文实验中将 prompt  $T$  通过  $\langle \text{sep} \rangle$  拼接在源语句  $X$  后作为生成式完形填空模型 GC-ASQP 输入序列  $U$ .

#### (8) GC-ASQP 模型中解码策略的影响

探索生成式完形填空模型 GC-ASQP 中不同解码策略对模型性能的影响, 主要考虑两种解码策略: 贪心解码策略和束搜索解码策略. 贪心解码在每个时间步都贪心地取概率最高的 token, 依次生成, 一个输入序列对应得到一个输出序列. 贪心解码策略的输出序列中包含多个情感四元组, 不同的四元组之间以特殊 token  $\langle \text{sep} \rangle$  进行切分, 由模型自行确定要输出的四元组数量.

束搜索解码策略需要预先设置超参数束宽, 记为  $\text{num\_beams}$ . 束搜索解码在每个时间步会保留概率最高的前  $\text{num\_beams}$  个序列, 因此解码后会生成

表8 Prompt位置对模型性能影响对比实验

Prompt 编号	Prompt 内容	Prompt 位置	Restaurant		
			P	R	F1
T1	$\langle c \rangle$ is $\langle s \rangle$ because $\langle a \rangle$ is $\langle o \rangle$	X-T	60.11	60.18	60.14
		T-X	60.00	59.82	59.91
T3	$\langle a \rangle$ is $\langle o \rangle$ so $\langle c \rangle$ is $\langle s \rangle$	X-T	60.68	60.58	60.63
		T-X	59.96	60.29	60.12
T5	The category of $\langle a \rangle$ is $\langle c \rangle$ and the sentiment of $\langle o \rangle$ is $\langle s \rangle$	X-T	60.89	60.12	60.50
		T-X	60.68	59.84	60.26
T7	The $\langle c \rangle$ such as $\langle a \rangle$ is $\langle o \rangle$ , so I feel $\langle s \rangle$	X-T	61.44	60.54	60.99
		T-X	60.87	60.24	60.55

num\_beams 个输出序列. 对于 ASQP 任务, 将 num\_beams 设置为大于单个评论语句中典型四元组的数量, 并让每个输出序列仅包含 1 个四元组信息, 即一个源语句通过束搜索解码生成 num\_beams 个候选四元组. 由于不同评论语句中实际含有的四元组数量不一样且一般小于 num\_beams, 因此需要对生成的 num\_beams 个候选四元组进行剪枝, 过滤掉其中无效的四元组.

本文采用 Mao 等人<sup>[8]</sup>提出的剪枝方案. 为了提升剪枝效果, 该方案引入了 1 个额外的元素  $v \in \{\text{"true"}, \text{"false"}\}$ , 旨在让模型通过对  $v$  值的学习来鉴别生成的四元组是否正确, 因此生成的四元组格式为 " $\langle c \rangle c_i \langle a \rangle a_i \langle o \rangle o_i \langle s \rangle s_i \langle v \rangle v_i$ ". 模型需要在有正负样本的训练数据集上进行学习, 数据集中已标注的四元组直接作为正样本, 对应的  $v$  设置为 "true"; 通过 Mao 等人<sup>[8]</sup>提出的数据增强方法获取  $v$  为 "false" 的负样本. GC-ASQP 束搜索解码策略下设计 prompt 时, 需要增加鉴别元素  $v$  对应的空白占位符  $\langle v \rangle$ , 我们选择性能最好的离散 prompt T7 进行实验, 即 prompt 为: "The  $\langle c \rangle$  such as  $\langle a \rangle$  is  $\langle o \rangle$ , so I feel  $\langle s \rangle$ . It is  $\langle v \rangle$ ". 剪枝时, 删除生成结果中  $v_i$  为 false 的方面情感四元组 ( $c_i, a_i, o_i, s_i$ ). 本节实验中分别将束搜索解码的 num\_beams 设置为 4、6、8 和 10, 在 Restaurant 数据集上进行实验, 实验结果如表 9 所示.

