



# 计算机科学

COMPUTER SCIENCE

## 基于prompt和知识增强的方面级情感分析

李阳, 唐积强, 朱俊武, 梁明轩, 高翔

引用本文

李阳, 唐积强, 朱俊武, 梁明轩, 高翔 [基于prompt和知识增强的方面级情感分析](#) [J]. 计算机科学, 2023, 50(6A): 220300279-7.

LI Yang, TANG Jiqiang, ZHU Junwu, LIANG Mingxuan, GAO Xiang. [Aspect-based Sentiment Analysis Based on Prompt and Knowledge Enhancement](#) [J]. Computer Science, 2023, 50(6A): 220300279-7.

---

## 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

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

### [基于多特征融合的GRU-LSTM大学生就业动态预测](#)

College Students Employment Dynamic Prediction of Multi-feature Fusion Based on GRU-LSTM  
计算机科学, 2023, 50(6A): 220500056-6. <https://doi.org/10.11896/jsjx.220500056>

### [基于深度学习的超高频标签识别系统](#)

Tag Identification for UHF RFID Systems Based on Deep Learning

计算机科学, 2023, 50(6A): 220200151-6. <https://doi.org/10.11896/jsjx.220200151>

### [CT影像阶段化目标检测方法研究](#)

Study on Phased Target Detection in CT Image

计算机科学, 2023, 50(6A): 220200063-10. <https://doi.org/10.11896/jsjx.220200063>

### [基于深度学习的摩托车车道实时检测](#)

Real-time Detection of Motorcycle Lanes Based on Deep Learning

计算机科学, 2023, 50(6A): 220200066-5. <https://doi.org/10.11896/jsjx.220200066>

### [基于改进YOLOv5的电动车头盔佩戴检测算法](#)

Electric Bike Helment Wearing Detection Alogrithm Based on Improved YOLOv5

计算机科学, 2023, 50(6A): 220500005-6. <https://doi.org/10.11896/jsjx.220500005>

# 基于 prompt 和知识增强的方面级情感分析

李阳<sup>1,2</sup> 唐积强<sup>3</sup> 朱俊武<sup>1</sup> 梁明轩<sup>1,2</sup> 高翔<sup>1,2</sup>

1 扬州大学信息工程学院 江苏 扬州 225000

2 中国科学院计算技术研究所 北京 100190

3 国家计算机网络应急技术处理协调中心 北京 100029

(1037759173@qq.com)

**摘要** 方面级情感分析是一种新兴的细粒度情感分析任务,旨在根据给定句子和方面词判断情感极性。目前广泛使用的预训练语言模型由于训练目标和方面级情感分析的目标有差异,分析结果不好。为了缓解预训练语言模型和情感分析目标的差异,prompt 被引入到方面级情感分析中,采用伪标签加方面词和意见词的方式创建 prompt 连续模板,并使用 prompt-encoder 训练伪标签使其拥有语义信息;然后,使用主题图注意力机制融合关于方面词和意见词的外部知识,根据融合外部知识的隐藏向量预测由情感词典组成的候选标签词;最后,采用求和置信度分数的方式将候选标签词的概率映射到情感极性分布空间上。实验表明,该模型在 SemEval 2014 任务的笔记本电脑数据集和餐厅数据集上将正确率分别提高了 1.53% 和 3.5%。

**关键词** 方面级情感分析;预训练语言模型;prompt;情感词典;知识增强;深度学习

中图法分类号 TP391

## Aspect-based Sentiment Analysis Based on Prompt and Knowledge Enhancement

LI Yang<sup>1,2</sup>, TANG Jiqiang<sup>3</sup>, ZHU Junwu<sup>1</sup>, LIANG Mingxuan<sup>1,2</sup> and GAO Xiang<sup>1,2</sup>

1 College of Information Engineering, Yangzhou University, Yangzhou, Jiangsu 225000, China

2 Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China

3 National Computer Network Emergency Response Technical Team/Coordination Center of China, Beijing 100029, China

**Abstract** sentiment analysis is an emerging fine-grained sentiment analysis task that aims to judge sentiment polarity based on given sentences and aspect words. Currently widely used pre-trained language models are different due to their training objectives and those of aspect-based sentiment analysis, resulting in poor analysis results. In order to alleviate the difference between the pre-trained language model and the sentiment analysis target, prompt is introduced into aspect-based sentiment analysis, using pseudo-labels plus aspect words and opinion words to create prompt continuous templates, and using prompt-encoder to train pseudo-labels to have semantic information; then, use the topic graph attention mechanism to fuse external knowledge about aspect words and opinion words, and predict candidate label words composed of sentiment dictionaries according to the hidden vector fused with external knowledge; The probabilities of label words are mapped onto the sentiment polarity distribution space. Experiments show that the model improves the accuracy by 1.53% and 3.5% on the Laptops dataset and Restaurants dataset of the SemEval 2014 task.

**Keywords** Aspect-based sentiment analysis, Pretrained language model, prompt, Sentiment dictionary, Knowledge enhancement, Deep learning

## 1 引言

随着科技的发展,互联网与每个人的生活紧密相关,每天互联网上都会产生大量的带有情感的文本,这些文本具有重大的价值,了解和分析这些文本中的情感信息可以对产品优化、决策支持提供重大帮助。文本情感分析就是对文本中的

情感信息进行识别和分析的研究。根据研究的文本粒度的不同,情感分析可以分为篇章级、句子级和方面级。其中,方面级情感分析(Asspect-based Sentiment Analysis)旨在提取文本中的方面 aspect 并识别出每个方面的情感极性。根据是否预设方面词,方面级情感分析可以分为方面术语情感分析(Asspect-term Sentiment Analysis)和方面类别情感分析(Asspect-

基金项目:国家 242 信息安全计划项目(2021A008);北京市科技新星计划交叉学科合作课题(Z191100001119014);国家重点研发计划重点专项(2017YFC1700300,2017YFB1002300);国家自然科学基金(61702234);江苏省(扬州大学)研究生科研与实践创新计划项目(SJXC21\_1551)

This work was supported by the National 242 Information Security Program(2021A008), Beijing NOVA Program(Z191100001119014), National Key Research and Development Program of China (2017YFC1700300, 2017YFB1002300), National Natural Science Foundation of China (61702234) and Postgraduate Research & Practice Innovation Program of Jiangsu Province(Yangzhou University)(SJXC21\_1551).

