

基于提示增强原型网络的小样本多标签方面类别检测

管超峰,白宇*,周贤雷

(沈阳航空航天大学 计算机学院,辽宁 沈阳 110136)

摘要:小样本多标签方面类别检测是细粒度情感分析的研究热点。在基于原型网络的方法中,训练数据缺乏以及与当前方面类别无关的噪音词严重影响了采用注意力机制生成原型向量的质量。针对这一问题,本文提出了基于提示增强原型网络模型,首先,利用提示学习对齐预训练任务与下游任务,同时借助提示信息指导模型进行句子表示,从而学习到更具有辨别性的向量,有效地促使类别信息易区分,并采用余弦相似度计算损失,降低高维向量空间的影响;其次,设计了减轻噪音对句子向量表示干扰的优化模型,促进相同方面类别的句子在嵌入空间中聚集。实验结果表明:该模型在三个公开数据集的F1值比当前最优(state-of-the-art, SOTA)模型分别提升了4.35%, 8.62%, 8.39%,可以有效的检测方面类别。

关键词:方面类别检测;原型网络;提示学习;多标签分类;元学习

中图分类号:O436

文献标志码:A

文章编号:0253-2395(2024)03-0494-12

Few-shot Multi-label Aspect Category Detection Based on Prompt-Enhanced Prototypical Network

GUAN Chaofeng, BAI Yu*, ZHOU Xianlei

(Shenyang Aerospace University Computer College, Shenyang 110136, China)

Abstract: Few-shot multi-label aspect category detection is a prominent area of research within fine-grained sentiment analysis. In methods based on prototypical networks, the lack of training data and the presence of irrelevant noise words severely impact the quality of prototype vector generation using attention mechanisms. Addressing this challenge, our study introduces the ProtPrompt model, an innovative prototype network enhanced through prompt learning. Firstly, we align pre-trained tasks with downstream tasks using prompt learning, guiding the model in precise sentence representation to learn more discriminative vectors. This effectively enhances category distinguishability. Meanwhile, we utilize cosine similarity to calculate the loss function instead of the commonly used Euclidean distance in prototypical networks, mitigating the influence of high-dimensional vector spaces. Secondly, we design an optimized framework to attenuate noise interference on sentence vector representation, thereby fostering the aggregation of sentences sharing the same aspect category within the embedding space. Experimental results validate the effectiveness of our proposed ProtPrompt model on three publicly available datasets. The experimental results show that the model improves the F1 score over the state-of-the-art (SOTA) model by 4.35%, 8.62%, and 8.39% for the three publicly available datasets, respectively. This substantiates its ability to efficiently detect aspect categories.

Key words: aspect category detection; prototypical network; prompt learning; multi-label classification; meta-learning

收稿日期:2023-12-15;接受日期:2024-01-12

基金项目:国家自然科学基金(U1908216)

作者简介:管超峰(1997-),男,山东潍坊人,硕士研究生,研究方向为自然语言处理。E-mail:chaofengya@163.com

* 通信作者:白宇(BAI Yu),E-mail:baiyu@sau.edu.cn

引文格式:管超峰,白宇,周贤雷.基于提示增强原型网络的小样本多标签方面类别检测[J].山西大学学报(自然科学版),2024,47(3):494-505. DOI:10.13451/j.sxu.ns.2024002

0 引言

方面类别检测 (Aspect Category Detection, ACD) 是细粒度情感分析 (Aspect-Based Sentiment Analysis, ABSA)^[1] 的一个重要子任务, 旨在从预定义的方面类别集合中检测评论句子中提及的方面类别, 从而更加简洁地描述观点目标。该任务的目标是将给定的评论句子分类为预定义的一组方面类别^[2]。方面类别检测在电子商务领域逐渐引起广泛关注, 对商业智能具有重要价值^[3-4]。

在方面类别检测的早期研究阶段, 研究者通常使用有监督分类器对提取的特征进行分类^[5]。受到大规模数据集的人工标注成本的限制, 这种方法在数据量匮乏的领域难以推广应用。小样本学习 (Few-Shot Learning, FSL)^[6] 提供了一种新的方法, 通过利用以往的类别中学到的先验知识来有效地识别新的类别, 从而打破标注数据匮乏的限制。

在小样本学习领域, 元学习 (Meta-Learning) 是一种代表性的方法^[7-8], 旨在通过学习跨不同任务的共享特征, 快速适应新任务。常用的元学习方法包括孪生网络 (Siamese Network)^[9]、匹配网络 (Matching Network)^[10]、原型网络 (Prototypical Network)^[11]、关系网络 (Relation Network)^[12] 等。其中, 原型网络的基本思想是将一种类别的一组样本向量映射为一个原型向量, 并通过计算查询样本与每个原型向量之间的距离进行分类。这种基于距离的小样本学习方法取得了令人瞩目的结果, 原型网络也因其原理清晰高效而成为小样本学习中最有效的方法之一。

原型网络的关键步骤之一是样本的向量表示, 并且十分依赖向量表示的质量^[13]。随着模型规模的增大, 对高质量向量表示的需求变得更加迫切。尽管预训练语言模型^[14]在一定程度上提高了向量表示的质量。但未处理的数据直接微调这些预训练模型, 导致预训练目标与下游任务之间存在差异, 限制了预训练语言模型的性能^[15-16]。

本文提出了一种基于提示增强的原型网络模型 (ProtPrompt), 旨在利用提示学习、注意力机制以及相似性度量来增强原型网络模型中句

子向量表示之间的辨别性, 实现小样本多标签方面类别检测。首先, 根据预训练任务使用提示模板, 将下游任务变成完形填空任务, 以促使模型学习更多与提示信息相关的特征和知识; 在此基础上, 通过注意力机制优化支持集各原型向量和查询向量的表示; 对于方面类别的判定, 将向量之间的欧氏距离度量转换为利用余弦相似度计算的相似性度量, 提升模型对分类特征分布的敏感性, 从而提高了模型的性能。

本文的主要贡献如下:

(1) 将提示学习方法引入原型网络的原型向量和查询向量表示, 并针对该表示方法提出利用查询向量与原型向量的余弦相似度作为方面类别判定依据。实验结果表明, 这一机制显著提高了模型的准确性, 使得模型能够更好地适应小样本多标签方面类别检测任务。

