基于层级软提示交互融合的少样本事件方面类别检测方法

艾传鲜[1,2,3]，郭军军[1,2*]，尹兆良[1,2,3]

（1.昆明理工大学信息工程与自动化学院，云南 昆明 650500；
2.昆明理工大学云南省人工智能重点实验室，云南 昆明 650500；
3.国家计算机网络应急技术处理协调中心云南分中心，云南 昆明 650000）

摘　要：事件方面类别检测（ACD）旨在识别出给定事件文本中的方面类别，相关研究可以辅助获取不同领域和事件文本中的关键信息，特别是在社交媒体舆情研究中具有实用价值。在社交媒体舆情事件发展前期，事件文本标记数据稀缺，如何基于少量标记数据实现准确的事件方面检测是一个亟待解决的问题。提出一种基于预训练模型的多层级软提示交互融合少样本事件方面类别检测方法，基于预训练构建多层级的软提示模板，分别与预训练模型进行层级语义表征和交互融合，并自适应地融合多层级的提示表征，从而提升少样本事件方面类别检测的效果。在自构建中文社交媒体少样本数据集和英文数据集上进行实验，实验结果证明所提方法明显优于其他基线方法，此外，消融实验和可视化结果验证了所提多层级提示交互融合模块的有效性。

关键词：少样本；提示学习；软提示；方面类别检测；多层级提示交互融合

源代码链接：https://github.com/aichuanxian/ACD.git

中图分类号：TP18　　**文献标志码：**A　　**DOI：**10.19678/j.issn.1000-3428.0069414

Method for Event Aspect Category Detection in Few Shot Scenarios via Hierarchical Soft Prompt Interaction Fusion

AI Chuanxian[1,2,3], GUO Junjun[1,2*], YIN Zhaoliang[1,2,3]

(1.Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, Yunnan, China;
2.Yunnan Key Laboratory of Artificial Intelligence, Kunming University of Science and Technology, Kunming 650500, Yunnan, China;
3.Yunnan Branch, National Computer Network Emergency Response Technical Team/Coordination Center of China, Kunming 650000, Yunnan, China)

【Abstract】Event Aspect Category Detection (ACD) aims to identify the aspect categories present in event text. Data need to be collected from various fields and textual events, particularly when researching public opinion on social media. The first phase of social media opinion events lacks sufficient data for labeling event text. The pressing issue is precisely detecting event aspects using a limited amount of labeled data. This paper presents a novel method for event ACD with limited samples. This method utilizes a pre-trained model to construct soft prompt templates, performs hierarchical semantic characterization and interaction fusion, and adaptively combines multilayer prompt characterizations. The objective is to enhance the accuracy of event ACD using limited samples. Experiments on a self-constructed Chinese social media dataset and English dataset demonstrate that the proposed method is significantly superior to other baseline methods. Further ablation experiments and visualizations confirm the effectiveness of the proposed multilayer prompt interaction fusion module.

【Key words】few-shot; prompt learning; soft prompt; Aspect Category Detection (ACD); multilayer prompt interaction fusion

0　引言

互联网技术快速发展和普及使得社交媒体成为网民接收各类互联网信息和表达自我观点及情绪的重要途径，社交媒体所体现出来的社会舆论影响力正受到日益关注。一般来说，社会舆情有 5 个发展阶段：萌发期、发展期、爆发期、衰退期和沉淀期。前 2 个阶段时间跨度较大，但相关数据稀缺，而爆发期会出现海量的与舆情事件相关的数据。充分挖掘和利用萌芽期和发展期的少量标记数据对舆情爆发期的海量文本未标记数据开展快速舆情分析，对于舆情监测和预警研判具有深远意义。

方面类别检测（ACD）是早期舆情检测的重要任务之一，其目标是从预定义的一组类别词中识别

收稿日期：2024-02-23　　**修回日期：**2024-04-12

基金项目：国家重点研发计划（202301AT070444）；国家自然科学基金（62366025）；云南省科技厅自然科学基金（202202AE090008-3）。

通信作者 E-mail：guojjgb@163.com

出给定文本所提到的方面类别[1-3]，能被用以获取各个领域的文本中的关键信息，如关键方面及对应的情感极性[4-5]。然而在仅有少量事件评论文本的情况下，如何获取事件语义信息，实现方面类别检测任务是一个巨大的挑战。

在以往的研究中，针对特定事件的舆情分析往往关注的是舆情事件整体的趋势和影响，但对于特定事件中的某些方面却少有关注。事实上，针对特定事件中的某一个或者几个方面开展检测和情感分析，能为舆情研判提供重要参考。然而，在少样本的条件下，针对特定事件中某一个或几个方面类别进行检测会变得更加困难。首先，少样本条件下的数据稀疏性大大增加了对类别不平衡问题的敏感度，可能导致过拟合，从而影响模型的泛化性能。其次，当数据量较小时，足够代表性的样本获取更为困难，因此可能无法全面地捕捉到对特定事件各个方面的观感变化，造成分析结果的偏差。最后，在有限的数据中抽取有效的特征，这对于深度学习模型的设计提出了更高的要求。

目前大多数解决方面类别检测任务的方法都严重依赖于大量有标注的数据，严重限制了这些方法的实用性，主要原因在于获取这些标注数据不仅费时费力，成本高昂，更在于这些数据仅仅局限于少数特定领域，在泛化至其他领域时，表现并不好。为解决这个问题，一些研究者探索将方面类别检测任务转化为少样本方面类别检测任务。例如：HU 等[2]提出一个名为 Proto-AWATT 的原型网络，该网络基于注意力机制提取语义信息实现少样本方面类别检测；ZHAO 等[6]和 WANG 等[7]在 HU 等研究基础上分别提出了一种标签驱动的降噪网络和一种句子级加权的方法用于解决 Proto-AWATT 中存在的噪声问题，进一步提升了模型在少样本方面类别检测上的性能。然而，这些研究主要聚焦于通用任务，对于特定事件的方面类别检测缺乏深入探索。不同于通用任务，特定事件的方面通常具有不同层级的事件知识。如何从特定事件少量样本数据中获取这些层级的有效信息并用于方面类别检测任务是本文的研究重点。