通讯作者:唐积强(tjq@cert.org.cn)

category Sentiment Analysis), 其中方面术语情感分析中的方面词属于文本, 方面类别情感分析中的方面词是预先设置的方面类别。对于图 1 中的实例, 方面术语情感分析首先需要提取方面词“interior decoration”和“chefs”, 然后识别情感极性; 方面类别情感分析需要识别“ambience”和“food”的情感极性。本文研究的主要内容是根据给定的文本和方面术语识别情感极性。

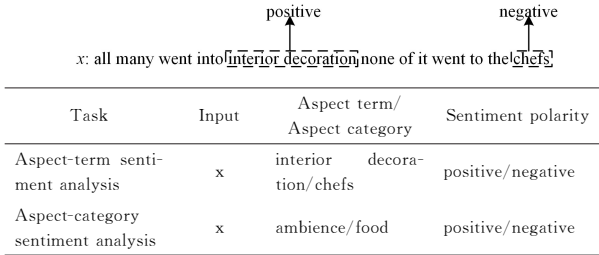
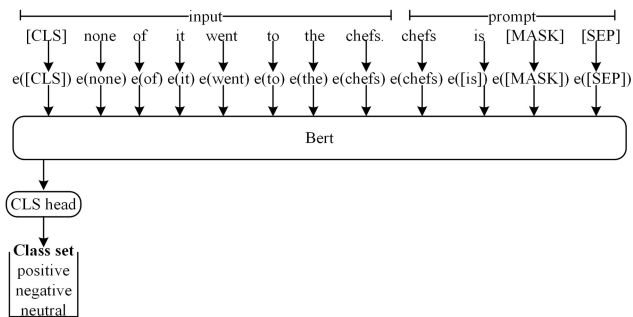


图 1 方面级情感分析的例子

Fig. 1 Example of aspect-based sentiment analysis

近年来, 预训练语言模型 (Pre-trained Language Model, PLM) 已经被广泛应用到自然语言处理任务中, 并且取得了较好的效果。比较常见的预训练模型包括 Bert, GPT, T5 和基于 Bert 改进的 RoBERTa<sup>[1-4]</sup>。这些预训练语言模型采用自监督学习的策略, 通过遮蔽文本中的一些词然后进行预测的方法获取文中的语言信息、语义信息、句法信息和世界知识<sup>[5-8]</sup>。因此, 只需要少量特定于任务的数据微调, 这些经过大量语料学习的预训练语言模型就能表现出惊人的性能。目前自然语言处理研究者的共识就是如何针对特定任务微调预训练语言模型, 而不是从头开始学习模型。图 2 给出了基于预训练语言模型微调范式。在方面级情感分析领域, Chen 等<sup>[9]</sup>和 Yan 等<sup>[10]</sup>也尝试将预训练语言模型应用到任务中, 并构建了统一框架。尽管语言模型和情感分析任务的融合研究已有长足发展, 但两者在客观形式上的巨大差距使得预训练语言模型学习到的知识无法被充分利用。



预训练语言模型微调的例子

Fig. 2 Example of pretrained language model fine-tuning

Prompt 微调是一种被用来弥合预训练任务和特定下游任务的新微调范式<sup>[11-14]</sup>。Prompt 微调由一个模板和一组候选标签词组成, 通过构建一个基于特定任务并带有 [MASK] 的模板, 将原本的下游分类任务转换为预测 [MASK] 的完形填空任务。预测遮蔽词的任务与预训练语言模型的任务一致可以充分激发模型的潜力, 预测出 [MASK] 位置的候选标签词, 再利用特定的映射函数将候选标签词转换成分类类别, 达到分类的效果。图 3 表现了 prompt 的微调范式。Prompt

微调为上游预训练任务和下游特定任务构建了桥梁, 改善了任务的效果, 但是构建契合特定任务的模板需要大量的人工成本, 并且具有固定词的模板在知识迁移过程中不能实现很好的效果, 因此如何自动构建模板并且提高性能是研究者的重要研究方向。Liu 等<sup>[15]</sup>提出 P-tuning, 该模型用伪标签和文本中的重要词构建模板, 实验证明, 经过 prompt-encoder 的训练, 伪标签也能达到人工构建模板的效果。

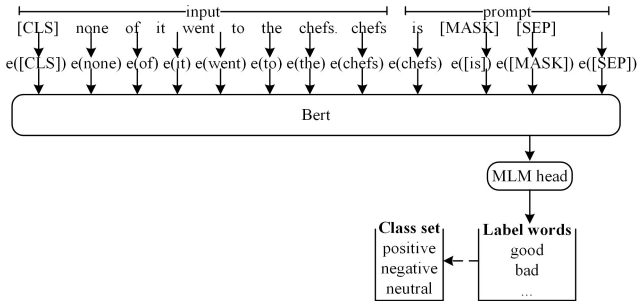


图 3 prompt 微调范式的例子

Fig. 3 Examples of prompt-tuning

基于以上问题和现状, 本文提出了一种基于 prompt 和知识增强的方面级情感分析方法 (PKABSA), 将 P-tuning 应用到方面级情感分析任务中, 利用方面词、意见词和一组伪标签构建模板, 将情感词典作为 [MASK] 位置的候选标签词集。在预测候选标签词时, 融合知识图谱中和方面词、意见词相关三元组的语义信息引导候选标签词的选择。获得候选标签词的概率后, 通过求和置信度分数的方式得到情感极性的概率。为了测试本文提出模型的有效性, 在 SemEval-2014 Task 4 的公开数据集 Laptops 和 Restaurants 上进行实验。