(2) 提出了一种基于多维注意力机制的句子向量表示优化模型。该模型通过多维注意力机制增强支持集样本间的共享信息并抑制噪声和不重要的信息, 从而提高了模型的准确性和鲁棒性。此外, 在查询向量优化过程中, 通过注意力计算引入原型向量信息, 增强了查询向量的多标签语义表示, 使其在小样本多标签数据集上表现出显著的有效性。

(3) 对本文提出的基于提示增强原型网络的小样本多标签方面类别检测模型在3个公开数据集上进行了实验评估, 结果表明, 本文提出的模型相较于当前最先进的模型获得了更好的表现。这进一步证明了本文方法的优越性和实用性, 以及在小样本多标签方面类别检测任务中的潜力。

1 相关工作

方面类别检测的相关研究主要包括两种类型: 无监督方法和有监督方法。无监督方法通过互信息^[17]或共现频率^[18-19]挖掘语义关联的方法^[20]来检测方面类别。这些方法需要大量的语料资源, 性能有限。监督方法通过手工设计特征^[21]、表示学习^[22]、主题-注意力网络^[23]或多任务学习^[24-25]来识别不同的方面类别。这些方法从预定义集合中检测方面类别, 但它们

严重依赖于通过大量标记数据训练一个分类器,且无法处理未知类别。这些挑战促进了小样本场景下的方面类别检测研究。

小样本学习^[26-27]借鉴了人类的学习过程,通过结合先验知识,能够在有限的监督信息下快速获取新知识。元学习已成功应用于解决小样本学习问题,主要包括基于模型的方法^[28-29]、基于优化的方法^[11, 30-31]和基于度量的方法。

基于度量的方法,如匹配网络、原型网络、关系网络等,通过在特征嵌入空间中测量实例之间的距离来评估它们之间的相似性。其中,原型网络由于其简单且有效,成为小样本学习中最受欢迎的方法之一。

传统的小样本学习主要关注单标签分类任务^[32]。Snell等^[33]基于原型网络,解决小样本场景下方面类别检测问题。该方法设计用于单标签小样本,它尝试利用注意机制减轻支持集和查询集其他方面带来的负面影响。

在许多实际场景中,一个样本往往具有多个标签,这就产生了小样本多标签学习。然而,小样本多标签的研究尚未充分开展。以往的工作主要集中在图像合成^[34]和信号处理^[35]方面。在自然语言处理领域,Hou等^[23]提出了在小样本场景中解决多标签意图检测任务的方法。基于核回归和逻辑回归的元校准阈值机制利用先前的领域经验和新的领域知识来估计阈值。

预训练语言模型(Pre-trained Language Model, PLM)充分利用大规模未标记文本来挖掘丰富的语义信息^[36-37]。但由于预训练目标与任务存在差异,可供微调的样本有限,此外在某些任务上受词汇差异和噪声信息的影响,导致其在向量表示效果不佳,影响句子向量表示的质量^[13, 16]。Liu等^[15]通过设计不同自然语言处理任务的特定描述(即提示),提升模型预测准确性。

提示学习在获取向量表示方面表现出色,在语义表达、模型性能和小样本学习等方面有显著优势,Davide等^[9]验证了提示学习在小样本学习中的有效性。Gu等^[39]通过在预训练阶段预训练软提示,提高了提示调整的效果。相

比传统的完整模型微调,在全数据和小样本学习情况下,模型在下游任务中表现出更好的性能。Hu等^[31]通过将外部知识库整合进来,扩展和优化标签词空间,使得在小样本情况下获得更好的性能。

基于这些问题,本文提出了一种利用提示学习增强的原型网络方法,通过使用提示模板来获得更好的向量表示,从而进一步提高方面类别检测任务的性能。

2 小样本多标签方面类别检测模型

2.1 问题定义

在小样本多标签方面类别检测任务中,目标是利用少量支持集样本正确地将查询样本进行分类。由于每个样本可能包含多个方面类别的信息,因此该任务可以定义为一个多标签文本分类问题。

在小样本方面类别检测任务中,首先在一组源域(任务) D 上训练,然后应用在一组不经过微调的目标域(任务) D' 中。具体来说,通过支持集学习句子嵌入向量空间的映射,将查询样本正确分类到新学习的类别中。在支持集和查询集中的样本共享 N 个类别 $D^n = \{x_1^n, \dots, x_k^n\}, n \in N, k \in K$,其中每类都包含 K 个样本(也称为 N -way K -shot)。原型向量由支持集映射生成 N 个原型向量用于表示类别,查询样本根据距离判断是否属于当前类别。目标是预测 $y \in Y$ 对应于每个查询样本的标签,其中 Y 是一组可能包含 N 个不同类的标签集合。

2.2 模型概览

本文设计了基于提示增强原型网络的小样本多标签方面类别检测模型。该模型通过结合提示学习、向量表示优化和余弦相似性度量,有效提升句子向量表示的质量和模型性能,增强了共享信息并抑制了噪声,进一步提升了模型的准确性和鲁棒性,同时在查询向量优化中引入了原型向量信息,增强了多标签语义表示效果。该方法的总体框架如图1所示。