借助大规模语料预训练模型，可以在表示特定事件文本时理解文本背后的事件知识，如果能够借助一定手段在事件文本表征的时候与预训练层级语义进行交互，就能够学习到事件不同层级的表征[8]。提示学习[9-11]是一种通过构建提示文本，缩小下游任务和预训练模型之间差距的方法。提示学习可以充分利用语言模型在学习大规模语料后得到的知识、模式以及文本生成能力，让预训练模型可以更好地适应下游任务，因此可以在训练数据很少甚至没有的情况下，提取特定事件不同层级的事件知识，实现更好的泛化结果[9,12-14]。针对少样本情况下缺少语义信息有效提取方法的情况，构建层级提示可能对事件表征有帮助。基于此，本文提出一种融合提示学习的中文社交媒体事件少样本方面类别检测方法，通过层级的提示表征捕获不同层预训练语言模型的表征，实现对不同事件层级知识表征，最终完成特定事件方面类别检测任务。本文主要贡献包括 3 个方面：

1）针对中文社交媒体早期舆情事件方面类别检测问题，提出基于预训练模型的多层级提示交互融合的中文社交媒体事件少样本方面类别检测方法，通过挖掘融合预训练模型中多层级表征，提升少样本事件方面类别检测能力，解决舆情发展早期数据量少和发展期数据大量爆发的问题，实现特定事件舆情事件监测。

2）提出多层级提示知识融合的少样本事件方面类别表征和分类方法。通过深度提示调整的方法实现多层级提示表征，并利用自注意力交叉机制和 CLS 标记捕获事件多层级语义信息，最后基于多层级交互融合，实现少样本方面类别检测。

3）完成在自建中文数据集和英文公共数据集上的实验，结果表明本文方法具有较好的泛化性，验证了方法的有效性。

1 相关工作

1.1 方面类别检测

自 2014 年 SemEval 引入方面类别检测任务后，研究者提出不同的方法来解决方面类别检测问题。目前，根据是否有标注标签，方面类别识别采用有监督的方法和弱监督的方法。随着 GPU 等更好的计算机硬件以及 Word2Vec[15]和 GloVe[16]等词嵌入方法的出现，深度学习的方法被引入自然语言处理任务，方面类别检测任务也开始利用任务的不同特征来提高性能，例如：使用卷积神经网络（CNN）提取局部特征[16]；使用循环神经网络（RNN）获取序列之间的关系[17-19]；使用注意力机制关注不同类别文本的不同部分[20]。最近几年，诸如 ELMo（Embeddings from Language Models）[21]、GPT（Generative Pre-trained Transformer）[22]、BERT（Bidirectional Encoder Representations from Transformers）[8]之类的基于 Transformer[23]的预训练模型不断涌现，这类预训练模型具备从大

规模无标注文本数据中学习通用语言表征并将所学知识迁移到下游任务的能力，使得预训练＋微调的范式在很多自然语言处理任务中都表现突出[20,23-24]。弱监督方法主要包括采用共现矩阵[25]、余弦相似度[26]等方法。然而，无论是有监督的方法还是无监督的方法，都需要大量的语料资源或者有标记的数据。

1.2 提示学习

BROWN 等[27]提出了 GPT-3 模型。基于该模型的理念，研究者提出了预训练-提示-预测的训练范式，这种范式在少样本学习方面表现突出[10,27]。传统的预训练-微调范式主要通过训练使模型适应下游任务，而预训练-提示-预测的范式是通过构建提示文本，激发下游任务更好地提取预训练模型中适用于任务的特征[10-11]。提示学习主要分为构造提示模板、答案输出和答案匹配 3 个主要步骤，其中构造提示模板对模型表现有较大影响。构造提示模板又称为提示工程，主要是通过人工或者自动的方式构造一个符合预定义形状的文本以便于与预训练语言模型进行交互。根据提示模板的形式可以将提示模板分为硬提示和软提示。硬提示是人工构造的离散词组成的文本提示，更便于人类识别、理解和调整，但因硬提示为人工构造，其模板内容对模型表现影响非常大。为解决人工构造模板带来的不确定性，研究者提出让语言模型自动学习提示模板的方法，如 LIU 等[28]提出 P-tuning 模型，该模型能自动学习适用于当前任务的最佳提示模板。有别于人工构造的离散模板，该方法得到的提示模板为连续的非自然语言模板，因此不具备可解释性。提示学习最主要的目的是通过构造合适的提示文本激发预训练模型中适用于任务的知识，减小下游任务和预训练模型之间的巨大差异，以达到提升模型表现的目的。

1.3 少样本学习

少样本学习[29]来源于对人类强大的推理能力和分析能力的模仿，是一种利用少量标注文本数据结合先验知识来实现快速获取新的知识的训练方法。少样本学习方法主要有基于元学习的方法和基于非元学习的方法，目前绝大部分少样本学习方法都为基于元学习的方法，主要包括基于度量的方法、基于优化的方法和基于模型的方法[2,6,30]。最近几年，基于这些方法，研究者在数据、算法和模型方面开展了一些创造性工作，并将它们应用于自然语言处理领域。尽管如此，少样本学习还是面对许多挑战，包括数据分布评估不准确、特征重用敏感性、泛化性问题和信息局限性[29,31]。如何充分利用少量样本中的特征信息，是提高少样本学习表现的重要研究方向。