本文的主要贡献可以总结如下:

- (1) 将 P-tuning 模型应用到方面级情感分析中, 利用可训练的连续模板微调方面级情感分析任务;
- (2) 将知识图谱融入模型, 帮助预测候选标签词, 融合知识图谱的先验知识使得隐藏向量拥有更多的语义信息, 实现更好的预测效果;
- (3) 将情感词典作为候选标签词集, 可以明确文本和情感极性之间的关系, 利用情感知识对候选标签词到情感极性的映射起到优化作用。

## 2 相关工作

### 2.1 方面级情感分析

方面级情感分析的主要目标是判断文本中特定方面词的情感极性。早期的方面级情感分析包括基于情感词典、逻辑规则和基于机器学习的建模方法。Ding 等<sup>[16]</sup>统计每个评论句子中靠近产品特征的积极和消极观点词的数量, 通过比较两者数量实现情感分析。Nguyen 等<sup>[17]</sup>应用基于构成树和依赖树的两个树核组合的新树核进行方面-观点关系提取, 通过自动识别方面-观点关系的算法进一步优化了基于情感词典的方面情感分类。Lipenkova 等<sup>[18]</sup>利用语言知识构建规则并融入到基于概率模型和情感词典的中文方面级情感分析中。Kiritchenko 等<sup>[19]</sup>将方面级情感分析任务建模成多分类任务, 并训练了一个线性 SVM 分类器。Rao 等<sup>[20]</sup>为方面级情感

分析任务提出一种半监督学习方法,他们定义了一个图,其节点由情感分类任务相关的单词或短语组成,边则编码了一些相似性的概念;然后制定规则,先标记一些节点种子示例,再根据这些种子导出其他节点的标签。

随着时代的发展,深度学习被广泛应用到方面级情感分析中。Sun 等<sup>[21]</sup>采用双向长短期记忆网络提取文本序列的词特征,利用图卷积神经网络提取文本的区域依存特征,最后选取方面词做平均池化操作,得到最终向量,进行情感分类。Wang 等<sup>[22]</sup>提出将方面级情感分析文本构建成基于特定方面的依赖树,并融合图卷积和注意力机制提取语义特征进行分类。Tang 等<sup>[23]</sup>发现注意力机制过度关注高频词,对低频词缺少关注,他们提出了一种新的增量监督注意力机制,并取得了更好的实验效果。

除了上述对方面级情感分析任务模型本身的优化,很多研究者也开始考虑通过引入其他相关任务与方面情感分析任务进行联合优化,以实现更好的实验效果。He 等<sup>[24]</sup>将方面级情感分析建模成一个序列标注问题,实现对方面抽取和方面情感分类的联合优化。Wu 等<sup>[25]</sup>构建了一个方面抽取和方面情感分类的联合模型,其中方面抽取被建模成序列标注任务,方面情感分析被建模成分类任务。Li 等<sup>[26]</sup>认为方面类别的情感是指示方面类别的词的情感的聚合,并基于此构建了一个多任务联合优化模型 AC-MIMLLN。Chen 等<sup>[27]</sup>认为方面抽取、意见抽取和方面情感分类 3 个子任务之间各有关系,他提出了关系感知协作学习模型(RACL),并为 3 个子任务定义了 4 种关系。

## 2.2 prompt 微调

随着各种预训练语言模型的兴起,自然语言处理任务的处理模式逐渐转变成预训练语言模型加微调范式。起初一些研究者会为下游任务添加一些提示语句用于充分激发预训练语言模型的语义知识,逐渐这些提示语句发展成 prompt 微调的范式。最早 Brown 等<sup>[28]</sup>提出微调范式需要大量数据训练的问题,他们通过加大模型和利用语言提示作为上下文的方式实现了更好的实验效果。受 GPT3 的启发,Petroni 等<sup>[29]</sup>将关系抽取任务转换成完形填空任务,并利用语言模型预测关系证明语言模型在完形填空任务中的有效性。在推断常识知识领域,Feldman 等<sup>[30]</sup>利用语言模型预测关系三元组中的遮蔽词,并证明这种无监督的方法有更好的泛化能力。在关系分类中,Han 等<sup>[31]</sup>提出将 prompt 和逻辑规则结合,首先为每个逻辑构建独立的 prompt,再利用规则组合这些 prompt 预测候选标签词。虽然人工构建的 prompt 能达到良好的效果,但是成本巨大,因此如何自动构建有效的 prompt 也得到广泛探索。Liu 等<sup>[32]</sup>探索了构建自动连续模板的可能性,为了保证有效性,他们用 prompt-encoder 训练伪标签。在生成任务中,Li 等<sup>[33]</sup>提出前缀微调,为特定生成任务搭配前缀向量序列,通过微调前缀实现优化模型的效果。Schick 等<sup>[34]</sup>将 prompt 应用到文本分类领域,并设计一种可以在给定少量训练数据的情况下自动找到映射的方法。

## 3 模型方法

本节首先对方面级情感分析的任务进行描述,然后根据

数据流动的方向介绍整个模型,包括 prompt-encoder 和主题图注意力机制。

### 3.1 任务描述

给定文本  $x = \{\omega_1, \omega_2, \dots, a, \dots, o, \dots, \omega_n\}$ , 其中  $a$  表示方面词,  $o$  表示意见词, 模型输出方面词  $a$  对应的情感极性  $y \in \{\text{positive}, \text{negative}, \text{neutral}\}$ 。在我们的模型中, 输入是文本  $x$  和 prompt 模板, 预训练语言模型 Bert 被用于生成和训练词向量。