2.3 提示增强

为了保持预训练任务目标和下游任务目标的一致性,本文将下游任务转化为掩码语言模

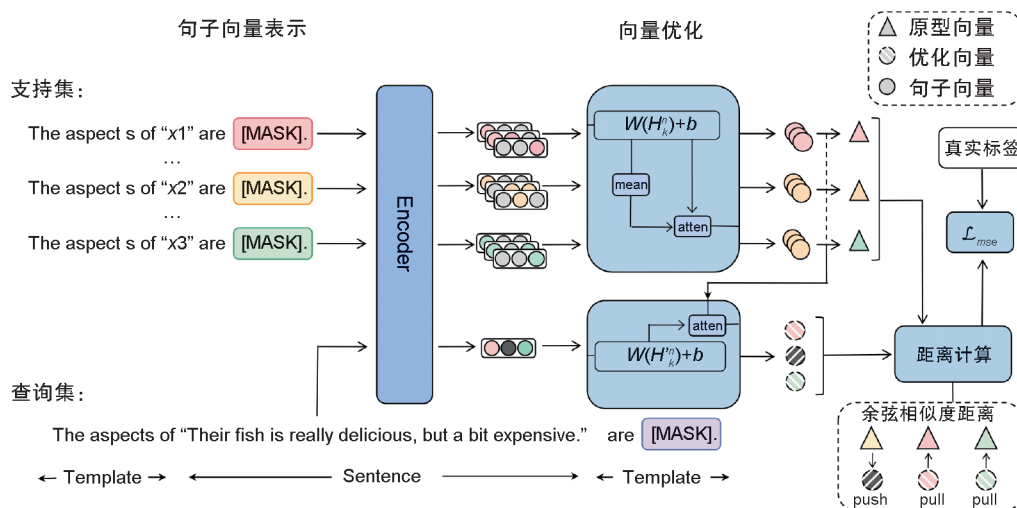


图1 提出方法的整体架构

左侧部分描述了支持和查询样本,其实彩色表示目标方面,灰色表示无关方面造成的噪音;右侧部分展示了向量优化的网络结构以及训练细节。

Fig. 1 The overall architecture of the proposed method

The left part describes the support and query samples, which are actually colored to indicate the target aspect and gray to indicate the noise caused by irrelevant aspects; The right part shows the network structure for vector optimization and the training details.

型的任务,其方法是使用模板改变输入为 $x_{\text{prompt}} = \text{"The aspects of } [x] \text{ are } [\text{MASK}] \text{"}$, 其中 $[x]$ 表示句子的占位符。句子的长度为 m 。将 x_{prompt} 输入到预训练模型获得 h 作为句子的向量。

提示学习常用的句子表示方法有两种,第一种是将 $[\text{MASK}]$ 位置的隐藏层状态向量作为该句子的向量,第二种是使用掩码语言模型在 $[\text{MASK}]$ 位置预测 $\text{top}k$ 个词,根据词预测的概率,进行加权求和表示该句子。从词预测的概率判断是否属于前 k 个集合的对应映射是不连续的,使得难以进行微调。因此选择第一种方式表示句子:

$$h = E(x_{\text{prompt}}). \quad (1)$$

借助任务特定的提示信息,提示学习能够指导模型对句子进行精确表示,提取文本中与提示高度相关的内容信息。

2.4 向量表示优化

在计算句子的向量表示时,本文针对隐藏层状态 $h_{[\text{MASK}]}$ 将句子中各个方面信息映射为一个句子向量提出了一种新方法。通过学习新的映射关系,将当前句子向量映射到多个维度,从而缓解了区分不同方面信息时信息相对有限的问题。随后采用注意力机制降低不相关维度

信息的影响,更好地区分不同方面信息。具体表达式如公式2所示。

$$v_i^n = \tanh(W \cdot h_i^n + b), \quad (2)$$

公式中 n 表示类别的数量, k 表示每个类别中句子的数量, $V = \{v_1^n, \dots, v_k^n\}$ 表示每个句子的向量集合, W 是权重参数, b 是偏置项。通过对句子的隐藏状态进行学习和优化,能够得到句子在不同类别下的语义表示,从而使其在不同类别之间具有更好的区分性。这种隐藏状态的学习过程使得句子在多个类别上的信息得以分解和凸显,有助于提高句子向量的区分能力和表达质量。

在公式3中,采用平均池化函数 g 对每个类别中的句子向量进行平均, $M = \{m^1, \dots, m^n\}$ 作为当前类别的初始化向量。通过这种方式能够有效地汇聚句子在同一类别下的信息,同时可以起到稀释作用,并将其用于初始化当前类别的向量。

$$m^n = g(v_1^n, \dots, v_k^n). \quad (3)$$

这种稀释操作能够提高句子向量的稳定性和表达质量,从而更好地支持多类别的句子表示和应用。

在模型进行标签预测时,每个句子信息对预测结果的贡献并不相同。为了解决这个问

题,公式4和公式5引入了注意力机制,通过为不同类别的句子分配权重,以更好地区分不同类别的信息:

$$z_i^n = m^n \cdot v_i^n, \quad (4)$$

$$\alpha_i^n = \frac{\exp(z_i^n)}{\sum_{j=1}^k \exp(z_j^n)}. \quad (5)$$

公式6通过对同一类别中的所有句子进行加权平均,得到该类别的原型向量 $C = \{c^1, \dots, c^n\}$ 。同样,对于查询样本公式7通过计算其与每个类别原型向量的注意力权重,并进行加权求和,得到查询向量 $Ct = \{ct^1, \dots, ct^n\}$ 。

$$c^n = \sum_{i=1}^k \alpha_i^n \cdot v_i^n, \quad (6)$$

$$ct^n = \sum_{i=1}^k \alpha_i^n \cdot ct_i^n. \quad (7)$$

这种方法能够更好地捕捉同一类别中每个句子的共性特征,并为距离度量提供更精准的句子表示。

本文采用多维度映射与注意力机制的方法,能够对句子中的所有信息的重要性进行量化,并更有效地生成句子的表示。即使句子包含多个方面的信息,也能更加注重有效信息来生成句子的最终表示,从而提高了原型向量的计算效果。通过引入注意力机制,将更多的注意力集中在与预测相关的信息上,有效地优化句子表示的质量,进而提升模型在复杂任务中的表现。

2.5 自适应阈值

在多标签分类任务中,阈值是评估性能指标的重要因素,直接影响分类结果的准确性。固定阈值难以适应不同标签的特征,而动态阈值可根据数据集的标签分布和模型的预测结果进行动态调整,更好地适应不同标签的特性,提高分类准确性,将阈值设置在最大值和最小值之间,以更加灵活地适应不同情况。阈值公式如下:

$$T(\hat{y}) = \theta \max(\hat{y}) + (1 - \theta) \min(\hat{y}). \quad (8)$$

通过动态调整阈值,模型能够更精确地决定每个标签的分类边界,更好地捕捉类别之间的关联信息,从而有效地提高分类结果的准确性和泛化能力。

2.6 损失函数

在以往的原型网络研究中,通常使用欧氏

距离^[33]。与此不同,BERT的预训练任务不使用欧氏距离,而是依赖于无监督学习的目标,学习文本中的词汇和语境表示,生成高维度的向量表示^[40]。在应用中,这些向量采用余弦相似度等度量方式来进行比较更合适,因为余弦相似度能够更好地捕捉文本语义的相似性,尤其在高维空间中更为稳健。因此,计算一个查询向量 ct 与每个原型向量 c^i 的相似度,并将其归一化,作为最终的预测结果 $\hat{y}$ 。

$$\hat{y} = \text{softmax}\left(\frac{\cos(c^i, ct) - \mu}{\sigma}\right), \quad (9)$$

