

北京邮电大学学报

*Journal of Beijing University of Posts and Telecommunications*

ISSN 1007-5321, CN 11-3570/TN

## 《北京邮电大学学报》网络首发论文

题目：基于元多任务提示学习的零样本谣言检测方法  
作者：石宇，于宁，孙亚伟，刘建毅  
DOI：10.13190/j.jbupt.2023-270  
收稿日期：2023-12-21  
网络首发日期：2024-06-05  
引用格式：石宇，于宁，孙亚伟，刘建毅. 基于元多任务提示学习的零样本谣言检测方法[J/OL]. 北京邮电大学学报. <https://doi.org/10.13190/j.jbupt.2023-270>



**网络首发：**在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

**出版确认：**纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

# 基于元多任务提示学习的零样本谣言检测方法

石宇<sup>1</sup>, 于宁<sup>1</sup>, 孙亚伟<sup>1</sup>, 刘建毅<sup>2</sup>

(1. 北京邮电大学 可信分布式计算与服务教育部重点实验室, 北京 100876;

2. 北京邮电大学 网络空间安全学院, 北京 100876)

**摘要:** 针对现有谣言检测方法在微调大语言模型时导致的高内存占用, 以及提示学习方法对初始点选择敏感的问题, 提出一种基于元多任务提示学习的零样本谣言检测方法。首先, 基于提示学习范式调整零样本谣言检测任务目标, 通过设计提示模板使这一任务目标与大语言模型的训练任务目标保持一致, 以充分利用大语言模型积累的先验知识。其次, 利用元学习的参数更新策略以定位适用于零样本谣言检测任务的提示模板初始点, 从不同的元任务中学习通用知识来实现参数优化。最后, 引入情感分析作为辅助元任务进一步调整和优化模型参数。在多个公开数据集上进行了对比实验, 结果表明所提方法在零样本谣言检测任务中表现出色, 其性能指标优于基准方法。

**关键词:** 谣言检测; 提示学习; 元学习; 多任务学习

中图分类号: TP183

文献标志码: A

## Zero-Shot Rumor Detection Via Meta Multi-Task Prompt Learning

SHI Yu<sup>1</sup>, YU Ning<sup>1</sup>, SUN Yawei<sup>1</sup>, LIU Jianyi<sup>2</sup>

(1. Key Laboratory of Trustworthy Distributed Computing and Service, Ministry of Education,  
Beijing University of Posts and Telecommunications, Beijing 100876, China;

2. School of Cyberspace Security, Beijing University of Posts and Telecommunications, Beijing 100876, China)

**Abstract:** To address the issue of the vast amount of memory usage associated with fine-tuning large language models in existing rumor detection methods, and to tackle the sensitivity of prompt learning to its initial point, a meta multi-task prompt learning method for zero-shot rumor detection is proposed. First, the zero-shot rumor detection task objective is modified based on the prompt learning, and the prompt template is designed to make its task objective align compatible with the training task objective of large language models. This way can effectively exploit the potential capabilities of large language models. Second, the parameter update strategy based on meta-learning is employed to rapidly identify suitable initial points of the prompt template for zero-shot rumor detection, and the meta-knowledge is learned from different meta-tasks to achieve parameter optimization. Finally, sentiment analysis is introduced as an auxiliary meta-task to further model parameter optimization. Extensive experiments conducted on real-world datasets demonstrate that the proposed model outperforms baseline methods in zero-shot rumor detection tasks, achieving the best performance across various metrics.