表9 不同解码策略对模型性能影响对比实验

解码策略	num_beams	Restaurant		
		P	R	F1
贪心解码	1	61.44	60.54	60.99
	4	59.70	57.23	58.44
	6	60.61	58.13	59.34
束搜索解码	8	59.39	59.50	59.44
	10	60.10	58.53	59.30

实验结果显示, GC-ASQP 模型使用贪心解码策略的 F1 值比使用束搜索解码策略时的最佳 F1 值高出 1.55%. 可能原因为:

① 相比束搜索解码并发且独立地生成多个四元组, 贪心解码策略通过一个序列生成多个四元组, 使相同哨兵 token 在序列中反复出现, 后续哨兵 token(待填空项)的嵌入表示能不断迭代更新, 有利于学习多个四元组之间的关系;

② 束搜索解码时需要通过剪枝策略<sup>[8]</sup>从 num\_beams 个候选四元组中筛选出有效的四元组, 模型最终性能受剪枝策略的影响. 结合表 7 中的对比结果, 本文的 GC-ASQP 模型采用效果更佳的贪心解码策略来生成输出序列.

## 7 结论与未来工作

本文将方面情感四元组预测(ASQP)建模为完形填空任务, 即原始评论语句为给定的阅读材料, prompt 为完形填空题目, 方面情感四元组中情感元素为填空的答案. 本文基于不同的模型策略提出 2 个完形填空模型 DC-ASQP 和 GC-ASQP, 模型的核心思想为: 通过 prompt 来挖掘不同情感元素之间关联语义; 通过 prompt 限定的语义关系, 将填入的情感元素关联为四元组. 模型在 Rest15、Rest16、Restaurant 和 Laptop 四个数据集上进行了多组测评和对比, 包括同类别基准模型横向对比、跨类别基准模型纵向对比. 横向对比结果显示, 本文提出的判别式完形填空模型 DC-ASQP 和生成式完形填空模型 GC-ASQP 均取得了显著的性能提升, 说明基于语义关系构建的 prompt 有助于提升方面情感四元组预测性能. 纵向对比结果表明生成类模型具有更强的方面情感四元组预测能力. 此外, 本文还探索了 prompt 位置、解码策略、不同的多 prompt 数据增

强策略以及 prompt 中情感元素的顺序等对 GC-ASQP 模型性能的影响。

未来工作可从以下方面考虑：

(1) 任务拓展. 将本文从元素语义关联角度设计 prompt 的方法应用到 ABSA 更多子任务上, 如方面词-情感极性预测、方面词-观点词预测、方面情感三元组预测等任务, 也可以拓展到其它相似的结构化信息抽取领域, 如事件抽取、实体关系抽取和通用信息抽取等。

(2) 性能提升. 非自回归并发架构的 DC-ASQP 模型在处理速度上具有显著优势, 提升该模型的性能是一项非常有意义的工作, 未来可以从以下角度进一步提升该模型性能, 包括: 基于依存句法树进一步增强情感元素的依赖关系; 针对不同的方面类别-情感极性对设计不同的 prompt. 对于 GC-ASQP 模型, 理论上基于最小熵<sup>[9,11]</sup>对 24 种连续 prompt 重新排序并进一步探索多连续 prompt 数据增强方案中参数  $Top_p$  的最优取值可进一步提升模型性能。

(3) 少样本、零样本下的 ASQP. 探索大语言模型、混合 prompt 等技术在 ASQP 任务上的应用, 探索 ASQP 任务在不同领域和场景中的迁移问题。