μ 是余弦相似度的均值, σ 是余弦相似度的方差, $i, j \in N$ 。给定原型向量集合 $C = \{c^1, \dots, c^n\}$ 和查询句子 $x \in Q$, 通过对 x 和每个原型向量进行注意力操作,得到 $Ct = \{ct^1, \dots, ct^n\}$ 作为去噪后的句子表示。将 Ct 与 C 之间的相似度得分进行优化,作为训练目标。对于这个目标,同一类别中查询句子 ct^n 和原型向量 c^n 之间的相似度得分应该高于 ct^n 和其他类别原型向量之间的相似度得分。

$$L = -\frac{1}{n} \sum [y \ln \hat{y} + (1 - y) \ln (1 - \hat{y})]. \quad (10)$$

这一目标使得每个 ct^n 向其所属类别的原型向量靠近。原型损失函数计算每个查询向量与其对应的类别原型向量之间的余弦相似度,并使用 softmax 函数将这些相似度转换为概率分布。该损失函数的训练目标是最小化交叉熵损失和原型之间的余弦相似度距离,以使具有相同标签的句子在嵌入空间中更紧密地聚集在一起。

3 实验与分析

本文对 ProtPrompt 模型进行了实验分析,展示了实验细节,用相应指标评估了模型的性能,并且对结果进行分析。

3.1 实验数据集

为了验证 ProtPrompt 模型的有效性,本文采用了三个公开的数据集^[28]进行实验测试:分别为 FewAsp (single), FewAsp (multi) 和 FewAsp:这三个数据集从 Yelp 方面^[41]构建了三个小样本方面类别检测数据集,各自拥有相同 100 个方面类别,其中 64 个方面类别用于训练,

16 个方面类别用于验证,20 个方面类别用于测试。

其中 FewAsp (single), FewAsp (multi) 和 FewAsp 三个数据集分别是仅一个方面类别、多个方面类别、混合(一个或者多个)类别类型的句子构成。该数据集具体分布情况如表 1。该数据集样例如表 2 所示,其中 #cls. 表示类别数量, #sent. 表示句子数量。

表 1 FewAsp(single), FewAsp(multi) 和 FewAsp 三个数据集的构成

Table 1 Composition of FewAsp(single), FewAsp(multi) and FewAsp datasets

Dataset	#cls.	#sent./cls.	#sent.
FewAsp(single)	100	200	20 000
FewAsp(multi)	100	400	40 000
FewAsp	100	630	63 000

注: #cls. 表示“类别数量(Number of classes)”; #sent. 表示“每类句子数(Number of sentences per class)”。# 通常用来表示“数量”或“数目”。

3.2 实验设置

3.2.1 评价指标

此任务可视为多标签分类任务,以 AUC (Area Under Curve) 和 macro F1 作为评价指标。

3.2.2 训练细节

针对小样本数据集的情况下,构建元任务有支持集和查询集,每个元任务从数据集中随机抽样类别和样本数为 5/10 作为支持集,同时随机抽取其类别中 5 个与支持集没有交集的样本作为查询集,支持集用于构建每个类别的原型向量,而查询集则用于测试模型的泛化能力。

3.2.3 参数设置

本研究中的实验采用 PyTorch 框架^[42]实现所有的模型和基线。选用 BERT 作为特征提取器。为了得到精确的原型向量,在模型中设置向量维度为 768,并使用 Adam 优化器^[43]来最小化损失函数。进行了 30 个 epochs 的训练,学习率设置为 1×10^{-6} ,并限制了最大序列长度为 100,以适应数据集。在训练过程中,使用了 sigmoid 激活函数和二元交叉熵损失函数来训练模型,并采用批量梯度下降算法进行优化。实验设置旨在提高模型的性能和泛化能力,以适应多标签分类任务的需求。

3.3 结果与分析

3.3.1 对比实验

本文研究方法与其他基线模型进行了比较,在 FewAsp (single), FewAsp (multi) 和 FewAsp 三个数据集上,采用了 5/10-way 和 5/10-shot 的实验条件进行了对比。在每个实验条件下,对测试集进行了 600 次测试,并将这些测试结果取平均值作为实验结果。实验结果汇总在表 3、表 4 和表 5 中,每个指标的最佳分数用粗体标示。实验结果表明,本文提出的方法在不同的数据集和实验条件下都表现出了有效性,并取得了优于其他方法的结果。对比模型如下:

Matching Network^[10]: 该网络结合了度量学习和记忆增强神经网络,并利用注意力和记忆机制来加速学习。

Prototypical Network^[11]: 通过学习浅层空间的向量表示的原型,可以在未见过的类别上进行分类。只需将新样本在 CNN 中向前传播一

表 2 3-way2-shot 示例:不同类别(A—C)之间可能存在交集

Table 2 3-way2-shot Example: there may be intersections between different categories (A—C)

支持集	
方面类别	样本
(A) room_cleanliness	(1) had to move because the first room was dirty. (2) people have mentioned, bed bugs on yelp !!
(B) service_front_desk	(1) breakfast is very unpleasant and no one provides any service. (2) desk and wheeling desk chair.
(C) staff_owner	(1) it was my first time tanning today and the owner completely took care of me! (2) I am a business owner and know customer service. this was not okay.
查询集	
(A)、(B) and (C)	(1) the carpet is dirty, the guest desk chairs are so filthy, I didn't even call the waiter and nobody paid attention to me. (2) the owners are extremely smart and worldly.