**Key words:** rumor detection; prompt learning; meta learning; multi-task learning

收稿日期: 2023-12-21

基金项目: 国家自然科学基金项目(U21B2020); 北京邮电大学中央高校基本科研业务费项目(2021XD-A11-3)

作者简介: 石宇(1997—), 女, 博士生。

通信作者: 刘建毅(1980—), 男, 教授, 邮箱: liujy@bupt.edu.cn。

在信息化时代，社交网络信息的快速传播在给公众带来便利的同时也为谣言的传播提供了有利环境。谣言通常指那些与公众关注的对象、事件或问题有关的未经证实的说法或陈述。当谣言大规模传播时，容易引起公众的误解和恐慌，威胁社会稳定<sup>[1,2]</sup>。由于基于神经网络的方法依赖大量标注数据，当面对新兴的、难以迅速获取大规模带标注数据的未知领域任务时，往往表现不佳。此外，谣言具有领域特异性，基于旧领域数据训练的模型在检测新领域谣言时效果并不理想，因此零样本谣言检测任务更具挑战性。

零样本任务的核心是使模型在未接触特定任务数据的情况下，快速适应新任务和领域，降低人工标注成本。目前相关研究主要集中在基于微调的迁移学习上。如 Lin 等<sup>[3]</sup>提出一种基于传播结构的检测方法，利用领域自适应技术对模型进行微调以适应新领域。然而大语言模型通常拥有上百亿参数，动态微调过程需要消耗大量内存导致训练和部署成本较高，同时通用语言模型在特定领域和任务上表现不佳<sup>[4,5]</sup>。为降低成本并提升自动化水平，有学者提出一种基于提示学习的新颖范式<sup>[6]</sup>。该范式通过调整下游任务形式使其与预训练任务形式保持一致，充分利用预训练语言模型的优势，实现零样本学习。如 Lin 等<sup>[7]</sup>提出一种基于多语言预训练模型的分层提示编码机制来学习提示和谣言数据的语言不可知的上下文表示以进行零样本谣言检测。然而，提示学习方法在零样本谣言检测任务中的研究有限且提示模版初始点会影响模型性能。

为此，提出一种元多任务提示学习（MMP, meta multi-task prompt learning）的零样本谣言检测方法，旨在结合体系化人工智能和大语言模型的优势，提高零样本谣言检测任务的准确性。通过引入提示学习调整谣言检测任务目标以充分利用预训练语言模型的先验知识，并利用原型策略来总结类别信息以减少噪声数据的影响。利用元学习确定提示模板的最佳初始点，并将情绪分析作为辅助元任务以进一步优化模型参数。在公开数据集上的实验结果表明，与现有方法相比，MMP 方法的检测准确率更高。

## 1 MMP

MMP 方法的整体架构如图 1 所示，主要包含特征编码、提示学习和元多任务学习三部分。

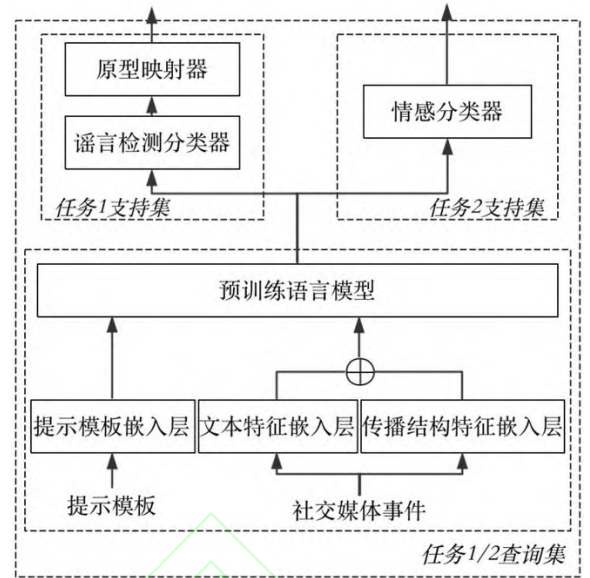


图 1 MMP 方法整体架构

### 1.1 特征编码

MMP 的特征编码阶段提取每个事件原帖和评论的文本特征及其传播结构特征。在此阶段选择大规模多语言预训练语言模型如跨语言鲁棒双向变换器模型 XLM-R（cross-lingual language model-RoBERTa）<sup>[8]</sup>进行特征编码。该阶段主要分为 2 步：嵌入表示和混合编码。