### 3.2 方法动机

基于预训练模型的微调作为上游任务与方面情感分析在任务目标上存在差异, 预训练语言模型以预测遮蔽词为训练目标, 方面情感分析则以预测方面词的情感极性为目标。基于 prompt 的微调与基于预训练语言模型的微调不同。假设模型的数据集为  $X$ , 其中每条文本为  $x$ 。类集为  $Y$ , 每个类别为  $y$ , 预训练语言模型为  $M$ 。在普通微调范式中, 首先对输入文本和特殊字符进行拼接, 得到输入序列  $x = \{[\text{CLS}], \omega_1, \omega_2, \dots, \omega_n, [\text{SEP}]\}$ , 然后将  $x$  输入预训练语言模型  $M$  中, 得到具有语义信息的隐藏向量表示  $\{h_{[\text{CLS}]}, h_1, \dots, h_n, h_{[\text{SEP}]}\}$ 。对于分类任务, 通常选择  $[\text{CLS}]$  的隐藏向量  $h_{[\text{CLS}]}$  作为模型  $M$  输出的最终表示, 使用 softmax 函数计算类集  $Y$  关于输入文本  $x$  的概率分布。Softmax 函数公式如下:

$$p(y|x) = \text{softmax}(W h_{[\text{CLS}]} + b) \quad (1)$$

其中,  $W$  和  $b$  是可学习的参数,  $W, b$  和  $M$  的所有参数通过最大化目标函数来调整。目标函数公式如下:

$$\text{argmax} \left( \frac{1}{|X|} \sum_{x \in X} \log p(y|x) \right) \quad (2)$$

在基于 prompt 微调范式中, 为了将分类任务转换成完形填空任务, 形式上, 我们需要构建一个模板  $T$ 、一组候选标签词集  $V$  和映射函数  $\mathcal{O}$ 。首先将文本和模板  $T$  拼接起来得到  $x_{\text{prompt}}$ , 该模板由一组固定的词和至少一个  $[\text{MASK}]$  组成, 其中  $[\text{MASK}]$  位置的词就是完形填空需要预测的词。通过将  $x_{\text{prompt}}$  输入预训练语言模型  $M$  得到  $[\text{MASK}]$  位置的隐藏向量  $h_{[\text{MASK}]}$ , 然后根据隐藏向量  $h_{[\text{MASK}]}$  计算所有候选标签词  $v$  的概率分布, 其中候选标签词  $v$  来自候选标签词集  $V$ 。为了桥接完形填空任务和分类任务, prompt 微调中还存在映射函数  $\mathcal{O}$ , 其连接了类集和候选标签词集, 将候选标签词的概率分布映射到类集的概率分布上, 因此学习目标可以形式化为以下公式:

$$\text{argmax} \left( \frac{1}{|X|} \sum_{x \in X} \log p([\text{MASK}] = \mathcal{O}(y) | x_{\text{prompt}}) \right) \quad (3)$$

通过普通微调范式和基于 prompt 的微调范式的流程, 可以清楚地看出两者之间的联系和差异。

显然, 基于 prompt 微调范式更容易实现预训练模型和方面情感分析的融合, 因此更适合方面级情感分析。受此启发, 本文提出了新的模型结构。

### 3.3 模型结构

本文的模型主要包括文本词嵌入层、prompt-encoder、融合外部知识的编码层和基于情感词典的输出层 4 个部分, 模型的具体结构如图 4 所示。

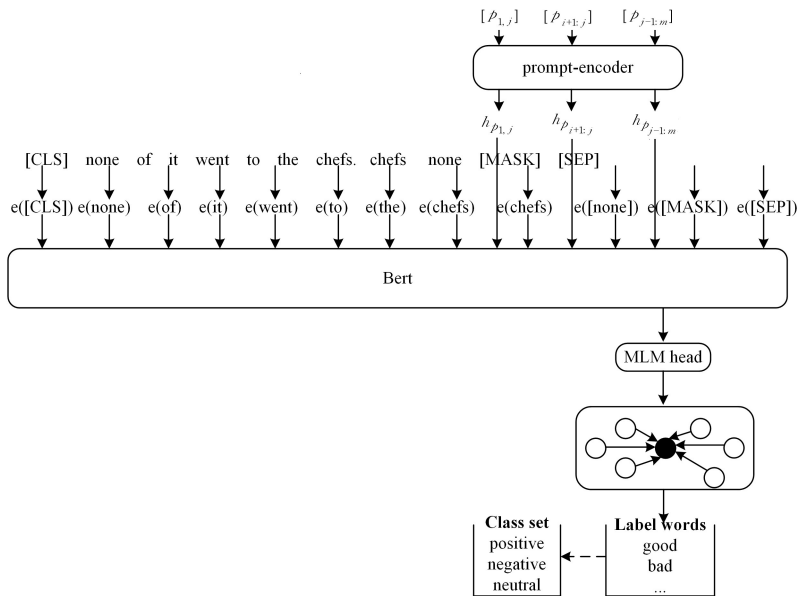


图4 基于 prompt 和知识增强的方面级情感分析模型的流程图

Fig. 4 Flow chart of aspect-based sentiment analysis model based on prompt and knowledge enhancement