遍,计算与原型的距离,即可分配到最近的类别。

Relation Network^[12]:通过在浅空间中学习原型的向量表示,可以对看不见的类别进行分类。只需通过CNN向前传播新样本,计算与原型的距离,分配到最接近的类别。

Graph Network^[42]:提出小样本学习作为监督消息传递任务,它使用图神经网络提供更好的数值性能和端到端训练。

无限混合原型(Infinite Mixture Prototypes, IMP)^[44]:是一种适应性地通过每个类别的变化聚类数来拟合数据分布的小样本学习模型。

Proto-HATT(Hybrid Attention)^[45]:提出一种学习新概念的小样本学习框架,结合度量学习和神经网络外部记忆,并利用从标记和未标记数据中学习的分层注意力传播机制。

Proto-AWATT(Aspect-Wise Attention)^[28]:提出了一种基于小样本多标签学习的原型网络用于情绪分析中的方面类别检测,利用有效的

注意机制减轻噪音。

标签驱动的去噪框架(Label-Driven Denoising Framework, LDF)^[30]:提出了将标签信息作为辅助信息加入训练过程中,利用标签指导提高原型的质量。

根据表3至表5中的数据,本文提出的ProtPrompt模型在所有数据集(FewAsp(single)、FewAsp(multi)和FewAsp)上表现出非常好的性能,每个条件下得分均超过了当前最优(SOTA)模型。该模型在三个公开数据集的F1值比SOTA模型分别提升了4.35%,8.62%,8.39%,特别是在表4中突出显示的FewAsp(multi)数据集上,模型的性能表现尤为突出。

在非常复杂的情况下,即FewAsp(multi)数据集10-way 10-shot条件下,ProtPrompt还取得了高达76.53%的F1分数,明显领先其他模型,例如次优的LDF模型仅有68.23%的F1分数。同时在模型表现方面,ProtPrompt的AUC值是95.94%,显著高于其他模型,较LDF提出的模

表3 FewAsp(single)多次实验结果在5/10-way 5/10-shot时的平均AUC(%)和F1(%)

Table 3 FewAsp (single) mean AUC (%) and F1 (%) of 5/10-way 5/10-shot results from multiple experiments

Models	5-way 5-shot		5-way 5-shot		5-way 5-shot		5-way 5-shot	
	AUC	F1	AUC	F1	AUC	F1	AUC	F1
Matching Network	97.05	81.89	97.49	84.62	96.30	70.95	96.72	73.28
Prototypical Network	96.49	83.30	97.53	86.29	95.97	74.23	96.71	76.83
Relation Network	93.31	75.79	90.86	72.02	91.81	63.78	90.54	61.15
Graph Network	96.54	81.45	97.46	85.04	95.45	70.75	96.97	77.84
IMP	96.65	83.69	97.47	86.14	96.00	73.80	96.91	77.09
Proto-HATT	96.45	83.33	97.62	86.71	95.71	73.42	97.00	77.65
Proto-AWATT	97.56	86.71	97.96	88.54	97.01	80.28	97.55	82.97
LDF	98.29	88.16	98.38	89.32	97.51	81.73	97.96	84.20
ProtPrompt (Ours)	98.80	93.23	99.03	94.15	98.70	88.95	99.03	90.58

表4 FewAsp(multi)多次实验结果在5/10-way 5/10-shot时的平均AUC(%)和F1(%)

Table 4 FewAsp (multi) mean AUC (%) and F1 (%) of 5/10-way 5/10-shot results from multiple experiments

Models	5-way 5-shot		5-way 5-shot		5-way 5-shot		5-way 5-shot	
	AUC	F1	AUC	F1	AUC	F1	AUC	F1
Matching Network	89.54	65.70	91.38	69.02	88.28	50.86	89.94	54.42
Prototypical Network	89.67	67.88	91.60	72.32	88.01	52.72	90.68	58.92
Relation Network	84.91	58.38	86.21	61.37	84.22	43.71	84.72	44.85
Graph Network	87.97	59.25	90.45	64.63	86.05	45.42	88.44	48.49
IMP	90.12	68.86	92.29	73.51	88.71	53.96	91.10	59.86
Proto-HATT	91.10	69.15	93.03	73.91	90.44	55.34	92.38	60.21
Proto-AWATT	91.45	71.72	93.89	77.19	89.80	58.89	92.34	66.76
LDF	92.67	73.38	93.89	78.81	90.87	62.06	92.93	68.23
ProtPrompt (Ours)	95.73	82.49	96.81	85.49	94.80	72.43	95.94	76.53

表 5 FewAsp 多次实验结果在 5/10-way 5/10-shot 时的平均 AUC(%) 和 F1(%)

Table 5 FewAsp mean AUC (%) and F1 (%) of 5/10-way 5/10-shot results from multiple experiments

Models	5-way 5-shot		5-way 5-shot		5-way 5-shot		5-way 5-shot	
	AUC	F1	AUC	F1	AUC	F1	AUC	F1
Matching Network	90.76	67.14	92.39	70.09	88.44	51.27	89.90	54.61
Prototypical Network	88.88	66.96	91.77	73.27	87.35	52.06	90.13	59.03
Relation Network	85.56	59.52	86.98	62.78	84.94	45.62	83.77	44.70
Graph Network	89.48	61.49	92.35	69.89	87.35	47.91	90.19	56.06
IMP	89.95	68.96	92.30	74.13	88.50	54.14	90.81	59.84
Proto-HATT	91.54	70.26	93.43	75.24	90.63	57.26	92.86	61.51
Proto-AWATT	93.35	75.37	95.28	80.16	92.06	65.65	93.42	69.70
LDF	94.65	78.27	95.71	81.87	92.74	67.13	94.29	71.97
ProtPrompt (Ours)	97.00	86.47	97.53	88.54	96.15	77.10	96.93	80.68

型的 AUC 仅为 92.93%。这些结果明确表明 ProtPrompt 在处理多标签任务方面的性能显著优于其他模型。

这些数据表明, ProtPrompt 在任务复杂性增加时仍保持高水平的性能。这归功于 ProtPrompt 模型与预训练任务的训练目标保持一致, 使得模型能够更好地利用预训练阶段学到的信息, 借助提示突出样本的重点, 增强了类别的辨别性。这一优势在处理多标签任务时尤为显著。总的来说, ProtPrompt 模型在所有指标上的优异表现充分证明了在处理多标签和复杂任务方面的卓越性能, 为未来的研究和应用提供了坚实的基础。

接下来, 本文将进一步讨论和分析 ProtPrompt 模型。为了方便观测结果, 在 FewAsp 数据集上挑战 10-way 5-shot 条件的实验。

3.3.2 消融实验

为了验证提示增强的原型的有效性, 在 10-way 5-shot 的条件下进行了消融实验, 比较不同消融条件的 AUC(%) 和 F1(%)。

表 6 FewAsp(multi) 的 AUC 与 Macro-F1 在 10-way 5-shot 条件下的消融研究比较

Table 6 Comparison of AUC and Macro-F1 for ablation study on the 10-way 5-shot scenario from FewAsp(multi)