2 本文方法

如图 1 所示（彩色效果见《计算机工程》官网 HTML 版，下同），本文通过构建连续的软提示文本，并将提示文本和预训练模型进行多层级表征和交互融合，以充分挖掘预训练模型的特征信息，从而提高少样本情况下模型的表现。

2.1 问题描述

针对少样本方面类别检测任务的数据集，有 2 个主要组成部分，分别是支持集和查询集。训练的目的是将查询集中的查询实例与支持集相匹配。少样本方面类别检测任务的训练数据一般遵循元学习 N-way K-shot 的设置方法，即每个任务有 N 个类别，每个类别有 K 条标注数据。例如，在一个 8-way 5-shot 任务中，有 8 个类别，每个类别有 5 条标注数据。在本文中，将数据集分为训练数据集（即支持集 $S_{\text{train}}=\{(x_i,y_i)\}_{i=1}^{k}$）和测试数据集（即查询集 $Q_{\text{test}}=\{(x_j,y_j)\}_{j=1}^{n}$），其中查询集包括了验证集和测试集。如式（1）所示，该任务通过训练集 S_{train} 训练一组参数 θ，使得该模型泛化至测试 Q_{test} 时取得较好的结果。

$$y=f(x;\theta)\ \text{where}\ (x,y)\in Q_{\text{test}} \tag{1}$$

式中：x 为输入文本；y 为类别标签，$y=\{y_1,y_2,\cdots,y_N\}$，其中，$y_N\in(0,1)$，N 为可能的类别数量；θ 为待训练优化的参数，$\theta=\underset{\theta}{\operatorname{argmin}}\sum_{(x,y)\in S_{\text{train}}}\mathcal{L}(f(x,\theta),y)$，其中，$\mathcal{L}(f(x,\theta),y)$ 为模型损失函数。

2.2 基于多层级提示知识融合的表征方法

2.2.1 层级软提示表征

在本文中，将少样本方面类别识别视为少样本分类任务。通过提示文本提取预训练模型中的先验知识，完成事件的特征提取。对于输入文本 x，通过提示函数 $f_{\text{prompt}}(\cdot)$ 将输入文本 x 转换为融合提示文本的文本 x'，然后将 x' 作为新的输入用以训练模型。如式（2）所示，融合提示知识方面类别检测方法通过将新的 x' 作为模型输入，训练优化模型得到最优的模型参数 θ'。在本文中，$f_{\text{prompt}}(x)=[\text{prompt}][x]$，即 $x'=[\text{prompt}][x]$，其中，prompt 为软提示文本，可以通过模型训练优化，以达到更好的表征效果。

$$y=f(x';\theta')\ \text{where}\ x'=f_{\text{prompt}}(x) \tag{2}$$

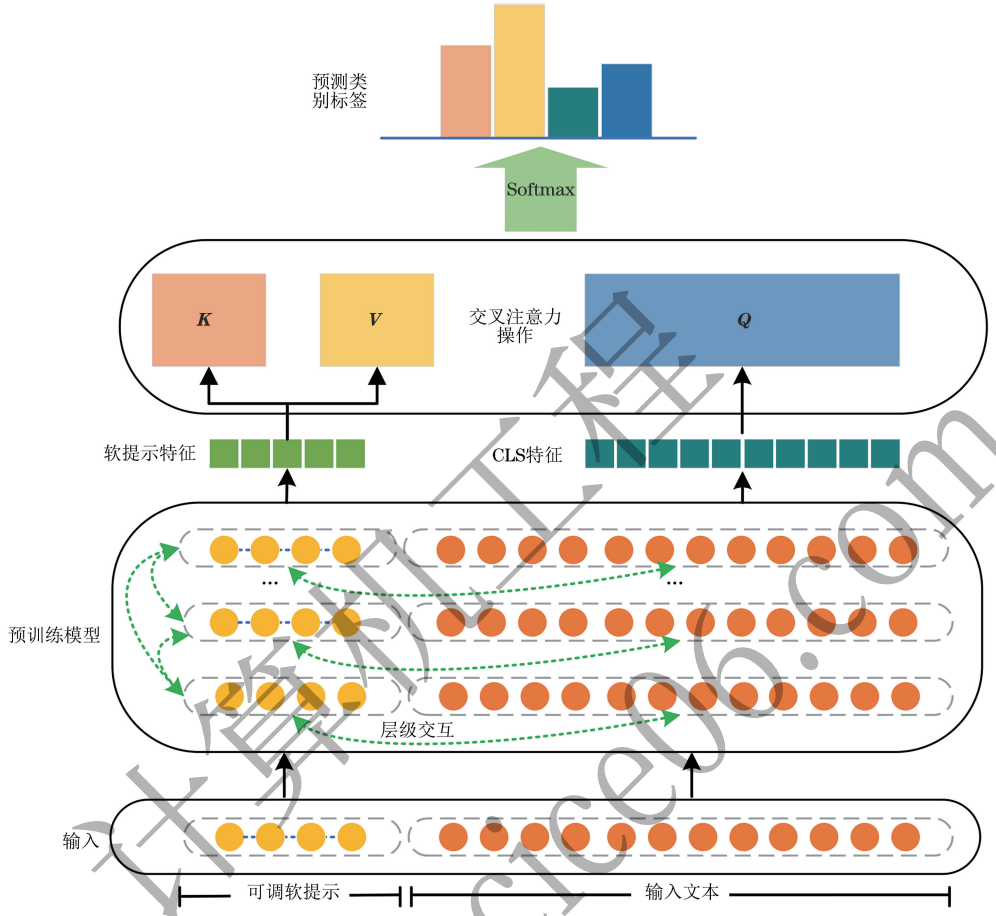

图 1 本文方法结构

Fig. 1 Structure of the proposed method

在深度神经网络中，深层的神经元对模型的输出影响更为直接，且拥有更多可以调节的参数。因此，一般情况下，模型训练的效果会更好。为了充分获取预训练模型与方面类别检测相关的先验知识，同时减少人工构造模板带来的不确定性，本文采用了深度软提示的方法实现多层级提示表征。如图 1 所示，首先将提示词初始化为固定长度的可以优化的向量，然后将这些向量拼接到每个嵌入输入的开头，并将组合序列输入模型。模型训练的时候，只更新可调的软提示向量，预训练模型的参数固定不训练。为了能有更多的可微调的参数，针对预训练模型的每一层都会插入软提示来微调，而不仅仅是输入层。这种方式使得提示文本能在各个层级中发挥作用，提供丰富而直观的语义信息。