对比图3和图4,图3采用人工构建的 prompt 模板;本文采用连续模板,由一组可训练伪标签和方面词、意见词组成,可以实现更好的效果。

### 3.3.1 文本词嵌入层

本文的输入  $x_{\text{prompt}}$  包括两个部分,分别是文本  $x$  和 prompt 模板  $T$ 。受 P-tuning 的模型启发<sup>[15]</sup>,其中模板  $T$  由可训练伪标签  $P_k$  和  $x$  的方面词与意见词  $a$  和  $o$  组成,  $P_k$  是一个随机初始化的待优化伪标签。本文采用的预训练语言模型是 Bert,因此还需要在头尾分别加上两个特殊标签 [CLS], [SEP]。  $x_{\text{prompt}}$  的构成如下:

$$\{[\text{CLS}], w_1, w_2, \dots, a, \dots, o, \dots, w_n, [p_{1,i}], a, [p_{i+1,j}], o, [p_{j+1,m}], [\text{MASK}], [\text{SEP}]\} \quad (4)$$

其中,意见词  $o$  与方面词  $a$  对应,例如在输入句子“All the money went into the interior decoration, none of it went to the chefs.”中,  $a$  为“chefs”,与  $a$  对应的  $o$  为“none”,  $a$  为“interior decoration”,与  $a$  对应的  $o$  为“all”。伪标签  $P_k$  由 prompt-encoder 训练,以获取更多的语义信息。

Bert 中的词嵌入层  $e$  将  $x_{\text{prompt}}$  映射成词矩阵  $e(x_{\text{prompt}})$ 。同时随机初始化伪标签  $P_k$  并通过 prompt-encoder 生成隐藏向量  $h_k$ ,其中  $h_k$  和  $e(x_{\text{prompt}})$  中的词向量维度一样,根据位置信息将  $h_k$  插入到  $x_{\text{prompt}}$  对应的伪标签位置。

### 3.3.2 prompt 编码层

本文采用可训练的伪标签构建连续模板,这种方式缓解了人工构建模板的大量成本,但是随机生成的伪标签缺少语义信息,并且和输入文本、情感分析任务的目标关联性很差。为了使伪标签获得更多的语义信息,增加与文本、任务目标的相关性,我们使用 p-tuning 模型中构建的 prompt-endoer 训练伪标签。prompt-encoder 由两层感知机 (MLP) 组成,每层感知机由双向长短期记忆模型 (BiLSTM) 和激活函数 RELU 构成。

### 3.3.3 融合外部知识的编码层

深度学习模型与外部知识的结合越来越紧密,利用外部知识可以提高深度学习模型的效果<sup>[35]</sup>。在方面级情感分类任务中,结合方面词  $a$  和意见词  $o$  的相关信息可以实现优化

模型的效果。本文从 wordnet 知识库中提取出两组与方面词和意见词有关的同义词集合  $syna$  和  $syno$ 。本文采用主题图注意力机制的方式融合同义词的语义信息。具体公式如下:

$$\beta_i = \begin{cases} \frac{(W_1 h_{[\text{MASK}]}) (W_2 syna_i)^T}{\sqrt{d}}, & syna_i \in syna \\ \frac{(W_1 h_{[\text{MASK}]}) (W_3 syno_i)^T}{\sqrt{d}}, & syno_i \in syno \\ \frac{(W_1 h_{[\text{MASK}]}) (W_4 h_{[\text{MASK}]})^T}{\sqrt{d}} \end{cases} \quad (5)$$

$$\alpha_i = \frac{\exp(\beta_i)}{\sum_{j \in \text{mm}} \exp(\beta_j)} \quad (6)$$

$$g_{[\text{MASK}]} = \sum_{i \in \text{mm}} \alpha_i syn_i \quad (7)$$

其中,  $h_{[\text{MASK}]}$  表示经过预训练语言模型 Bert 编码后的具有语义信息的 [MASK] 位置的隐藏向量,  $syna_i$  表示方面词  $a$  第  $i$  个同义词,  $syno_i$  表示意见词  $o$  第  $i$  个同义词,  $W_1, W_2, W_3, W_4$  是可训练参数,  $g_{[\text{MASK}]}$  表示融合外部知识的隐藏向量。  $d$  表示隐藏向量的维度,由于向量点积结果的数字过大,本文参考 transformer 中自注意力机制的方式,通过除以  $\sqrt{d}$  的方式对结果进行缩小。另外,本文还将隐藏向量  $h_{[\text{MASK}]}$  也加入注意力计算中,这种方式可以动态控制是否需要引入外部知识,控制隐藏向量  $h_{[\text{MASK}]}$  对于外部知识的注意力,平衡预训练语言模型训练出的语义信息和外部知识中的语义信息。

### 3.3.4 基于情感词典的输出层

本文采用情感词典作为 prompt 中的候选标签词集  $V$ 。融合外部知识的编码层生成隐藏向量  $g_{[\text{MASK}]}$ 。  $g_{[\text{MASK}]}$  通过神经网络全连接层转换为维度为  $|V|$  的向量,用 softmax 函数对其做归一化,获得每个候选标签词的概率  $P([\text{MASK}] = v_i | x_{\text{prompt}})$ ,其中  $v_i$  代表标签候选词集  $V$  的第  $i$  个词。最后,将候选标签词的概率通过求和置信度分数的方式映射到情感极性上。具体公式如下:

$$P(y|x) = \sum_{i \in |V_y|} \lambda_i P([\text{MASK}] = v_i | x_{\text{prompt}}) \quad (8)$$

其中,  $\lambda$  表示候选标签词是  $v_i$  的概率的置信度,  $V_y$  表示候选标签词中属于情感极性  $y$  的所有词的集合。本文的损失函数使用