Models	5-way 5-shot	
	AUC/%	F1/%
ProtPrompt (Ours)	96.15	77.10
Prompt (w/Last avg.)	94.59	73.29
Prompt (w/h _[CLS])	94.48	72.12
Denoising Model (w/o)	95.91	76.18
CosSim (w/ED)	93.11	66.28

在表格 6 中, 展示了在 10-way 5-shot 条件下, 本文提出的 ProtPrompt 模型的消融实验结果, 其 AUC 得分为 96.15%, F1 得分为 77.10%。对模型进行的消融实验表明, 使用平均句向量或 h_[CLS] 的隐藏层向量替换提示学习向量表示会导致得分降低至 AUC 94.59% 和 F1 73.29%, 以及 AUC 94.48% 和 F1 72.12%; 去除降低噪音的优化模型会导致得分降至 AUC 95.91% 和 F1 76.18%; 使用欧氏距离而非余弦相似度进行距离计算会导致得分降至 AUC 93.11% 和 F1 66.28%。在过去的原型网络研究中, 通常使用欧氏距离来计算向量之间的距离。欧氏距离并不符合预训练目标, 并且受到向量维度的影响较大。相比之下, 余弦相似度对维度变化不太敏感。因此基于 BERT 的预训练语言模型更倾向于使用余弦相似度来衡量向量之间的语义关系。这些结果明确地突显了 ProtPrompt 模型的每个组件在其高性能中起着关键作用。

为了更好地观察本文提出的方法对原型向量生成的影响, 在 5-way 5-shot 条件下从 FewAsp(multi) 测试集中随机采样了 50 次, 并使用 tSNE^[46] 对所得到的原型向量进行了可视化。图 2 展示了通过提示学习对句子嵌入空间的原型向量可视化结果, 其中具有相同颜色的原型(数据点)包含相同的方面信息。

本文提出的方法所得到的原型可视化如图 2 所示, 展示了与其他方法相比更清晰的分布。Class 代表的类别标签分别是 procedure_relax_spa、room_interior、food_food、experience_wait 和 sport_pool。这表明在使用提示学习的情况

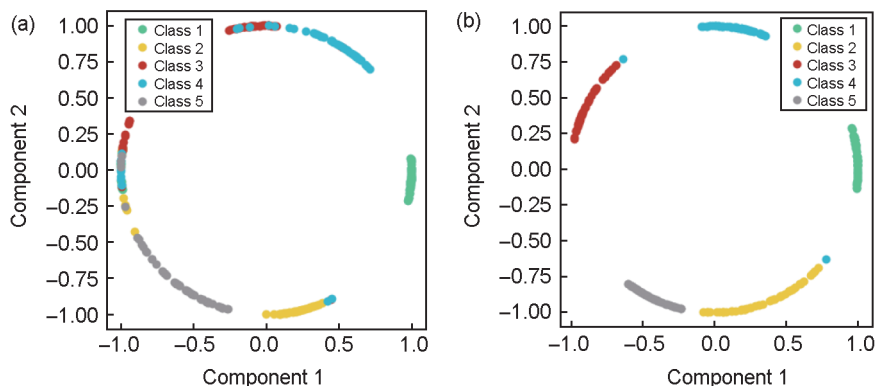


图2 原型嵌入可视化

(a)展示了未使用提示学习时的原型嵌入可视化,(b)展示了使用提示学习后的原型嵌入可视化

Fig. 2 Visualization of prototype embeddings

(a) the prototype embedding visualization without learning with hints, (b) the prototype embedding visualization after learning with hints

下,可以根据提示获得更有辨别性的向量,模型能够更好地分离和聚合句子向量。这种区分和聚合的能力使得模型能够生成高质量原型向量从而更容易地区分不同类别的原型向量。

3.3.3 提示模板实验

基于提示的任务的一个关键问题是找到合适的模板。在这方面,本文讨论了两种类型的模板:硬模板和软模板。硬模板将模板与句子连接起来,从而改变输入。而软模板则提供一个或者多个可学习的向量<prompt_n>,让模型自行学习。

在表7中,本文针对不同的提示模板对10-way 5-shot 条件的分类任务进行了评估,使用AUC和F1得分作为评估指标。实验结果显示,所有模板都取得了非常好的性能,其中表现最好的模板是“Category [MASK]: [x]”,其AUC为96.15%,F1为77.10%。此外,使用软提示的模型在性能方面也优于其他方法,其AUC为96.06%,F1为76.47%。通过分析表7中所列模板,使用模板的方法平均提升4.15% F1,这表明提示学习可以有效地解决小样本多标签分类问题,并且选择合适的提示模板对于不同任务和应用场景可以提高模型性能。

4 结论

本研究针对多标签方面类别任务的小样本学习挑战,提出了一种基于提示增强的原型网络模型(ProtPrompt)。该模型充分利用了预训练语言模型的知识 and 提示学习的机制,以提升

表7 多种硬模板和软模板的AUC与Macro-F1在10-way 5-shot 条件下的实验比较

Table 7 Experimental comparison of AUC of multiple hard and soft templates with Macro-F1 under 10-way 5-shot conditions

Template	5-way 5-shot	
	AUC/%	F1/%
The aspects of [x] are [MASK].	95.85	75.86
What are the aspects of [x] [MASK].	95.88	76.19
[x] are related to [MASK].	96.01	76.28
[x] can be described as [MASK].	96.04	76.45
Category [MASK]: [x].	96.15	77.10
This sentence [x] means [MASK].	95.88	76.15
This [x] means [MASK].	95.90	75.72
Soft Template	96.06	76.47

句子向量表示的辨别性。通过将下游任务转化为完形填空任务,模型能够更好地学习与提示信息相关的特征。借助多维度映射与注意力机制,支持集和查询向量的表示得以优化,进一步提升了模型性能。在方面类别判定方面,采用余弦相似度计算代替传统的欧氏距离度量,从而有效提高了模型对分类特征分布的敏感性等问题,显著提升了模型的性能。