**致谢** 由衷地感谢论文评审专家和编辑对本文所提出的修改意见!

## 参 考 文 献

- [1] Liu B. Sentiment analysis and opinion mining. California, USA: Morgan & Claypool Publishers, 2012
- [2] Pontiki M, Galanis D, Pavlopoulos J, et al. SemEval-2014 task 4: Aspect based sentiment analysis//Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval). Dublin, Ireland, 2014: 27-35
- [3] Zhang W, Li X, Deng Y, et al. A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(11): 11019-11038
- [4] Chen Zhuang, Qian Tie-Yun, Li Wan-Li, et al. Low-resource aspect-based sentiment analysis: A survey. Chinese Journal of Computers, 2023, 46(07): 1445-1472 (in Chinese)  
(陈壮, 钱铁云, 李万理等. 低资源方面级情感分析研究综述. 计算机学报, 2023, 46(07): 1445-1472)
- [5] Nazir A, Rao Y, Wu L, et al. Issues and challenges of aspect-based sentiment analysis: A comprehensive survey. IEEE Transactions on Affective Computing, 2022, 13(2): 845-863
- [6] Cai H, Xia R, Yu J. Aspect-category-opinion-sentiment quadruple extraction with implicit aspects and opinions//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP). Online, 2021: 340-350
- [7] Zhang W, Li X, Deng Y, et al. Towards generative aspect-based sentiment analysis//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP). Online, 2021: 504-510
- [8] Mao Y, Shen Y, Yang J, et al. Seq2Path: generating sentiment tuples as paths of a tree//Findings of the Association for Computational Linguistics (FindingsACL) Dublin, Ireland, 2022: 2215-2225
- [9] Hu M, Wu Y, Gao H, et al. Improving aspect sentiment quad prediction via template-order data augmentation//Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP). Abu Dhabi, United Arab Emirates, 2022: 7889-7900
- [10] Zhang W, Deng Y, Li X, et al. Aspect sentiment quad prediction as paraphrase generation//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP). Punta Cana, Dominican Republic, 2021: 9209-9219
- [11] Gou Z, Guo Q, Yang Y. MvP: Multi-view prompting improves aspect sentiment tuple prediction//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL). Toronto, Canada, 2023: 4380-4397
- [12] Petroni F, Lewis P, Piktus A, et al. How context affects language models' factual predictions//Proceedings of the 2nd Conference on Automated Knowledge Base Construction (AKBC). Online, 2020: 1-15
- [13] Zhang Z, Wang B. Prompt learning for news recommendation//The 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR). Taipei, China, 2023: 227-237
- [14] Lester B, Al-Rfou R, Constant N. The power of scale for parameter-efficient prompt tuning//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP). Punta Cana, Dominican Republic, 2021: 3045-3059
- [15] Gu Y, Han X, Liu Z, et al. PPT: Pre-trained prompt tuning for few-shot learning//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL). Dublin, Ireland, 2022: 8410-8423
- [16] Tang D, Qin B, Liu T. Aspect level sentiment classification with deep memory network//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP). Austin, USA, 2016: 214-224
- [17] Wang Y, Huang M, Zhu X, et al. Attention-based LSTM for aspect-level sentiment classification//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP). Austin, USA, 2016: 606-615
- [18] Chen P, Sun Z, Bing L, et al. Recurrent attention network on memory for aspect sentiment analysis//Proceedings of the 2017