嵌入表示包含文本嵌入表示和传播结构嵌入表示。为了将离散文本  $\mathbf{X}_i$  映射到连续表示空间，查询 XLM-R 的语义嵌入表获取文本嵌入表示  $\mathbf{H}^x$ 。给定帖子  $\mathbf{x}_i$ ，沿传播树路径计算其与源贴  $\mathbf{x}_0$  之间的距离作为传播位置。传播结构嵌入表示由可训练学习的参数  $\mathbf{W}$  计算得到，即

$$\mathbf{H}^{\text{pos}} = \mathbf{W} \mathbf{d}_{\text{pos}} \quad (1)$$

其中  $\mathbf{d}_{\text{pos}}$  为帖子的传播位置。

在混合编码阶段，参考文献[6]，将  $\mathbf{H}^x$  和  $\mathbf{H}^{\text{pos}}$  相加得到最终事件的嵌入表示  $\hat{\mathbf{H}}$ ，将其输入至预训练语言模型中进行编码。

### 1.2 提示学习

提示学习过程可概括为 2 步：① 根据任务设计有效的提示模板并将数据嵌入至该模板中；② 构建提示学习映射器用于将模型的预测输出  $\hat{\mathbf{y}}$  映射到相应的标签  $\mathbf{y}$ ，实现谣言检测。

#### 1.2.1 提示模板

构建提示模板的方法有两类：手工设计模板和自动生成模板<sup>[9]</sup>。手工设计模板通常采用自然语言形式，由词汇表中的离散词汇组成，其设计过程耗时且性能在很大程度上依赖预定义提示模板的选择。自动生成模板则自动选取适合模型的模板来优化任务

效果,其性能在很大程度受初始设置点的影响,不恰当的初始点会影响模型的性能。

为了提升模型对自然语言的理解能力,采用自动生成模板的方法,同时结合元学习在连续的提示空间中寻找提示模板的初始点。具体而言,将提示模板 $\mathbf{p}$ 与事件特征表示 $\mathbf{H}$ 进行拼接形成最终输入表示,并将最终输入表示传递到预训练语言模型中,得到特征表示 $\mathbf{H}^m$ 。接着进行元多任务学习得到输出表示 $\hat{\mathbf{y}}$ ,将 $\hat{\mathbf{y}}$ 输入到提示学习映射器中,得到最终预测分类结果。

### 1.2.2 提示学习映射器

映射器的主要功能是将输出表示 $\hat{\mathbf{y}}$ 映射到预设标签空间,计算概率值得到事件的最终预测标签 $\mathbf{y}$ 。理想的映射器应能封装类级语义特征,然而在缺乏特定任务先验知识的情况下,选择合适的映射器是一项具有挑战性的任务。原型指从实例中总结出来的包含更广泛语义信息的类级语义向量,代表了类别的整体信息。通过估算每个类别的原型向量并将其作为映射器,可以减轻模型对单个样本的依赖<sup>[10]</sup>。具体而言,在训练阶段,为每个类别构建一个可学习的原型映射器 $\mathbf{V}_y$ ,从原型级和实例级 2 个维度引入对比学习,优化模型并提高其准确率。

1) 原型级对比损失。隶属同一标签的事件表示与原型映射器展现出更高的相似度。给定事件表示 $\hat{\mathbf{y}}_i$ ,采用最小化负对数似然作为原型级对比损失函数以量化原型与事件表示之间的相似度:

$$L_p = -\ln \frac{\exp(S(\hat{\mathbf{y}}_i, \mathbf{V}_{y_i}))}{\sum_j \exp(S(\hat{\mathbf{y}}_i, \mathbf{V}_{y_j}))} \quad (2)$$