实验结果表明,本文提出的基于提示增强的原型网络模型在三个公开数据集上相较于当前最先进的基线模型显著提升性能,每个条件下得分均超过了SOTA模型。在三个数据集上平均提升了7.626%的F1值。提示增强的原型网络通过余弦相似度损失计算,合理权衡样本和类别信息的重要性,减少了高维向量空间的影响,且在不同提示模板下均有效提高了分类

性能。有效解决了小样本多标签方面类别检测任务的挑战,并具备较强的可迁移性和更好的语义关系捕捉能力。

在未来的工作中,计划探索更多的提示学习方法,以提高模型的性能和泛化能力。此外,将扩展模型到零样本学习,并在更多的评论数据集上验证其性能。

参考文献:

- [1] PONTIKI M, GALANIS D, PAVLOPOULOS J, *et al.* SemEval-2014 Task 4: Aspect Based Sentiment Analysis [C]//Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). Dublin, Ireland: Association for Computational Linguistics, 2014:27-35.
- [2] SHI T, LI L Q, WANG P, *et al.* A Simple and Effective Self-supervised Contrastive Learning Framework for Aspect Detection[EB/OL]. arXiv Preprint: 2009.09107, 2020. <http://arxiv.org/abs/2009.09107>.
- [3] KARN A L, KARNA R K, KONDAMUDI B R, *et al.* Customer Centric Hybrid Recommendation System for E-commerce Applications by Integrating Hybrid Sentiment Analysis[J]. *Electron Commer Res*, 2023, **23**(1): 279-314. DOI: 10.1007/s10660-022-09630-z.
- [4] MOVAHEDI S, GHADERY E, FAILI H, *et al.* Aspect Category Detection via Topic-attention Network[EB/OL]. arXiv Preprint: 1901.01183, 2019. <http://arxiv.org/abs/1901.01183>.
- [5] CHEBOLU S U S, ROSSO P, KAR S, *et al.* Survey on Aspect Category Detection[J]. *ACM Comput Surv*, 2023, **55**(7): 1-37. DOI: 10.1145/3544557.
- [6] WANG Y Q, YAO Q M, KWOK J T, *et al.* Generalizing from a few Examples: A Survey on Few-shot Learning [J]. *ACM Comput Surv*, 2020, **53**(3): 63. DOI: 10.1145/3386252.
- [7] VANSCHOREN J. Meta-learning: a Survey[EB/OL]. arXiv Preprint: 1810.03548, 2018. <http://arxiv.org/abs/1810.03548>.
- [8] FINN C, ABBEEL P, LEVINE S. Model-agnostic Meta-learning for Fast Adaptation of Deep Networks[EB/OL]. arXiv Preprint: 1703.03400, 2017. <http://arxiv.org/abs/1703.03400.pdf>.
- [9] CHICCO D. Siamese Neural Networks: an Overview[J]. *Methods Mol Biol*, 2021, **2190**: 73-94. DOI: 10.1007/978-1-0716-0826-5_3.
- [10] VINYALS O, BLUNDELL C, LILLICRAP T, *et al.* Matching Networks for one Shot Learning[C]//Proceedings of the 30th International Conference on Neural Information Processing Systems. New York: ACM, 2016: 3637-3645. DOI: 10.5555/3157382.3157504.
- [11] REN M Y, LIAO R J, FETAYA E, *et al.* Incremental Few-shot Learning with Attention Attractor Networks [EB/OL]. arXiv Preprint: 1810.07218, 2018. <http://arxiv.org/abs/1810.07218>.
- [12] HU H, GU J Y, ZHANG Z, *et al.* Relation Networks for Object Detection[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 3588-3597. DOI: 10.1109/CVPR.2018.00378.
- [13] JIANG T, HUANG S H, ZHANG Z H, *et al.* Promptbert: Improving Bert Sentence Embeddings with Prompts[C]//Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. Abu Dhabi: Association for Computational Linguistics, 2022: 8826-8837. DOI: 10.18653/v1/2022.emnlp-main.603.
- [14] SUN C, QIU X P, XU Y G, *et al.* How to Fine-tune BERT for Text Classification?[C]//China National Conference on Chinese Computational Linguistics. Cham: Springer, 2019: 194-206. DOI: 10.1007/978-3-030-32381-3_16.
- [15] LIU X, ZHENG Y N, DU Z X, *et al.* GPT Understands, Too[C]. *AI Open*, 2021: 232269696. DOI: 10.1016/j.aiopen.2023.08.012.
- [16] 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 Comput Surv*, **55**(9): 195. DOI: 10.1145/3560815.
- [17] SU J L, CAO J R, LIU W J, *et al.* Whitening Sentence Representations for Better Semantics and Faster Retrieval[EB/OL]. arXiv Preprint: 2103.15316, 2021. <http://arxiv.org/abs/2103.15316>.
- [18] HAI Z, CHANG K Y, KIM J J. Implicit Feature Identification via Co-occurrence Association Rule Mining[C]//Proceedings of the 12th International Conference on Computational Linguistics and Intelligent Text Processing, 2011: 393-404. DOI: 10.5555/1964799.1964836.
- [19] SU Q, XIANG K, WANG H F, *et al.* Using Pointwise Mutual Information to Identify Implicit Features in Customer Reviews[C]//International Conference on Computer Processing of Oriental Languages. 2006, **4285**: 22-30. DOI: 10.1007/11940098_3.
- [20] SCHOUTEN K, VAN DER WEIJDE O, FRASINCAR F, *et al.* Supervised and Unsupervised Aspect Category Detection for Sentiment Analysis with Co-occurrence Data[C]//IEEE Transactions on Cybernetics. IEEE, 2018: 1263-1275. DOI: 10.1109/TCYB.2017.2688801.
- [21] KIRITCHENKO S, ZHU X D, CHERRY C, *et al.* Nre-