- Conference on Empirical Methods in Natural Language Processing (EMNLP). Copenhagen, Denmark, 2017: 452-461
- [19] Li X, Bing L, Lam W, et al. Transformation networks for target-oriented sentiment classification//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL). Melbourne, Australia, 2018: 946-956
- [20] Huang B, Carley K. Parameterized convolutional neural networks for aspect level sentiment classification//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP). Brussels, Belgium, 2018: 1091-1096
- [21] Zhang C, Li Q, Song D. Aspect-based sentiment classification with aspect-specific graph convolutional networks//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong, China, 2019: 4568-4578
- [22] Zeng Yi-fu, Lan Tian, Wu Zu-feng, et al. Bi-memory based attention model for aspect level sentiment classification. Chinese Journal of Computers, 2019, 42(08): 1845-1857(in Chinese)  
(曾义夫, 蓝天, 吴祖峰等. 基于双记忆注意力的方面级别情感分类模型. 计算机学报, 2019, 42(08): 1845-1857)
- [23] Wang K, Shen W, Yang Y, et al. Relational graph attention network for aspect-based sentiment analysis//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL). Online, 2020: 3229-3238
- [24] Chen Z, Qian T. Relation-aware collaborative learning for unified aspect-based sentiment analysis//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL). Online, 2020: 3685-3694
- [25] Zhang M, Qian T. Convolution over hierarchical syntactic and lexical graphs for aspect level sentiment analysis//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Online, 2020: 3540-3549
- [26] Du Cheng-yu, Liu Peng-yuan. Helical attention networks for aspect-level sentiment classification. Journal of Chinese Computer Systems, 2020, 34(09): 70-77(in Chinese)  
(杜成玉, 刘鹏远. 基于螺旋注意力网络的方面级别情感分析模型. 中文信息学报, 2020, 34(09): 70-77)
- [27] Zhang X, Xu J, Cai Y, et al. Detecting dependency-related sentiment features for aspect-level sentiment classification. IEEE Transactions on Affective Computing, 2023, 14 (01) : 196-210
- [28] Li R, Chen H, Feng F, et al. Dual graph convolutional networks for aspect-based sentiment analysis//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP). Online, 2021: 6319-6329
- [29] Song Wei, Wen Zi-jian. Feature dual distillation network for aspect-based sentiment analysis. Journal of Chinese Computer Systems, 2021, 35(07): 126-133(in Chinese)  
(宋威, 温子健. 基于特征双重蒸馏网络的方面级情感分析. 中文信息学报, 2021, 35(7): 126-133)
- [30] Xie Jun, Wang Yu-Zhu, Chen Bo, et al. Aspect-based sentiment analysis model with bi-guide attention network. Journal of Computer Research and Development, 2022, 59 (12): 2831-2843(in Chinese)  
(谢珺, 王雨竹, 陈波等. 基于双指导注意力网络的属性情感分析模型. 计算机研究与发展, 2022, 59(12): 2831-2843)
- [31] Yang Chun-xia, Song Jin-jian, Yao Si-cheng. A weighted dependency tree convolutional networks for aspect-based sentiment analysis. Journal of Chinese Computer Systems, 2022, 36(05): 125-132(in Chinese)  
(杨春霞, 宋金剑, 姚思诚. 面向方面级情感分析的加权依存树卷积网络. 中文信息学报, 2022, 36(5): 125-132)
- [32] Ma Yuan-yuan, Yu Long, Tian Sheng-wei, et al. Metaphorical aspect sentiment analysis based on RoBERTa and attention mechanism. Journal of Chinese Computer Systems, 2023, 44 (10): 2236-2241(in Chinese)  
(马圆圆, 禹龙, 田生伟等. 融合RoBERTa和注意力机制的隐喻方面级情感分析. 小型微型计算机系统, 2023, 44(10): 2236-2241)
- [33] Wan Yu-jie, Chen Yu-zhong. Knowledge-enhanced bi-graph convolutional network for aspect-level sentiment analysis. Journal of Chinese Computer Systems, 2024, 45(01): 37-44(in Chinese)  
(万宇杰, 陈羽中. 一种用于方面级情感分析的知识增强双图卷积网络. 小型微型计算机系统, 2024, 45(01): 37-44)
- [34] Zhao Zhi-ying, Shao Xin-hui, Lin Xing. GCN-aware attention networks for aspect-based sentiment analysis. Journal of Chinese Computer Systems, 2022, 36 (07) : 154-163 (in Chinese)  
(赵志影, 邵新慧, 林幸. 用于方面情感分析的结合图卷积神经网络的注意力模型. 中文信息学报, 2022, 36(07): 154-163)
- [35] Fu Chao-yan, Huang Xian-ying, Liu Han-kai, et al. Aspect level sentiment analysis of syntactic feature fusion enhancement. Journal of Chinese Computer Systems, 2023, 44(04): 682-689 (in Chinese)  
(付朝燕, 黄贤英, 刘瀚楷等. 句法特征融合增强的方面级情感分析. 小型微型计算机系统, 2023, 44(04): 682-689)
- [36] Yang Chun-xia, Xu Ben, Chen Qi-gang, et al. Aspect level sentiment analysis model based on deep BiGRU and global graph convolution. Journal of Chinese Computer Systems, 2023, 44(01): 132-139(in Chinese)  
(杨春霞, 徐奔, 陈启岗等. 融合深度BiGRU与全局图卷积的方面级情感分析模型. 小型微型计算机系统, 2023, 44(01): 132-139)
- [37] Li Wei-jiang, Wu Yu-chen. Aspect level sentiment analysis based on syntactic structure and mixed attention mechanism. Journal of Chinese Computer Systems, 2023, 37(05): 143-156 (in Chinese)  
(李卫疆, 吴宇宸. 基于句法结构树和混合注意力网络的方面级情感分类. 中文信息学报, 2023, 37(05): 143-156)
- [38] Liu P, Joty S, Meng H. Fine-grained opinion mining with recurrent neural networks and word embeddings//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP). Lisbon, Portugal, 2015: 1433-1443