其中:  $y_i, y_j$  分别为第  $i, j$  个事件的真实标签,  $S(\cdot)$  为余弦相似度函数。

2) 实例级对比损失。隶属同一类别的事件表示 $\hat{\mathbf{y}}$ 具有较高的相似性,即通过最小化实例级对比损失增加同类事件间的相似性且减少不同类别事件间的差异性。给定事件表示 $\hat{\mathbf{y}}_i$ ,实例级对比损失为

$$L_i = -\frac{1}{N_{y_i}} \sum_j \mathbb{I}_{[i \neq j]} \mathbb{I}_{[y_i = y_j]} \ln \frac{\exp(S(\hat{\mathbf{y}}_i, \hat{\mathbf{y}}_j))}{\sum_{j'} \mathbb{I}_{[i \neq j']} \exp(S(\hat{\mathbf{y}}_i, \hat{\mathbf{y}}_{j'}))} \quad (3)$$

其中:  $N_{y_i}$  为标签为  $y_i$  的事件数量,  $\mathbb{I}$  为示性函数。

提示学习映射器的最终优化目标 $L_{\text{ver}}$ 由 $L_p$ 和 $L_i$ 相加得到。

## 1.3 元多任务学习

为优化自动生成模板过程,使用元学习方法在

连续提示空间中寻找模板初始点并进行迭代更新。元学习的核心在于训练一系列任务,每个任务的训练周期包含内循环和外循环 2 个阶段,多次迭代以获取快速适应的元知识。通过这种策略,设计一个既能提取谣言特性,又能理解情感的有效模型,进而在谣言检测任务中达到更高的准确度。

模型无关元学习<sup>[11]</sup>是一种基于梯度的高效的元学习方法,其核心在于从可获取的源数据中抽样出多个元提示任务以寻找适合模型的初始化参数,实现对新任务的快速适应。考虑到情感在谣言检测中的重要性,在元训练阶段构建了谣言检测和情感分析 2 个元任务。可将模型参数划分为跨任务共享的元参数 $\phi$ 和特定的任务参数 $\theta^k, k \in \{1, 2\}$ 为任务序号,分别对应谣言检测和情感分析。元参数 $\phi$ 包含提示模板,在元训练过程中进行优化并在任务之间实现知识转移和共享,任务参数 $\theta^k$ 则针对相应的任务进行调整以捕捉该任务的特殊性。

在元学习过程中,为每个任务分别构建支持集和查询集,采用内外循环交错训练的机制为不同任务获得有效的初始化参数。

在内循环阶段,对于特定任务 $k$ ,其任务参数 $\theta^k$ 基于该任务支持集的特征输出 $\mathbf{H}^m$ 进行初始化

$$\theta^k = \mathbf{W}^k f_{\text{svd}}(\mathbf{H}^m) \quad (4)$$

其中:  $\mathbf{W}^k$  为可训练的参数权重,  $f_{\text{svd}}(\cdot)$  为用于提取重要特征的奇异值分解函数。任务 $k$ 的输出表示为

$$\hat{\mathbf{y}}^k = f_{\theta^k}(\mathbf{H}^m) \quad (5)$$

其中:  $f_{\theta^k}(\cdot)$  为特定于 $k$ 任务的函数,结合梯度下降法计算最小化损失以找到每个任务的最优参数。

对于谣言检测任务,采用交叉熵损失来计算预测输出表示与真实标签之间的差异:

$$L_{\text{rd}} = -\sum_{i=1} y_i \ln \hat{y}_i \quad (6)$$

其最终损失函数由 $L_{\text{rd}}$ 和 $L_{\text{ver}}$ 联合构成:

$$L^{k=1} = \text{mean}(L_{\text{rd}} + L_{\text{ver}}) \quad (7)$$

对于情感分析任务,利用文献[12]的方法获取事件情感作为真实情感标签,计算预测输出与真实情感标签之间的交叉熵损失 $L^{k=2}$ 以衡量情感分布差异。

特定任务参数 $\theta^k$ 在元学习的内循环阶段依赖支持集进行更新,避免对单一任务过拟合且节省内存。

在外循环阶段,元参数 $\phi$ 则通过优化每个任务查询集上的损失梯度实现更新

$$\phi = \phi - \beta(\nabla_{\phi} L^k) \quad (8)$$

其中:  $\beta$  为外循环阶段的学习率,  $\nabla_{\phi}$  为梯度。这一阶



段旨在增强参数 $\phi$ 的适应性,使其在元测试阶段更好地适应目标任务。

## 2 实验结果与分析

### 2.1 数据集

为验证所提出方法的有效性,使用 4 个来自主流社交媒体平台的公开数据集进行实验,分别为 Twitter、Weibo、Twitter-COVID19 和 Weibo-COVID19。其中, Twitter 和 Twitter-COVID19 数据集基于 Twitter Academic API 接口完成收集,原帖及其传播路径通过 twarc2 工具完成获取。Weibo 和 Weibo-COVID19 数据集包括新浪社区管理中心收集的谣言事件和随机筛选的未被报道为谣言的非谣言事件。在实验设置中,将与 COVID-19 相关的突发事件数据集视为零样本目标数据集。当 Twitter-COVID19 数据集为目标数据集时,将 Weibo 数据集用作源数据集;当 Weibo-COVID19 数据集为目标数据集时,将 Twitter 数据集用作源数据集。上述 4 个数据集均包含 2 种二进制标签,其中,谣言为可能包含不准确或误导性信息的言论;非谣言为不包含不准确或误导性信息的言论。

### 2.2 实验参数设置与评价指标

实验的计算环境为 NVIDIA Tesla P100,所有方法均基于 PyTorch 框架实现。在模型训练阶段,关键参数设置如下:学习率为  $10^{-5}$ ,批处理大小为 16,使用 AdamW 优化器对模型参数进行优化。为了全面评估方法的性能,选用准确率和 F1 值作为主要评估指标,其中,准确率用于衡量方法在所有事件中正确预测谣言事件比例;F1 值是精确度和召回率的调和平均值,是整体评估方法性能的指标。

### 2.3 基准方法

为验证 MMP 方法在零样本谣言检测任务的优势,选取多个基线方法进行对比分析,这些方法可被分为 2 类:基于微调的方法和基于提示学习的方法。

基于微调的方法利用有标签的源数据集对模型进行针对性微调。标准微调(VFT, vanilla fine-tune)方法<sup>[13]</sup>在预训练语言模型顶部添加特定任务的分类模型以进行谣言检测。基于翻译的微调(TFT, fine-tune based on translate)方法<sup>[14]</sup>将源数据集中的事件翻译成目标语言,使用翻译后的谣言数据微调模型以适应目标数据集。基于对比学习的微调(CFT, fine-tune based on contrastive learning)方法<sup>[3]</sup>综合考虑传播结构和对比学习以检测跨语言和跨领域的谣言,

实现在零样本场景下对模型微调。由于谣言检测的任务目标与预训练语言模型的任务目标存在差异,这些依赖迁移学习技术微调模型的零样本谣言检测方法不能充分利用预训练语言模型的先验知识。

基于提示学习的方法是一种新兴学习范式,核心思想是将下游任务目标转化为与预训练语言模型的任务目标一致的形式。基于源数据集的提示调整(SPT, prompt-based tuning in source)方法<sup>[15]</sup>通过源数据集和手工设计提示模板来训练模型,应用至目标数据集。基于目标数据集的提示调整(TPT, prompt-based tuning in target)方法<sup>[16]</sup>通过源数据集和手工设计提示模板来训练模型,之后将提示模板其翻译为目标语言以适应目标数据集。基于可调令牌的提示调整(Soft-PT, prompt-based tuning based on tunable tokens)方法<sup>[17]</sup>采用自动生成提示模板的方法,即利用可调令牌代替离散令牌作为提示模板。基于回复感知的提示学习(RPL, response-aware prompt learning)方法<sup>[7]</sup>兼具手工设计模板和自动生成提示模板的优点,通过微调多语言预训练模型以生成与语言无关的提示模板表征向量。