- canada-2014: Detecting Aspects and Sentiment in Customer Reviews[C]//International Workshop on Semantic Evaluation. Dublin: Association for Computational Linguistics, 2014: 437–442. DOI: 10.3115/v1/S14-2076.
- [22] JIANG Q N, CHEN L, XU R F, *et al.* A Challenge Dataset and Effective Models for Aspectbased Sentiment Analysis[C]//Empirical Methods in Natural Language Processing International Joint Conference on Natural Language Processing, 2019: 6280–6285. DOI: 10.18653/v1/D19-1654.
- [23] HOU Y T, LAI Y K, WU Y S, *et al.* Few-shot Learning for Multi-label Intent Detection[EB/OL]. arXiv Preprint: 2010.05256, 2020. <http://arxiv.org/abs/2010.05256>.
- [24] HU MENGTING, ZHAO SHIWAN, ZHANG LI, *et al.* CAN: Constrained Attention Networks for Multi-aspect Sentiment Analysis[C]//Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, 2019, 1: 4601–4610. DOI: 10.18653/v1/D19-1467.
- [25] XUE W, ZHOU W B, LI T, *et al.* MTNA: A Neural Multi-task Model for Aspect Category Classification and Aspect Term Extraction on Restaurant Reviews[C]//Eighth International Joint Conference on Natural Language Processing, 2017, 2: 151–156. DOI: 10.1007/978-3-031-15931-2_23.
- [26] REN M Y, TRIANTAFILLOU E, RAVI S, *et al.* Meta-learning for Semi-supervised Few-shot Classification [EB/OL]. arXiv Preprint: 1803.00676, 2018. <http://arxiv.org/abs/1803.00676>.
- [27] LI F F, FERGUS R, PERONA P. One-shot Learning of Object Categories[C]//IEEE Transactions on Pattern Analysis and Machine Intelligence. Washington: IEEE, 2006: 594–611. DOI: 10.1109/TPAMI.2006.79.
- [28] HU MENGTING, ZHAO SHIWAN, GUO HONGLEI, *et al.* Multi-label Few-shot Learning for Aspect Category Detection[C]//Association for Computational Linguistics International Joint Conference on Natural Language Processing. 2021, 1: 6330–6340. DOI: 10.18653/v1/2021.acl-long.495.
- [29] LEE C W, FANG W, YEH C K, *et al.* Multi-label Zero-shot Learning with Structured Knowledge Graphs[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 1576–1585. DOI: 10.1109/CVPR.2018.00170.
- [30] ZHAO F, SHEN Y C, WU Z, *et al.* Label-Driven Denoising Framework for Multi-Label Few-Shot Aspect Category Detection [C]//Association for Computational Linguistics Conference on Empirical Methods in Natural Language Processing, 2022: 2390–2402. DOI: 10.18653/v1/2022.findings-emnlp.177.
- [31] HU S D, DING N, WANG H D, *et al.* Knowledgeable Prompt-tuning: Incorporating Knowledge into Prompt Verbalizer for Text Classification[EB/OL]. arXiv Preprint: 2108.02035, 2021. <http://arxiv.org/abs/2108.02035>.
- [32] WANG Y Q, YAO Q M, KWOK J T, *et al.* Survey on Few-shot Learning[C]//Association for Computing Machinery Computing Surveys, 2020, 53: 349–369. DOI: 10.1145/3386252.
- [33] SNELL J, SWERSKY K, ZEMEL R S. Prototypical Networks for Few-shot Learning[EB/OL]. arXiv Preprint: 1703.05175, 2017. <https://doi.org/10.48550/arXiv.1703.05175>.
- [34] ALFASSY A, KARLINSKY L, AIDES A, *et al.* LaSO: Label-set Operations Networks for Multi-label Few-shot Learning[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019: 6541–6550. DOI: 10.1109/CVPR.2019.00671.
- [35] CHENG K H, CHOU S Y, YANG Y H. Multi-label Few-shot Learning for Sound Event Recognition[C]//2019 IEEE 21st International Workshop on Multimedia Signal Processing (MMSp). Kuala Lumpur: IEEE, 2019: 1–5. DOI: 10.1109/MMSp.2019.8901732.
- [36] JILL BURSTEIN, CHRISTY DORAN, THAMAR SOLORIO, *et al.* Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding[C]//Association for Computational Linguistics Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis: Association for Computational Linguistics. 2018: 4171–4186. DOI: 10.18653/v1/N19-1423.
- [37] LIU Z, LIN W, SHI Y, *et al.* A Robustly Optimized BERT Pre-training Approach with Post-training[C]//China National Conference on Chinese Computational Linguistics. Cham: Springer. 2021: 471–484. DOI: 10.1007/978-3-030-84186-7_31.
- [38] BROWN T B, BENJAMIN M, RYDER N, *et al.* Language Models are Few-shot Learners[C]//Advances in Neural Information Processing Systems, 2020: 1877–1901. DOI: 10.48550/arXiv.2005.14165.
- [39] GU Y X, HAN X, LIU Z Y, *et al.* PPT: Pre-trained Prompt Tuning for Few-shot Learning[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg, PA, USA: Association for Computational Linguistics. 2022: 8410–8423. DOI: 10.18653/v1/2022.acl-long.576.
- [40] LI B, ZHOU H, HE J X, *et al.* On the Sentence Embed-

- dings from Bert for Semantic Textual Similarity[C]// Conference on Empirical Methods in Natural Language Processing. 2020: 9119–9130. DOI: 10.18653/v1/2020.emnlp-main.733.
- [41] GARCIA V, BRUNA J. Few-shot learning with graph neural networks[EB/OL]. arXiv Preprint: 1711.04043, 2018. <https://doi.org/10.48550/arXiv.1711.04043>.
- [42] PASZKE A, GROSS S, MASSA F, *et al.* PyTorch: an Imperative Style, High-performance Deep Learning Library[EB/OL]. arXiv Preprint: 1912.01703, 2019. <http://arxiv.org/abs/1912.01703>.
- [43] KINGMA D P, BA J. Adam: A Method for Stochastic Optimization[EB/OL]. arXiv preprint: 1412.6980, 2014. <http://arxiv.org/abs/1412.6980.pdf>.
- [44] ALLEN K R, SHELHAMER E, SHIN H, *et al.* Infinite Mixture Prototypes for Few-shot Learning[EB/OL]. arXiv preprint: 1902.04552, 2019. <http://arxiv.org/abs/1902.04552>.
- [45] GAO T Y, HAN X, LIU Z Y, *et al.* Hybrid Attention-based Prototypical Networks for Noisy Few-shot Relation Classification[J]. *Proc AAAI Conf Artif Intell*, 2019, **33**(1): 6407–6414. DOI: 10.1609/aaai.v33i01.33016407.
- [46] VAN DER MAATEN L, HINTON G. Visualizing Data Using t-SNE[J]. *J Mach Learn Res*, 2008, **9**(86): 2579–2605.