2.2.2 层内语义交互

在本文的模型中，层内语义交互主要通过 BERT 模型实现。BERT 模型内部有多个 Transformer 层，本文通过利用 Transformer 层的自注意力机制实现层内语义交互。这种交互方式使得模型能够关注到输入序列中的重要部分，从而提高模型的性能。此外，本文还引入了提示知识，通过在输入序列中添加提示标记，使模型能够更好地理解和学习类别信息。将输入文本 $X=[x_1,x_2,\cdots,x_T]$ 和提示文本 $P=[p_1,p_2,\cdots,p_n]$ 按照 $f_{\text{prompt}}(x)=[\text{prompt}][x]$ 进行拼接得到最终的输入序列 $\boldsymbol{H}_0=[h_{0,1},h_{0,2},\cdots,h_{0,T+n}]$。

如式(3)所示，输入序列经过预训练模型实现语义交互，在每一层 $l=1,2,\cdots,L$ 中，使用自注意力机制对 $\boldsymbol{H}_{l-1}$ 进行加权处理得到自注意力输出 $\boldsymbol{O}_l$：

$$\begin{aligned}\boldsymbol{Q}_l&=\boldsymbol{W}_{\boldsymbol{Q}}^l\cdot\boldsymbol{H}_{l-1}\\ \boldsymbol{K}_l&=\boldsymbol{W}_{\boldsymbol{K}}^l\cdot\boldsymbol{H}_{l-1}\\ \boldsymbol{V}_l&=\boldsymbol{W}_{\boldsymbol{V}}^l\cdot\boldsymbol{H}_{l-1}\\ \boldsymbol{A}_l&=\text{Softmax}\left(\frac{\boldsymbol{Q}_l\boldsymbol{K}_l^{\mathrm{T}}}{\sqrt{d_k}}\right)\\ \boldsymbol{O}_l&=\boldsymbol{A}_l\cdot\boldsymbol{V}_l\end{aligned}\tag{3}$$

式中：$\boldsymbol{W}_{\boldsymbol{Q}}^l$、$\boldsymbol{W}_{\boldsymbol{K}}^l$、$\boldsymbol{W}_{\boldsymbol{V}}^l$ 为第 l 层的可学习权重矩阵；d_k 为键向量的维度。

2.2.3 多层级交互融合

为了增强提示知识和预训练模型的关联性，本文设计了一个特殊的多层级交互融合策略。具体来说，首先从每一层获取相应的提示特征和 CLS 标记的特征，然后对两者进行交叉注意力操作，以捕获它

们之间的交互信息，此过程可以表示为：

$$\boldsymbol{C}=\text{CrossAttention}(\boldsymbol{H}_{\text{CLS}},\boldsymbol{P}) \tag{4}$$

式中：$\boldsymbol{H}_{\text{CLS}}$ 是 CLS 标记的隐藏状态；$\boldsymbol{P}$ 是所有层级上的提示特征。

接着，采用残差连接将交叉注意力特征与 CLS 标记的隐藏状态进行融合，形成最终的特征表示：

$$\boldsymbol{F}=\boldsymbol{C}+\boldsymbol{H}_{\text{CLS}} \tag{5}$$

这种多层级交互融合策略能够使模型更好地理解和捕获提示文本与原始输入文本之间的语义关系，从而进一步提升模型的性能。

2.3 融合多层级提示知识的分类

为完成最终的类别分类，对融合后的最终特征通过全连接层和 Softmax 函数，将最终特征表示转化为类别预测概率：

$$\hat{y}=\text{Softmax}(\text{Classifier}(\boldsymbol{F})) \tag{6}$$

在训练阶段，本文采用交叉熵损失函数来度量模型预测结果与真实标签之间的误差，计算公式如下：

$$\mathcal{L}=-\frac{1}{N}\sum_{i=1}^{N}y_i\ln\hat{y}_i \tag{7}$$

式中：y_i 是真实标签；$\hat{y}_i$ 是预测的概率；N 是样本总数。

综上所述，本文提出的基于 Transformer 的多层级提示信息融合模型有效地利用了多层级的提示知识，从而在少样本情境下显著提升了文本分类任务的性能。在后续的实验部分，将展示该模型在各项任务中的性能优势，证明其实际应用价值。

3 实验结果分析

3.1 数据集

微博平台作为现阶段最大最活跃的中文社交媒体平台，拥有丰富的评论文本。本文通过爬虫技术爬取微博平台特定事件评论文本，经过清洗标注，构建一个中文社交媒体事件少样本方面类别检测数据集。该数据集共有 6 937 条文本，涉及 8 个方面类别，其分布如图 2 所示。8-way 5-shot和8-way 10-shot分别从