### 2.4 基准方法对比实验

所提 MMP 方法和基线方法的参数量及在目标数据集 Weibo-COVID19 和 Twitter-COVID19 上对于谣言检测的性能如表 1 所示。可见,第 1 组基于微调的方法谣言检测的性能表现较差,这是因为这组方法在进行微调时严重依赖下游任务目标,缺乏泛化能力,导致方法在目标数据集上的性能不佳。第 2 组基于提示学习的方法在性能上优于第 1 组,但其性能仍受到限制,主要是基于源数据集训练的提示方法在适应目标数据集时出现偏差。其次当涉及到翻译提示模板时,机器翻译的不准确性可能会降低 TPT 方法的性能。此外在目标数据集上训练提示模板需要大量已标注的目标数据才能实现充分优化,而在零样本学习场景下获取这些数据是不切实际的。

此外,在参数量方面, MMP 方法在资源占用方面有明显优势。尽管 RPL 方法在个别性能指标上具有竞争力,但其参数量远远大于 MMP 方法,原因在于 RPL 方法对 2 个包含嵌入层的编码器进行了微调,分别从提示模板和数据两方面学习与语言无关的上下文表示。这意味着 MMP 方法在计算资源和存储需求方面更有优势,即在保持较低复杂度的同时,实现与更复杂的 RPL 方法相媲美甚至更优的性能。此外, MMP 方法具有更高的灵活性,与 RPL 方法不同,

MMP 方法不需要对预训练语言模型进行微调便能够轻松地将预训练语言模型替换为其他大语言模型,这一特性使其在实际应用中更加便捷。

表 1 各谣言检测方法的参数量及在 2 个目标数据集上检测性能

方法	参数量 /MB	Weibo-COVID19		Twitter-COVID19	
		准确率	F1 值	准确率	F1 值
VFT	278.05	0.623	0.585	0.603	0.602
TFT	278.05	0.639	0.567	0.634	0.574
CFT	279.16	0.656	0.582	0.653	0.644
SPT	278.05	0.664	0.648	0.670	0.616
TPT	278.05	0.650	0.489	0.674	0.651
Soft-PT	278.05	0.652	0.574	0.685	0.652
RPL	513.12	<b>0.713</b>	0.675	0.715	<b>0.689</b>
MMP	278.20	0.712	<b>0.693</b>	<b>0.724</b>	0.682

总体来说, MMP 方法不仅在有效适应目标数据集方面具有显著优势, 而且能够充分利用大模型知识, 实现无需增加新模块的体系化人工智能设计思想。在保持方法复杂度和参数数量的同时, 展现出优越的性能, 这对进一步探索和改进基于提示的学习方法具有重要的意义。

2.5 预训练语言模型对比实验

现有大模型大致可分为生成式和判别式两类<sup>[18]</sup>。大规模多语言预训练文本到文本转换器 (mT5, massively multilingual pre-trained text-to-text transformer)<sup>[19]</sup>作为生成式预训练模型, 与判别式模型 XLM-R 在模型构架和训练目标上存在差异。为了验证 MMP 方法在不同预训练语言模型中的稳健性和普适性, 使用 mT5 替换 XLM-R 进行特征编码。在 2 种预训练语言模型下, MMP 方法和对比方法在目标数据集 Weibo-COVID19 和 Twitter-COVID19 上的性能表现如表 2 所示。

表 2 使用不同预训练语言模型各谣言检测方法在 2 个目标数据集上的检测性能

方法	Weibo-COVID19		Twitter-COVID19	
	准确率	F1 值	准确率	F1 值
XLM-R+SPT	0.664	0.648	0.670	0.616
XLM-R+Soft-PT	0.652	0.574	0.685	0.652
XLM-R+MMP	<b>0.712</b>	<b>0.693</b>	<b>0.724</b>	<b>0.682</b>
mT5+SPT	0.712	0.705	0.783	0.778
mT5+Soft-PT	0.692	0.691	0.809	0.802
mT5+MMP	<b>0.816</b>	<b>0.773</b>	<b>0.864</b>	<b>0.860</b>