交叉熵,通过最小化交叉熵误差训练模型,具体公式如下:

$$Loss = - \sum_{j \in |X|} \sum_{i \in |Y|} y_i \log P(y_i | x) \quad (9)$$

其中,  $|X|$  表示样本数量,  $|Y|$  表示情感极性数量。

## 4 实验

### 4.1 实验数据集及实验设置

本文采用两个公开的方面级情感分析数据集进行实验,分别是 SemEval 2014 任务中的笔记本电脑(Laptops)和餐厅(Restaurants)评论数据集<sup>[39]</sup>,包含 3 个情感极性:积极、消极和中立。另外,为了获取与方面词对应的意见词,我们收集 peng 等标注的意见词信息<sup>[36]</sup>。但是据我们观察,两个数据集之间的数据有差异,因此我们将 peng 标注的意见词整合到笔记本电脑和餐厅数据集中。两个数据集的具体统计数据如表 1 所列。

表 1 实验数据集

Table 1 Experimental datasets

dataset	positive		neutral		negative	
	Train set	Test set	Train set	Test set	Train set	Test set
Laptops	994	341	870	128	464	169
Restaurants	2164	728	807	196	637	196

本文采用 wordnet 知识库提取方面词和情感词的同义词集。采用的深度模型框架为 pytorch,生成词向量采用预训练语言模型 Bert,生成的词向量维度为 786,训练和测试的 batch\_size 为 8,学习衰退率(decay\_rate)为 0.98,权重衰退(weight\_decay)为  $1 \times 10^{-2}$ ,模型训练的学习率(Learning\_rate)为  $5 \times 10^{-6}$ 。

### 4.2 评价指标

采用正确率(Acc)和宏平均的 F1 值(Macro-F1)作为模型的评价指标。具体公式如下:

$$Acc = \frac{predict\_true\_num}{total\_num} \quad (10)$$

$$Macro-F1 = \frac{1}{|Y|} \sum_{i \in |Y|} F1_i \quad (11)$$

$$F1_i = 2 \frac{Recall_i \times Precision_i}{Recall_i + Precision_i} \quad (12)$$

$$Precision_i = \frac{predict\_true\_num_i}{predict\_num_i} \quad (13)$$

$$Recall_i = \frac{predict\_true\_num_i}{true\_num_i} \quad (14)$$

其中,  $true\_num_i$  表示实际属于情感极性  $i$  的数量,  $predict\_num_i$  表示预测结果是情感极性  $i$  的数量,  $predict\_true\_num_i$  表示正确预测情感极性  $i$  的数量。

### 4.3 比较实验

选取如下模型进行对比。

(1)LSTM<sup>[37]</sup>:通过两个依赖目标的 LSTM 整合目标词和上下文的关系进行情感分类。

(2)TD-LSTM<sup>[38]</sup>:该模型是对 LSTM 模型的改进,由两个 LSTM 组成,分别编译方面词的左右上下文合并进行情感分类。

(3)ATAE-LSTM<sup>[39]</sup>:最早将 LSTM+Attention 应用到方面级情感分析中的模型,将文本和方面词输入 LSTM 学习语义信息,再利用 Attention 关注重要的词,实现更好的分类效果。

(4)IAN<sup>[40]</sup>:提出交互式注意力网络来交互式学习上下文和目标中的注意力进行情感分类。

(5)MemNet<sup>[41]</sup>:利用深度记忆网络捕捉每个上下文词的重要性进行情感分类。

(6)RAM<sup>[42]</sup>:应用新的多重注意力机制提取句子结构中的特征进行情感分类。

(7)Cabasc<sup>[43]</sup>:提出能够全局角度捕获信息的内容注意力机制和能够提取单词顺序、相关性的方面注意力机制模型,该模型善于处理多方面和句法复杂的句子。

为了证明模型的有效性,对比这些模型的实验数据,下列模型的实验数据来自 Deng 的相关论文<sup>[44]</sup>。

如表 2 所列,PKABSA 实现了最佳性能,在两个数据集上的正确率分别达到 76.6% 和 84.04%,Macro-F1 分别达到 73.48 和 76.82。PKABSA 将 prompt 微调的方式应用到方面级情感分类任务中,这样可以更最大限度地开发预训练语言模型的学习能力。用情感词典作为候选标签词集,一方面引导 Bert 将[MASK]预测成情感词;另一方面,情感词与情感极性有明确的映射关系,通过求和置信度的方式映射到情感极性空间上。最后,将由方面词和情感词的同义词集组成的外部知识融入到[MASK]的隐藏向量中,也会帮助模型预测出正确的情感极性。

表 2 比较模型在两个公开数据集上的实验结果

Table 2 Experimental results of comparative models on two public datasets

Comparative models	Laptops		Restaurants	
	Acc/%	Macro-F1	Acc/%	Macro-F1
LSTM	66.77	61.78	74.29	62.58
TD-LSTM	67.71	60.25	75.36	64.48
ATAE-LSTM	68.50	61.52	77.32	64.99
IAN	72.10	63.10	78.60	67.40
MemNet	71.79	67.07	79.96	69.09
RAM	74.49	71.35	80.23	70.80
Cabasc	75.07	70.13	80.54	70.76
PKABSA	76.60	73.48	84.04	76.82

在所有比较模型中,Cabasc 模型在两个数据集的正确率分别达到 75.07% 和 80.54%,这是因为其提出了能够在全局角度捕获信息的文档注意力机制,这种注意力可以缓解在长文本序列中记忆模型无法学习正确语义信息的问题,因此效果比单纯改进 LSTM+Attention 的 ATAE-LSTM,IAN,MemNet 和 RAM 要好。

LSTM 和 TD-LSTM 在两个数据集的正确率分别达到 66.77%,74.29% 和 67.71%,75.36%。其中 TD-LSTM 改进了 LSTM 模型,但是因为 LSTM 无法体现方面词和文本句子的交互信息,另外 LSTM 按文本的序列顺序处理句子,学习到的语义信息不够全面,并且太长的句子会造成梯度下降缓慢的问题。因此,这两个模型的实验效果与其他模型有差距。