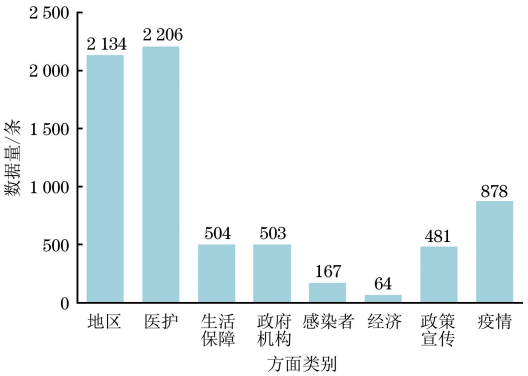

图 2 自建数据集方面类别分布

Fig. 2 Distribution of aspect categories in self-constructed dataset

每个类别对应文本数据中随机选取 5/10 条数据构成训练数据集，剩余数据按照 50%的比例分为验证数据集和测试数据集，其分布如表 1 所示。

表 1 自建中文社交媒体少样本数据集

Table 1 Self-constructed Chinese social media few-shot dataset 单位：条

设置方法	总数据量	训练集	验证集	测试集
10-shot	6 937	80	3 429	3 428
5-shot	6 937	40	3 449	3 448

为验证本文所提模型的泛化性，选取 SemEval 2014 restaurant 数据集构建少样本数据集。同自建中文数据集一样，8-way 5-shot 和 8-way 10-shot 的英文数据集分别从每个类别对应文本数据中随机选取 5/10 条数据构成训练数据集，剩余数据按照 50%的比例分为验证数据集和测试数据集，其分布如表 2 所示。

表 2 SemEval 2014 restaurant 少样本数据集

Table 2 SemEval 2014 restaurant few-shot dataset 单位：条

设置方式	总数据量	训练集	验证集	测试集
10-shot	3 832	50	800	832
5-shot	3 832	25	800	832

3.2 实验参数设置

1）评估指标。

在以往的单标签少样本研究中，一般采用准确率作为评价指标。为更好地评价各个类别的准确率和召回率，本文还采用宏观 F1 值（Macro F1，$F_{\text{Macro\_F1}}$）作为评估指标：

$$F_{\text{Macro\_F1}}=\frac{1}{N}\sum\frac{2\times P_i\times R_i}{P_i+R_i} \tag{8}$$

式中：$P_i=N_{\text{TP}\_i}/(N_{\text{TP}\_i}+N_{\text{FP}\_i})$；$R_i=N_{\text{TP}\_i}/(N_{\text{TP}\_i}+N_{\text{FP}\_i})$；$N_{\text{TP}\_i}$ 表示预测为正样本且分类正确的样本数；N_{FP_i} 表示实际为负且分类错误的样本数。

2）训练细节。

实验配置一块 32 GB NVIDIA V100 显卡，运行平台为 Linux，代码基于 PyTorch 框架实现。对于中文数据集的评测，采用 12 层网络结构的 bert-base-chinese 预训练模型。对于英文数据集的评测，采用 12 层网络结构的 bert-base-uncased 预训练模型，并使用 AdamW 优化器进行模型参数的优化。

3.3 主实验

通过与以下一些常用的方法进行对比，以验证本文方法的有效性。

1）基于 BERT 的预训练-微调方法[32]：预训练-微调的方法常被用于方面级情感分析任务，该方法通过微调预训练模型，以适应下游任务，从而完成类别检测。

2）Proto-AWATT[2]：使用一种基于原型网络

的多标签少样本方面类别检测方法，并使用了支持集和查询集注意力机制以减少噪声的影响。

3）Proto-SLWLA[7]：在原型网络的基础上，该方法设计了句子级关注机制，给予支持集中每个实例不同的权重，通过加权平均来计算原型，并执行方面类别检测任务。

4）LDF[6]：在 Proto-AWATT 方法的基础上，该方法构建了一个标签驱动的降噪框架，旨在减少原型网络噪声影响，从而提升模型性能。

表 3 总结了不同方法的实验结果，加粗表示最优的结果，下同。可以看出，本文方法在所有的少样本数据集上均取得了超过基线模型的效果，这证明本文采用的技术路径是可行的，并且可以明显看出，本文方法无论是在中文数据集还是在英文数据集上都取得了较基线方法更好的效果，这表明本文方法具备较好的泛化性。值得注意的是，10-shot 的表现均要好于 5-shot，这也更符合直觉，也就是样本量越多，训练效果会越好。此外，SemEval 数据集的效果要明显好于自建中文数据集，这与数据集质量有关：在 SemEval 数据集中，各类文本讨论的主题均比较集中，并且很有可能在预训练模型训练过程中被用于训练资料；而自建中文数据集来源于微博评论，虽然讨论的是具体的事件，但由于其差异性大且主观性强，导致模型在自建中文数据集上的表现不如英文数据集。

表 3 本文方法与其他方法对比结果

Table 3 Comparison results of the proposed method with other methods %

方法	自建中文数据集				SemEval 数据集			
	8-way 5-shot		8-way 10-shot		5-way 5-shot		5-way 10-shot	
	准确率	Macro F1	准确率	Macro F1	准确率	Macro F1	准确率	Macro F1
预训练-微调	35.07	25.12	48.83	34.08	58.88	49.33	59.75	53.95
Proto-SLWLA	39.43	28.96	54.62	43.29	59.26	51.61	60.47	57.16
Proto-AWATT	39.02	27.45	55.23	44.37	60.98	**51.90**	63.78	57.89
LDF	39.58	**30.83**	55.26	46.89	61.03	50.43	64.40	58.46
本文方法	**41.51**	30.70	**57.38**	**47.99**	**61.83**	51.83	**66.25**	**59.77**

3.4 消融实验

3.4.1 层级提示消融

表 4 总结了层级提示消融实验的结果。从表中可以看出，使用层级提示时，模型准确率最多提升 7.55 百分点，这证明了层级提示在本文方法中起到关键性的作用。这个模块主要帮助模型更好地理解和捕获提示文本与原始输入文本之间的语义关系。通过这种方式，模型可以在处理复杂任务时，利用已有的知识进行推理，从而大大提升模型的性能。此外，该模块还有效地缓解了数据稀疏性问题，特别是对于那些只有少量标注样本可用的任务，可以通过引导模型更加关注与任务相关的信息，从而提高模型的泛化能力。这表明多层级交互融合模块在本文的模型设计中起到了核心作用，并且对于提高模型的性能有着至关重要的影响。