由表 2 可知, 与 SPT 和 Soft-PT 方法相比, MMP 方法在集成 mT5 和 XLM-R 这 2 种不同类型的预训练语言模型时, 均展示了卓越的性能, 证明了 MMP 方法的有效性和灵活的适应能力。其中, 集成了 mT5 的 MMP 方法表现优于使用 XLM-R 的 MMP 方法, 这是因为 mT5 使用了规模更大、内容更多样化的多语言数据集进行预训练, 提供了更灵活的文本到文本的任务适应能力。这一比较展示了 MMP 方法的优势, 为未来研究提供了重要参考。

2.6 消融实验

在完整 MMP 方法的基础上, 设置 3 种消融实验, 通过去除特定模块来验证其对 MMP 方法整体性能的影响。(1) -w/o 元学习: 去除元学习模块, 改为使用随机初始化提示模板; (2) -w/o 任务 2: 不执行任务 2, 即去掉情感分析任务仅利用谣言检测任务完成参数调优; (3) -w/o 原型: 移除原型模块, 改用标签代替原型映射器。将以上方法在目标数据集 Weibo-COVID19 和 Twitter-COVID19 进行谣言检测实验, 结果如表 3 所示。

表 3 MMP 方法各模块的消融实验结果

方法	Weibo-COVID19		Twitter-COVID19	
	准确率	F1 值	准确率	F1 值
-w/o 元学习	0.677	0.651	0.684	0.634
-w/o 任务 2	0.697	0.684	0.702	0.657
-w/o 原型	0.689	0.670	0.691	0.668
MMP	<b>0.712</b>	<b>0.693</b>	<b>0.724</b>	<b>0.682</b>

由表 3 可知, 与完整的 MMP 方法相比, 去除元学习模块会导致准确率和 F1 值下降, 这表明元学习对于寻找合适的提示模板初始点至关重要。当去除任务 2 时, 所提方法性能有所下降, 突显情感在谣言检测中的重要作用。当原型策略被移除时, 所提方法 MMP 性能同样下降, 这表明包含更多信息的原型提示映射器能将输出表示映射到正确标签。这一结果表明, MMP 中的所有模块在提高零样本谣言检测任务性能方面都发挥了重要作用, 证明了方法设计中每一步的必要性和有效性。

3 结束语

为有效利用大语言模型的先验知识, 同时提高模型在提示学习初始化阶段的适应性, 提出 MMP 零样本谣言检测方法。通过引入提示学习调整谣言检测任务目标, 以更好地适应预训练语言模型的任务目标, 充分利用大语言模型的知识。借助元学习有效

确定最佳的提示模板初始点,并将情感分析作为辅助元任务进一步优化模型参数。在多个公开数据集上的实验结果表明,MMP方法在零样本谣言检测任务中表现优异,且无需对大语言模型进行参数微调,有效地降低了成本。

未来将进一步优化 MMP 方法,探讨其在处理更复杂的谣言检测任务时的可扩展性和适应性。同时,将深入分析不同事件背后的逻辑、因果关系和事实证据等信息,以丰富和完善当前研究成果。