- [39] Yin Y, Wei F, Dong L, et al. Unsupervised word and dependency path embeddings for aspect term extraction// Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI). New York, USA, 2016: 2979-2985
- [40] Xu H, Liu B, Shu L, et al. Double embeddings and CNN-based sequence labeling for aspect extraction//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL). Melbourne, Australia, 2018: 592-598
- [41] Ma D, Li S, Wu F, et al. Exploring sequence-to-sequence learning in aspect term extraction//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL). Florence, Italy, 2019: 3538-3547
- [42] Chen Z, Qian T. Enhancing aspect term extraction with soft prototypes//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Online, 2020: 2107-2117
- [43] Wei Z, Hong Y, Zou B, et al. Don't eclipse your arts due to small discrepancies: boundary repositioning with a pointer network for aspect extraction//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL). Online, 2020: 3678-3684
- [44] Li K, Chen C, Quan X, et al. Conditional augmentation for aspect term extraction via masked sequence-to-sequence generation//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL). Online, 2020: 7056-7066
- [45] Zhang Ming-fang, Xiang Yan, Shao Dang-guo, et al. Aspect terms extraction based on multiple attention mechanisms. Journal of Chinese Computer Systems, 2022, 36(03): 136-145 (in Chinese)  
(张名芳, 相艳, 邵党国等. 结合多种注意力机制的方面词提取方法. 中文信息学报, 2022, 36(03): 136-145)
- [46] Yang Y, Zhao H. Aspect-based sentiment analysis as machine reading comprehension//Proceedings of the 29th International Conference on Computational Linguistics (COLING). Gyeongju, Korea, 2022: 2461-2471
- [47] Li X, Bing L, Li P, et al. A unified model for opinion target extraction and target sentiment prediction//Proceedings of the AAAI Conference on Artificial Intelligence (AAAI). Honolulu, Hawaii, USA, 2019: 6714-6721
- [48] Zhou Y, Huang L, Guo T, et al. A span-based joint model for opinion target extraction and target sentiment classification// Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI). Macao, China, 2019: 5485-5491
- [49] Hu M, Peng Y, Huang Z, et al. Open-domain targeted sentiment analysis via span-based extraction and classification// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL). Florence, Italy, 2019: 537-546
- [50] Phan M H, Ogunbona P O. Modelling context and syntactical features for aspect-based sentiment analysis//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL). Online, 2020: 3211-3220
- [51] Liu De-xi, Nie Jian-yun, Wan Chang-xuan, et al. A classification based sentiment words extracting method from microblogs and its feature engineering. Chinese Journal of Computers, 2018, 41(07): 1574-1597(in Chinese)  
(刘德喜, 聂建云, 万常选等. 基于分类的微博新情感词抽取方法和特征分析. 计算机学报, 2018, 41(07): 1574-1597)
- [52] Fan Z, Wu Z, Dai X Y, et al. Target-oriented opinion words extraction with target-fused neural sequence labeling// Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL). Minneapolis, USA, 2019: 2509-2518
- [53] Wu Z, Zhao F, Dai X Y, et al. Latent opinions transfer network for target-oriented opinion words extraction//Proceedings of the AAAI Conference on Artificial Intelligence (AAAI). New York, USA, 2020: 9298-9305
- [54] Pouran Ben Veyseh A, Nouri N, Derroncourt F, et al. Introducing syntactic structures into target opinion word extraction with deep learning//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Online, 2020: 8947-8956
- [55] Yan H, Dai J, Ji T, et al. A unified generative framework for aspect-based sentiment analysis//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP). Online, 2021: 2416-2429
- [56] Chen H, Zhai Z, Feng F, et al. Enhanced multi-channel graph convolutional network for aspect sentiment triplet extraction// Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL). Dublin, Ireland, 2022: 2974-2985
- [57] Liu Xin-yi, Ning Bo, Wang Ming, et al. Fine-grained sentiment triplet extraction method based on syntactic enhancement. Journal of Computer Research and Development, 2023, 60(7): 1649-1660(in Chinese)  
(刘欣逸, 宁博, 王明等. 基于句法增强的细粒度情感三元组抽取方法. 计算机研究与发展, 2023, 60(7): 1649-1660)
- [58] Peng H, Xu L, Bing L, et al. Knowing what, how and why: A near complete solution for aspect-based sentiment analysis// Proceedings of the AAAI Conference on Artificial Intelligence (AAAI). New York, USA, 2020: 8600-8607
- [59] Xu L, Li H, Lu W, et al. Position-aware tagging for aspect sentiment triplet extraction//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Online, 2020: 2339-2349
- [60] Wu Z, Ying C, Zhao F, et al. Grid tagging scheme for aspect-oriented fine-grained opinion extraction//Findings of the Association for Computational Linguistics (FindingsACL). Online, 2020: 2576-2585
- [61] Chen S, Wang Y, Liu J, et al. Bidirectional machine reading comprehension for aspect sentiment triplet extraction// Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI). Online, 2021: 12666-12674