### 4.4 模型模块分析

#### 4.4.1 情感词典和主题图注意力机制的影响

本节讨论情感词典和主题图注意力对模型实验效果的影响。如表 3 所列,模型的基本结构是预训练语言模型 Bert 加 prompt,在两个数据集上的正确率和 Macro-F1 分别达到 72.02%,71.12 和 79.81%,70.74。加入情感词典后,两个数据集的正确率和 Macro-F1 分别提高了 3.21%,1.58% 和 1.75,1.75。prompt 最大的作用是将下游的情感分类任务

转变成完形填空任务,因此[MASK]位置的候选标签词集会对 prompt 产生影响,情感词典和情感分类任务高度相关,将情感词典作为候选标签词集,预训练语言模型 Bert 预测的情感词会帮助下游情感极性的识别。另外,情感词典中的词都带有情感极性,可以直接映射成情感极性,因此我们对情感词典中的词使用求和置信度的方式将概率映射到情感极性的概率空间上,可以实现优化模型的效果。

表3 情感词典和主题图注意力对模型的影响

Table 3 Effects of sentiment dictionary and topic graph attention on models

model	Laptops		Restaurants	
	Acc/%	Macro-F1	Acc/%	Macro-F1
Bert+prompt	72.02	71.12	79.81	70.74
Bert+prompt+SD	75.23	72.87	81.39	72.49
Bert+prompt+SD+TGA	76.60	73.48	84.04	76.82

Bert+prompt+SD+TGA 模型在两个数据集上的正确率和 Macro-F1 分别达到 76.6%, 73.48 和 84.04%, 76.82。与不加入主题图注意力的模型相比,正确率和 Macro-F1 分别提高了 1.37%, 2.65% 和 0.61, 4.33。这是因为主题图注意力将外部知识融入到模型中,首先通过 wordnet 知识库获取句子中方面词和意见词的同义词集,通过主题图注意力将这些外部知识融入到模型的隐藏向量中,并且为了平衡外部知识和模型学习到的知识,将隐藏向量本身也加入到注意力计算中。通过注意力机制,模型可以选择对正确情感分类有帮助的知识。

#### 4.4.2 prompt 模板对模型的影响

本节讨论不同长度的 prompt 对方面级情感分析任务的影响,在两个公开数据集上对不同的 prompt 进行了实验。我们简化了 prompt 模板的训练,将模板 3 个位置伪标签的数量固定成同样长度。假设 3 个位置的伪标签数量为相同的  $N$ ,对  $N=0, 1, 2, 3$  进行比较。实验结果如图 5 所示,当一个  $N=0, 1, 2, 3$  时,更多的伪标签提供更多可训练的参数,增强了模型的能力,但需要更多的数据和时间来正确训练,最终本文选择  $N=3$  作为模型伪标签数量。

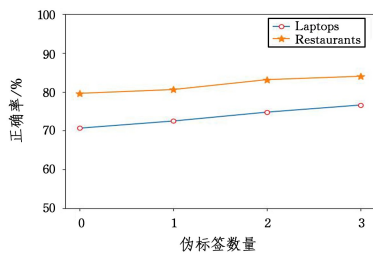


图5 两个数据集在不同伪标签数量下的正确率

Fig. 5 Accuracy of two datasets with different numbers of pseudo-labels

**结束语** 本文提出了一个基于外部知识和 prompt 的方面级情感分析模型。该模型使用可训练的伪标签、方面词和意见词构建模板,并将文本句子和模板拼接输入到预训练语言模型生成具有语义信息的隐藏向量。为了保证随机生成伪标签的有效性,采用两层由长短期记忆模型和 RELU 组成的感知机训练伪标签。然后,将通过 wordnet 知识库获取到的外部知识通过主题图注意力机制的方式融入到隐藏向量中。为了平衡外部知识和模型学习的语义信息知识,将隐藏向量

本身也参与注意力计算中。最后,通过全连接层获得候选标签词的概率,并通过求和置信度的方式将候选标签词的概率映射到对应情感极性空间上。在两个公开方面级情感分析数据集中,获得了比对比模型更高的正确率和 Macro-F1 值,证实了模型的有效性。本文通过模型模块分析证明了外部知识和 prompt 模块的作用。下一步,我们将考虑如何更加有效地利用外部知识,以期能达到更好的实验效果。

## 参考文献

- [1] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv:1810.04805, 2018.
- [2] RADFORD A, NARASIMHAN K, SALIMANS T, et al. Improving language understanding by generative pre-training [J]. OpenAI Blog, 2018.
- [3] RAFFEL C, SHAZEER N, ROBERTSA, et al. Exploring the limits of transfer learning with a unified text-to-text transformer [J]. J. Mach. Learn. Res., 2020, 21(140): 1-67.
- [4] LIU Y, OTT M, GOYAL N, et al. Roberta: A robustly optimized bert pretraining approach[J]. arXiv:1907.11692, 1907.
- [5] JAWAHAR G, SAGOT B, SEDDAH D. What does bert learn about the structure of language? [C] // Proceedings of ACL. 2019: 3651-3657.
- [6] YENICELIK D, SCHMIDT F, KILCHER Y. How does bert capture semantics? a closer look at polysemous words [C] // Proceedings of BlackboxNLP. 2020: 156-162.
- [7] HEWITT J, MANNING C D. A structural probe for finding syntax in word representations [C] // Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). 2019: 4129-4138.
- [8] PETRONI F, ROCKTÄSCHEL T, LEWIS P, et al. Language models as knowledge bases? [J]. arXiv:1909.01066, 2019.
- [9] 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.
- [10] YAN H, DAI J, QIU X, et al. A unified generative framework for aspect-based sentiment analysis [J]. arXiv: 2106.04300, 2021.
- [11] JIANG Z, XU F F, ARAKI J, et al. How can we know what language models know? [J]. Transactions of the Association for Computational Linguistics, 2020, 8: 423-438.
- [12] SHIN T, RAZEGHI Y, LOGAN IV R L, et al. Autoprompt: Eliciting knowledge from language models with automatically generated prompts [J]. arXiv:2010.15980, 2020.
- [13] SCHICK T, SCHÜTZE H. Exploiting cloze questions for few shot text classification and natural language inference [J]. arXiv:2001.07676, 2020.
- [14] HAN X, ZHAO W, DING N, et al. Ptr: Prompt tuning with rules for text classification [J]. arXiv:2105.11259, 2021.
- [15] LIU X, ZHENG Y, DU Z, et al. GPT understands, too [J]. arXiv:2103.10385, 2021.
- [16] DING X, LIU B, YU P S. A holistic lexicon-based approach to opinion mining [C] // Proceedings of the 2008 International Conference on Web Search and Data Mining. 2008: 231-240.