## 参考文献:

- [1] GUO B, DING Y S, YAO L N, et al. The future of false information detection on social media: new perspectives and trends[J]. *ACM Computing Surveys*, 2020, 53(4): 68:1-36.
- [2] HU L M, WEI S Q, ZHAO Z W, et al. Deep learning for fake news detection: a comprehensive survey[J]. *AI Open*, 2022, 3: 133-155.
- [3] LIN H Z, MA J, CHEN L L, et al. Detect rumors in microblog posts for low-resource domains via adversarial contrastive learning[C]// *Findings of the Association for Computational Linguistics: NAACL 2022*. Stroudsburg: ACL, 2022: 2543-2556.
- [4] 车万翔, 窦志成, 冯岩松, 等. 大模型时代的自然语言处理:挑战、机遇与发展[J]. *中国科学:信息科学*, 2023, 53(9): 1645-1687.  
CHE W X, DOU Z C, FENG Y S, et al. Towards a comprehensive understanding of the impact of large language models on natural language processing: challenges, opportunities and future directions[J]. *Scientia Sinica(Informationis)*, 2023, 53(9):1645-1687.
- [5] 武俊宏, 赵阳, 宗成庆. ChatGPT 能力分析与未来展望[J]. *中国科学基金*, 2023, 37(5): 735-742.  
WU J H, ZHAO Y, ZONG C Q. Analysis of ChatGPT's capabilities and future prospects[J]. *Bulletin of National Natural Science Foundation of China*, 2023, 37(5): 735-742.
- [6] SCHICK T, SCHÜTZE H. Exploiting cloze-questions for few-shot text classification and natural language inference[C]// *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics*. Stroudsburg: ACL, 2021: 255-269.
- [7] LIN H Z, YI P Y, MA J, et al. Zero-shot rumor detection with propagation structure via prompt learning[C]// *Proceedings of the AAAI Conference on Artificial Intelligence*. Menlo Park: AAAI, 2023, 37: 5213-5221.
- [8] CONNEAU A, KHANDELWAL K, GOYAL N, et al. Unsupervised cross-lingual representation learning at scale[C]// *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Stroudsburg: ACL, 2020: 8440-8451.
- [9] LIU X, JI K X, FU Y C, et al. P-tuning: prompt tuning can be comparable to fine-tuning across scales and tasks[C]// *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*. Stroudsburg: ACL, 2022: 61-68.
- [10] CUI G Q, HU S D, DING N, et al. Prototypical verbalizer for prompt-based few-shot tuning[C]// *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*. Stroudsburg: ACL, 2022: 7014-7024.
- [11] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks[C]// *Proceedings of the 34th International Conference on Machine Learning*. New York: PMLR, 2017: 1126-1135.
- [12] ZHANG X Y, CAO J, LI X R, et al. Mining dual emotion for fake news detection[C]// *Proceedings of the Web Conference 2021*. New York: ACM Press, 2021: 3465-3476.
- [13] DEVLIN J, CHANG M-W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C]// *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Stroudsburg: ACL, 2019: 4171-4186.
- [14] DU J S, DOU Y T, XIA C Y, et al. Cross-lingual COVID-19 fake news detection[C]// *2021 International Conference on Data Mining Workshops (ICDMW)*. Piscataway, NJ: IEEE Press, 2021: 859-862.
- [15] LIN X V, MIHAYLOV T, ARTETXE M, et al. Few-shot learning with multilingual generative language models[C]// *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. Stroudsburg: ACL, 2022: 9019-9052.
- [16] ZHAO M J, SCHÜTZE H. Discrete and soft prompting for multilingual models[C]// *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Stroudsburg: ACL, 2021: 8547-8555.
- [17] LESTER B, AL-ROUFU R, CONSTANT N. The power of scale for parameter-efficient prompt tuning[C]// *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Stroudsburg: ACL, 2021: 3045-3059.
- [18] HAN J L, ZHAO S, CHENG B, et al. Generative prompt tuning for relation classification[C]// *Findings of the Association for Computational Linguistics: EMNLP 2022*. Stroudsburg: ACL, 2022: 3170-3185.
- [19] XUE L T, CONSTANT N, ROBERTS A, et al. mT5: a massively multilingual pre-trained text-to-text transformer[C]// *Proceedings of the 2021 Conference of the*

North American Chapter of the Association for  
Computational Linguistics: Human Language Technologies.  
Stroudsburg: ACL, 2021: 483-498.