表 4 层级提示消融实验结果

Table 4 Results of hierarchical prompt ablation experiment %

方法	8-way 5-shot		8-way 10-shot	
	准确率	Macro F1	准确率	Macro F1
本文方法	**41.51**	**30.70**	**56.38**	**46.69**
无层级提示	35.07	25.12	48.83	34.08

3.4.2 多层级交互融合模块消融

表 5 总结了层级提示消融实验的结果。从表中可以看出，在没有多层级交互融合模块的情况下，模型的准确率最多下降了 5.44 百分点，而 Macro F1 值最多下降了 12.16 百分点。这一结果显著表明了多层级交互融合模块在本文方法中的重要性，说明其能够帮助模型更好地理解和捕获提示文本与原始输入文本之间的语义关系，从而显著提升模型的性能。

表 5 多层级交互融合模块消融实验结果

Table 5 Results of multilayer interactive fusion module ablation experiment %

方法	8-way 5-shot		8-way 10-shot	
	准确率	Macro F1	准确率	Macro F1
本文方法	**41.51**	**30.70**	**56.38**	**46.69**
无多层级交互融合	39.96	29.77	50.94	34.44

3.5 定量实验

3.5.1 不同预训练模型比较

本文使用不同的预训练模型在自建数据集上进行对比。考虑到实际应用时硬件、资金投入等因素，采用目前参数量较小，适合在资源有限的环境下使用的 BERT 模型作为基线。此外，为了能在中文方面类别情感分析中取得更好的效果，采用基于中文语料库训练的 bert-base-chinese 这个经典且被广泛使用的模型作为基线模型。作为对比，在使用了全词 Mask 策略且基于更大规模中文语料库的 chinese-roberta-wwm-ext 模型上进行对比实验。从表 6 中可以看出，参数量更大、模型更复杂的 chinese-roberta-wwm-ext 预训练模型相较于 bert-base-chinese 预训练模型效果更好，这表明提升模

型参数量和复杂度能使得模型更好地学习和理解完整词语的语义，从而提升模型表现。

表 6 自建中文数据集上不同预训练模型对比结果

Table 6 Comparison results of different pre-trained models on self-constructed Chinese dataset %

模型	8-way 5-shot		8-way 10-shot	
	准确率	Macro F1	准确率	Macro F1
chinese-roberta-wwm-ext	41.07	**31.32**	**57.27**	**49.45**
bert-base-chinese	**41.51**	30.70	56.38	46.69

3.5.2 提示长度的影响

为了研究提示长度对模型性能的影响，本文在固定其他参数的情况下，改变提示长度，并记录模型的表现。图 3 表明，当提示文本较短时，模型的准确率和 Macro F1 值相较于长提示文本更好。当提示长度超过一定阈值后，模型的准确率和 Macro F1 值均下降，这与过长的提示会引入噪声使得模型难以从中学习到有效的信息有关。此外，过长的提示词也会引发数据复杂度上升，超过了所选择的预训练模型的复杂度，导致模型表现受到影响。因此，选择合适的提示长度对于模型的性能至关重要。

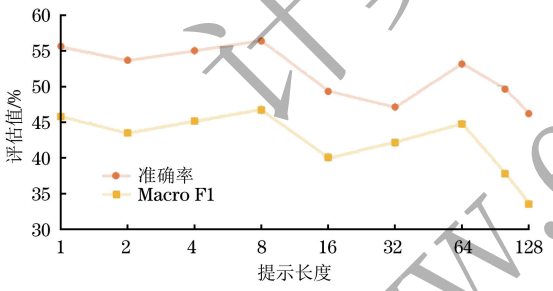

图 3 不同提示长度对模型性能的影响

Fig. 3 Impact of different prompt lengths on model performance

3.5.3 不同样本数量的影响实验

本文还探索了不同样本数量对模型性能的影响。在本文的实验中，分别尝试了 1-shot～25-shot 场景，并比较了模型的性能。图 4 表明，模型的表现和样本量整体呈现正相关的趋势，但是随着样本量的增多，模型效果提升的趋势越来越小。这个结果符合预期，因为在更多样本场景下，模型可以访问更多的样本数据进行学习，从而提升模型的整体表现。但是，过多的样本反而会引入噪声，导致模型性能在引入一定样本量后并没有得到显著提升。因此，选择适当的样本数量也是一个重要的问题，需要进一步研究。

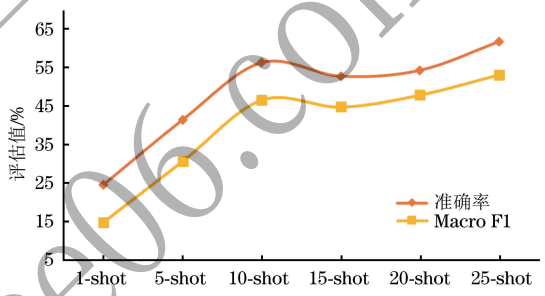

图 4 不同样本数量对模型性能的影响

Fig. 4 Impact of different numbers of shots on model performance

3.6 可视化分析

本文进行了多层级交互融合权重的可视化分析，使用注意力热图来描绘不同层级在模型中的注意力分布。通过比较热图，可以直观地看出每个层级对于最后模型预测结果的贡献，以及它们之间的差异。如图 5 所示，一些层级的注意力得分较高，说