- [17] NGUYEN T H,SHIRAI K. Aspect-based sentiment analysis using tree kernel based relation extraction[C]// International Conference on Intelligent Text Processing and Computational Linguistics. Cham;Springer,2015:114-125.
- [18] LIPENKOVA J. A system for fine-grained aspect-based sentiment analysis of Chinese[C]// Proceedings of ACL-IJCNLP 2015 System Demonstrations. 2015:55-60.
- [19] KIRITCHENKO S,ZHU X,CHERRY C,et al. Detecting aspects and sentiment in customer reviews[C]// 8th International Workshop on Semantic Evaluation(SemEval). 2014:437-442.
- [20] RAO D,RAVICHANDRAN D. Semi-supervised polarity lexicon induction[C]//Proceedings of the 12th Conference of the European Chapter of the ACL(EACL 2009). 2009:675-682.
- [21] SUN K,ZHANG R,MENSAH S,et al. Aspect-level sentiment analysis via convolution over dependency tree[C]// Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019: 5679-5688.
- [22] WANG K,SHEN W,YANG Y,et al. Relational Graph Attention Network for Aspect-based Sentiment Analysis[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020:3229-3238.
- [23] TANG J,LU Z,SU J,et al. Progressive Self-Supervised Attention Learning for Aspect-Level Sentiment Analysis[C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019:557-566.
- [24] HE R,LEE W S,NGH T,et al. An Interactive Multi-Task Learning Network for End-to-End Aspect-Based Sentiment Analysis[C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019:504-515.
- [25] WU C,XIONG Q,YI H,et al. Multiple-element joint detection for Aspect-Based Sentiment Analysis[J]. Knowledge-Based Systems,2021,223:107073.
- [26] LI Y,YIN C,ZHONG S,et al. Multi-Instance Multi-Label Learning Networks for Aspect-Category Sentiment Analysis [C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020: 3550-3560.
- [27] 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.
- [28] BROWN T,MANN B,RYDER N,et al. Language models are few-shot learners[J]. Advances in Neural Information Processing Systems,2020,33:1877-1901.
- [29] PETRONI F,ROCKTÄSCHEL T,LEWIS P,et al. Language models as knowledge bases? [J]. arXiv:1909. 01066,2019.
- [30] DAVISON J,FELDMAN J,RUSH A M. Commonsense knowledge mining from pretrained models[C]// Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing(EMNLP-IJCNLP). 2019:1173-1178.
- [31] HAN X,ZHAO W,DING N,et al. Ptr:Prompt tuning with rules for text classification[J]. arXiv:2105. 11259,2021.
- [32] LIU X,ZHENG Y,DU Z,et al. GPT understands , too [ J ] . arXiv:2103. 10385,2021.
- [33] LI X L,LIANG P. Prefix-tuning:Optimizing continuous prompts for generation[J]. arXiv:2101. 00190,2021.
- [34] SCHICK T,SCHMID H,SCHÜTZE H. Automatically identifying words that can serve as labels for few-shot text classification[J]. arXiv:2010. 13641,2020.
- [35] BENGIO Y. From system 1 deep learning to system 2 deep learning[C]//Neural Information Processing Systems. 2019.
- [36] 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.
- [37] TANG D,QIN B,FENG X,et al. Effective LSTMs for target-dependent sentiment classification[J]. arXiv:1512. 01100,2015.
- [38] DONG L,WEI F,TAN C,et al. Adaptive recursive neural network for target-dependent twitter sentiment classification[C]// Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (volume 2: Short papers). 2014: 49-54.
- [39] WANG Y,HUANG M,ZHU X,et al. Attention-based LSTM for aspect-level sentiment classification[C]// Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. 2016:606-615.
- [40] MA D,LI S,ZHANG X,et al. Interactive attention networks for aspect-level sentiment classification [J]. arXiv: 1709. 00893, 2017.
- [41] TANG D,QIN B,LIU T. Aspect level sentiment classification with deep memory network[J]. arXiv:1605. 08900,2016.
- [42] CHEN P,SUN Z,BING L,et al. Recurrent attention network on memory for aspect sentiment analysis[C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017:452-461.
- [43] LIU Q,ZHANG H,ZENG Y,et al. Content attention model for aspect based sentiment analysis[C]// Proceedings of the 2018 World Wide Web Conference. 2018:1023-1032.
- [44] DENG L M,WEI J J,WU Y B,et al. Based on knowledge map and circular attention network Perspective Level Emotional Analysis[J]. Pattern Recognition and Artificial Intelligence,2020, 33(6):9.



**LI Yang**, born in 1994, master. His main research interest is natural language processing.



**TANG Jiqiang**, born in 1981, master. His main research interests include natural language processing and internet security.