- [62] Mao Y, Shen Y, Yu C, et al. A joint training dual-mrc framework for aspect based sentiment analysis//Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI). Online, 2021: 13543-13551
- [63] Xu L, Chia Y K, Bing L. Learning span-level interactions for aspect sentiment triplet extraction//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP). Online, 2021: 4755-4766
- [64] Oh S, Lee D, Whang T, et al. Deep context- and relation-aware learning for aspect-based sentiment analysis//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP). Online, 2021: 495-503
- [65] Wu C, Xiong Q, Yi H, et al. Multiple-element joint detection for aspect-based sentiment analysis. Knowledge-Based Systems, 2021, 223(01): 107073-107082
- [66] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL). Minneapolis, USA, 2019: 4171-4186
- [67] Qiu G, Liu B, Bu J, et al. Opinion word expansion and target extraction through double propagation. Computational Linguistics, 2011, 37(1): 9-27
- [68] Wan H, Yang Y, Du J, et al. Target-aspect-sentiment joint detection for aspect-based sentiment analysis//Proceedings of the AAAI Conference on Artificial Intelligence (AAAI). New York, USA, 2020: 9122-9129
- [69] Wang W, Pan S J, Dahlmeier D, et al. Coupled multi-layer attentions for co-extraction of aspect and opinion terms//Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI), San Francisco, USA, 2017: 3316-3322
- [70] Bao X, Zhongqing W, Jiang X, et al. Aspect-based sentiment analysis with opinion tree generation//Proceedings of the 31st International Joint Conference on Artificial Intelligence (IJCAI). Vienna, Austria, 2022: 4044-4050
- [71] Zhang N, Li L, Chen X, et al. Differentiable prompt makes pre-trained language models better few-shot learners//Proceedings of the 10th International Conference on Learning Representations (ICLR). Online, 2022: 1-19
- [72] Xiang W, Wang Z, Dai L, et al. Connprompt: Connective-cloze prompt learning for implicit discourse relation recognition//Proceedings of the 29th International Conference on Computational Linguistics (COLING). Gyeongju, Korea, 2022: 902-911
- [73] Schick T, Schütze H. Exploiting cloze-questions for few-shot text classification and natural language inference//Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL). Online, 2021: 255-269
- [74] Gao T, Fisch A, Chen D. Making pre-trained language models better few-shot learners//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP). Online, 2021: 3816-3830
- [75] Cui L, Wu Y, Liu J, et al. Template-Based Named Entity Recognition Using BART//Findings of the Association for Computational Linguistics (FindingsACL). Online, 2021: 1835-1845
- [76] Chen X, Li L, Deng S, et al. LightNER: A Lightweight Tuning Paradigm for Low-resource NER via Pluggable Prompting//Proceedings of the 29th International Conference on Computational Linguistics (COLING). Gyeongju, Korea, 2022: 2374-2387
- [77] Dai L, Wang B, Xiang W, et al. Bi-directional iterative prompt-tuning for event argument extraction//Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP). Abu Dhabi, United Arab Emirates, 2022: 6251-6263
- [78] Ding N, Chen Y, Han X, et al. Prompt-learning for fine-grained entity typing//Findings of the Association for Computational Linguistics (FindingsACL). Abu Dhabi, United Arab Emirates, 2022: 6888 - 6901
- [79] Goswami A, Bhat A, Ohana H, et al. Unsupervised relation extraction from language models using constrained cloze completion//Findings of the Association for Computational Linguistics (FindingsACL). Online, 2020: 1263-1276
- [80] Li Z, Fu L, Wang X, et al. RFBFN: A relation-first blank filling network for joint relational triple extraction//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL). Dublin, Ireland, 2022: 10-20
- [81] Han X, Zhao W, Ding N, et al. PTR: Prompt tuning with rules for text classification. AI Open, 2022, 3(01): 182-192
- [82] Petroni F, Rocktäschel T, Riedel S, et al. Language models as knowledge bases? //Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong, China, 2019: 2463-2473
- [83] Brown T, Mann B, Ryder N, et al. Language models are few-shot learners. Advances in Neural Information Processing Systems, 2020, 33: 1877-1901.
- [84] Shin T, Razeghi Y, Logan IV R L, et al. AutoPrompt: Eliciting knowledge from language models with automatically generated prompts//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Online, 2020: 4222-4235
- [85] 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//Proceedings of the 29th International Conference on Computational Linguistics (COLING). Gyeongju, Korea, 2022: 7002-7012
- [86] Liu S, Li K, Li Z. A robustly optimized BMRC for aspect sentiment triplet extraction//Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL). Seattle, USA,