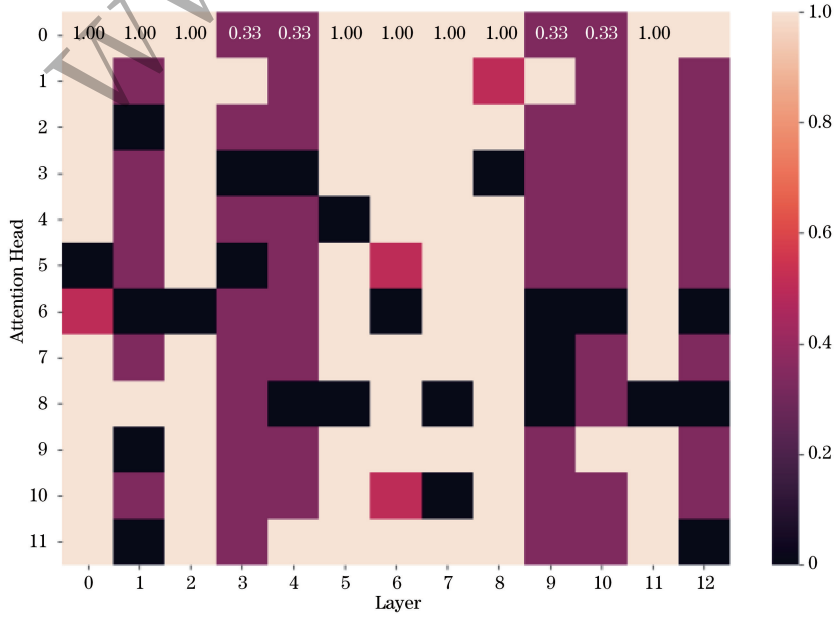

图 5 可视化分析

Fig. 5 Visualization and analysis

明模型在做预测时更依赖这些层级的信息。

3.7 实例分析

为了更好地理解本文模型，详细分析了一些具体的案例。在表 7 中，展示了一些输入样本、模型的预测结果以及它们的真实标签。从这些样本中可以看到，虽然在一些复杂的情况下，模型会做出错误的预测，但大多数情况下，模型能够准确地识别出正确的方面类别。

本文还从错误预测的样本中找出了一些常见的失败模式，例如，模型有时会困惑于文本中的歧义词，或者在面对少见的方面类别时无法做出正确的预测。这为未来改进模型提供了一些线索。总体而言，通过实例分析，可以深入理解模型的优点和缺点，从而为未来的工作提供启示。

表 7 实例分析

Table 7 Case study

测试文本	真实值	预测值	是否正确
希望我们可以攻克难关，心系武汉，众志成城!!!	地区	地区	是
#小汤山抗非典医生请战新型肺炎# 不管是出于什么目的，这样一支专业的团队能置个人安危于一边投入疫区奋战，即使戴上全套护具，也不能百分百的防护，还是有被感染的风险！给这些人点个赞，这些人才是最美天使。	医护	医护	是
我在武汉，粑粑开出租车。我给他车上备了很多口罩免费提供给乘客。我心很小，希望身边人都健康。希望武汉早点渡过这段时期！	疫情	地区	否
一定会平安度过的。#武汉加油# 我们一起，打赢这场防御战！向战斗在一线的医护人员致敬！	医护	疫情	否

4 结束语

针对实际应用中方面类别检测方法所面临的资源匮乏问题，本文提出了一种基于多层级软提示交互融合的少样本事件方面类别检测策略。通过构建软提示文本，在事件文本表征时与预训练模型的层级语义进行深度多层交互，从而充分挖掘有限样本下文本的特征信息，达成了在少样本环境中完成方面类别检测任务的目标。经过严谨的实验表明，相较之前的研究方法，本文方法在处理中文社交媒体事件的少样本方面类别检测任务时表现更佳。值得一提的是，当应用到英文的少样本数据集上时，该方法展示出了较好的泛化能力，这为其在跨语言场景中的适用性和可靠性提供了有力证据。本文工作能够对解决少样本问题，尤其是在中文特定事件方面类别检测领域，提供新的视角和思路。模型的表现与模型复杂度、提示长度和样本数量息息相关，如何找到最佳的提示长度和更经济的样本量是值得探索的方向。此外，针对已识别出的方面类别开展情感分析，为社交媒体舆情研究提供更丰富的支撑，也是下一步要探索的内容。

参考文献

[1] PONTIKI M, GALANIS D, PAPAGEORGIOU H, et al. SemEval-2016 task 5: aspect based sentiment analysis[C]// Proceedings of International Workshop on Semantic Evaluation. San Diego, USA: [s. n.], 2016: 1-10.

[2] HU M T, ZHAO S W, GUO H L, et al. Multi-label few-shot learning for aspect category detection[EB/OL]. [2023-12-12]. https://arxiv.org/abs/2105.14174v1.

[3] KIRANGE D K, DESHMUKH R R, KIRANGE M D K. Aspect based sentiment analysis SemEval-2014 task 4[J]. Asian Journal of Computer Science & Information Technology, 2014, 4(8): 1-10.

[4] JANG H, REMPEL E, ROE I, et al. Tracking public attitudes toward COVID-19 vaccination on tweets in Canada: using aspect-based sentiment analysis[J]. Journal of Medical Internet Research, 2022, 24(3): e35016.

[5] SAMUEL J, ALI G G M N, RAHMAN M M, et al. COVID-19 public sentiment insights and machine learning for tweets classification[J]. Information, 2020, 11(6): 314.

[6] ZHAO F, SHEN Y C, WU Z, et al. Label-driven denoising framework for multi-label few-shot aspect category detection [EB/OL]. [2023-12-12]. https://arxiv.org/abs/2210.04220v1.

[7] WANG Z Y, IWAIHARA M. Few-shot multi-label aspect category detection utilizing prototypical network with sentence-level weighting and label augmentation [M]// STRAUSS C, AMAGASA T, KOTSIS G, et al. Database and expert systems applications. Berlin, Germany: Springer, 2023: 363-377.

[8] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[EB/OL]. [2023-12-12]. http://arxiv.org/abs/1810.04805.

[9] LESTER B, AL-RFOU R, CONSTANT N, et al. The power of scale for parameter-efficient prompt tuning[EB/OL]. [2023-12-12]. https://arxiv.org/abs/2104.08691v2.

[10] LIU P F, YUAN W Z, FU J L, et al. Pre-train, prompt, and predict: a systematic survey of prompting methods in natural language processing[J]. ACM Computing Surveys, 2023, 55(9): 1-35.

[11] 顾勋勋，刘建平，邢嘉璐，等. 文本分类中 Prompt Learning 方法研究综述[J]. 计算机工程与应用，2024，60(11)：50-61.
GU X X, LIU J P, XING J L, et al. A summary of the research on Prompt Learning methods in text classification [J]. Computer Engineering and Applications, 2024, 60(11): 50-61. (in Chinese)

[12] LIU X, JI K X, FU Y C, et al. P-tuning v2: prompt tuning can be comparable to fine-tuning universally across scales and tasks[EB/OL]. [2023-12-12]. https://arxiv.org/abs/2110.07602v3.
[13] ZHANG N Y, LI L Q, CHEN X, et al. Differentiable prompt makes pre-trained language models better few-shot learners[EB/OL]. [2023-12-12]. https://arxiv.org/abs/2108.13161v7.
[14] SEOH R, BIRLE I, TAK M, et al. Open aspect target sentiment classification with natural language prompts[EB/OL]. [2023-12-12]. https://arxiv.org/abs/2109.03685v1.
[15] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[EB/OL]. [2023-12-12]. https://arxiv.org/abs/1301.3781v3.
[16] TOH Z, SU J. NLANGP: supervised machine learning system for aspect category classification and opinion target extraction [C] // Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015). Stroudsburg, USA: Association for Computational Linguistics, 2015: 1-10.
[17] SAEIDI M, BOUCHARD G, LIAKATA M, et al. SentiHood: targeted aspect based sentiment analysis dataset for urban neighbourhoods [C] // Proceedings of the 26th International Conference on Computational Linguistics: Technical Papers. Osaka, Japan: The COLING 2016 Organizing Committee, 2016: 1546-1556.
[18] TANG D Y, QIN B, LIU T. Document modeling with gated recurrent neural network for sentiment classification [C] // Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, USA: Association for Computational Linguistics, 2015: 1-10.
[19] WANG W Y, PAN S J, DAHLMEIER D, et al. Recursive neural conditional random fields for aspect-based sentiment analysis[EB/OL]. [2023-12-12]. https://arxiv.org/abs/1603.06679v3.
[20] LIANG B, DU J C, XU R F, et al. Context-aware embedding for targeted aspect-based sentiment analysis [EB/OL]. [2023-12-12]. https://arxiv.org/abs/1906.06945v1.
[21] PETERS M E, NEUMANN M, IYYER M, et al. Deep contextualized word representations [C] //Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). New Orleans, USA: Association for Computational Linguistics, 2018: 2227-2237.
[22] RADFORD A, NARASIMHAN K, SALIMANS T, et al. Improving language understanding by generative pre-training [EB/OL]. [2023-12-12]. https://www.mikecaptain.com/resources/pdf/GPT-1.pdf.
[23] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C] // Proceedings of the 31st International Conference on Neural Information Processing Systems. New York, USA: ACM Press, 2017: 6000-6010.
[24] KALYAN K S, RAJASEKHARAN A, SANGEETHA S. AMMUS: a survey of transformer-based pretrained models in natural language processing [EB/OL]. [2023-12-12]. https://arxiv.org/abs/2108.05542v2.
[25] SCHOUTEN K, FRASINCAR F, JONG F. COMMIT-p1wp3: a co-occurrence based approach to aspect-level sentiment analysis [C] //Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). Dublin, Ireland: Association for Computational Linguistics, 2014: 203-207.
[26] GHADERY E, MOVAHEDI S, FAILI H, et al. An unsupervised approach for aspect category detection using soft cosine similarity measure[EB/OL]. [2023-12-12]. https://arxiv.org/abs/1812.03361v2.
[27] BROWN T B, MANN B, RYDER N, et al. Language models are few-shot learners [EB/OL]. [2023-12-12]. http://arxiv.org/abs/2005.14165.
[28] LIU X, ZHENG Y N, DU Z X, et al. GPT understands, too [J]. AI Open, 2024, 5: 208-215.
[29] SONG Y S, WANG T, CAI P Y, et al. A comprehensive survey of few-shot learning: evolution, applications, challenges, and opportunities[J]. ACM Computing Surveys, 2023, 55(13s): 1-40.
[30] 李子成, 常晓琴, 李雅梦, 等. 基于联合学习的少样本多类别情感分类方法[J]. 北京大学学报(自然科学版), 2023, 59(1): 57-64.
LI Z C, CHANG X Q, LI Y M, et al. A joint learning approach to few-shot learning for multi-category sentiment classification[J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2023, 59(1): 57-64. (in Chinese)
[31] 李睿凡, 魏志宇, 范元涛, 等. 增强提示学习的少样本文本分类方法[J]. 北京大学学报(自然科学版), 2024, 60(1): 1-12.
LI R F, WEI Z Y, FAN Y T, et al. Enhanced prompt learning for few-shot text classification method [J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2024, 60(1): 1-12. (in Chinese)
[32] GAO T, FISCH A, CHEN D. Making pre-trained language models better few-shot learners [EB/OL]. [2023-12-12]. http://arxiv.org/abs/2012.15723.

编辑 金胡考