- 2022; 272-278
- [87] Zhai Z, Chen H, Feng F, et al. COM-MRC: A context-masked machine reading comprehension framework for aspect sentiment triplet extraction//Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP). Abu Dhabi, United Arab Emirates, 2022; 3230-3241
- [88] Lu Y, Lin H, Xu J, et al. Text2Event: Controllable sequence-to-structure generation for end-to-end event extraction//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP). Online, 2021; 2795-2806
- [89] Liu P, Yuan W, Fu J, et al. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 2023, 55(9): 1-35
- [90] Raffel C, Shazeer N, Roberts A, et al. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research*, 21(1): 1532-4435
- [91] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need//Advances in Neural Information Processing Systems (NeurIPS). Long Beach, California, USA, 2017; 5998-6008
- [92] Sui D, Chen Y, Liu K, et al. Joint entity and relation extraction with set prediction networks. *IEEE Transactions on Neural Networks and Learning Systems*, 2023; 1-12
- [93] Seoh R, Birlle I, Tak M, et al. Open aspect target sentiment classification with natural language prompts//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP). Online and Punta Cana, Dominican Republic, 2021; 6311-6322
- [94] Chen P, Bogoychev N, Heafield K, et al. Parallel sentence mining by constrained decoding//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL). Online, 2020; 1672-1678
- [95] Pontiki M, Galanis D, Papageorgiou H, et al. SemEval-2015 task 12: Aspect based sentiment analysis//Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval). Denver, USA, 2015; 486-495
- [96] Pontiki M, Galanis D, Papageorgiou H, et al. SemEval-2016 task 5: Aspect based sentiment analysis//Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval). San Diego, USA, 2016; 19-30



**PENG Wen-Zhong**, Ph. D. candidate, assistant research fellow. His research interests include information extraction and sentiment analysis.

**XIA Jia-Li**, Ph. D., professor, Ph. D. supervisor. Her research interests include data mining and big data processing.

**WAN Qi-Zhi**, Ph. D, lecturer. His current research interests include deep learning, information extraction,

natural language processing, and data mining.

**LIU De-Xi**, Ph. D., professor, Ph. D. supervisor. His research interests include social media processing, information retrieval, and natural language processing.

**WAN Ben-Ting**, Ph. D, associate professor. His research interests include intelligence decision making and intelligence computing.

**CAO Zhong-Hua**, Ph.D, lecturer. His research interests include information extraction, and machine learning.

**XIA Chi-Yu**, Ph.D. candidate, assistant research fellow. His research interests include data mining and time series analysis.

## Background

The aspect sentiment quad prediction (ASQP) is the latest subtask of aspect-based sentiment analysis (ABSA). It aims to extract all aspect-level sentiment elements from review sentences, including aspect terms, aspect categories, opinion terms and sentiment polarities. The ASQP task contributes to a comprehensive understanding of how users evaluate different aspects of a product or service.

Existing ASQP methods suffer from the following limitations: (1) Discriminative models fail to capture the semantic relations between sentiment elements using prompts; (2) Generative models either merely combine sentiment element type labels to form prompts, lacking contextual understanding of

label semantics, or use a discrete template as input to the decoder, preventing the encoder from capturing the semantic relations between sentiment elements in the template.

This paper proposes two types of prompts, discrete and continuous, based on the cloze-style methodology. These prompts provide context for understanding the semantics of four sentiment element types, aiding the model in capturing semantic relations between sentiment elements more effectively. To better utilize the prompts, we introduce the C-ASQP framework, consisting of the discriminative model DC-ASQP and the generative model GC-ASQP. Experimental results demonstrate that DC-ASQP significantly outperforms similar discriminative



baseline models, and GC-ASQP surpasses all baseline models. These results not only validate the effectiveness of modeling ASQP as cloze tasks but also prove the efficacy of the designed two types of prompts and the C-ASQP framework.

This team has conducted extensive research in structured information extraction, with a primary focus on sentiment analysis and event extraction. Their research has been published in renowned academic journals and conferences, including the Chinese Journal of Computers, ACM Transactions on Information Systems, Association for Computational Linguistics, Information Sciences, and other prestigious publications.

The work is supported by the National Natural Science

Foundation of China (No. 62272206, No. 62272205, No. 62076112), the Jiangxi Provincial Academic and Technical Leaders Training Program for Leading Talents in Major Disciplines (No. 20213BCJL22041), the Natural Science Foundation of Jiangxi Province (No. 20212ACB202002), the Science and Technology Research Project of Jiangxi Provincial Education Department (No. GJJ2200560, No. GJJ2200501), the Social Science Foundation of Jiangxi Province (No. 23TQ02), and the Humanities and Social Sciences Research Project of the Ministry of Education (No. 22YJA880051). These projects mainly focus on information extraction and data mining for texts in the psychological or e-commerce fields, in which ABSA plays an essential role.