基于主题提示学习的零样本立场检测方法

陈子潇, 梁斌, 徐睿峰*
(哈尔滨工业大学（深圳）计算机科学与技术学院, 广东省 深圳市 518071
{chenzixiao, bin.liang}@stu.hit.edu.cn
xuruifeng@hit.edu.cn)

摘要

零样本立场检测目的是针对未知目标数据进行立场极性预测。一般而言，文本的立场表达是与所讨论的目标主题是紧密联系的。针对未知目标的立场检测，本文将立场表达划分为两种类型：一类在说话者面向不同的主题和讨论目标时表达相同的立场态度，称之为目标无关的表达；另一类在说话者面向特定主题和讨论目标时才表达相应的立场态度，本文称之为目标依赖的表达。对这两种表达进行区分，有效学习到目标无关的表达方式并忽略目标依赖的表达方式，有望强化模型的可迁移能力，使其更加适应零样本立场检测任务。据此，本文提出了一种基于主题提示学习的零样本立场检测方法。具体而言，受自监督学习的启发，本文为了零样本立场检测设置了一个代理任务框架。其中，代理任务通过掩盖上下文中的目标主题词生成辅助样本，并基于提示学习分别预测原样本和辅助样本的立场表达，随后判断原样本和辅助样本的立场表达是否一致，从而在无需人工标注的情况下判断样本的立场表达是否依赖于目标的代理标签。然后，将此代理标签提供给立场检测模型，对应学习可迁移的立场检测特征。在两个基准数据集上的大量实验表明，本文提出的方法在零样本立场检测任务中相比基线模型取得了更优的性能。

关键词： 零样本立场检测 ； 提示学习 ； 代理任务

A Topic-based Prompt Learning Method for Zero-Shot Stance Detection

Zixiao Chen, Bin Liang, Ruifeng Xu*
(School of Computer Science and Technology,
Harbin Institude of Technology (Shenzhen), Shenzhen,Guangdong 518071,China
{chenzixiao, bin.liang}@stu.hit.edu.cn
xuruifeng@hit.edu.cn)

Abstract

Zero-shot stance detection (ZSSD) aims to detecting the stance of previously unseen targets during the inference stage. It is generally believed that the stance expression in a sentence is closely related to the stance target and topics discussed.We divide stance expressions of speakers into two categories: target-invariant and target-specific categories. Target-invariant stance expressions carry the same stance polarity regardless of the targets they are associated with. On the contrary, target-specific stance expressions only co-occur with certain targets. As such, it is important to distinguish these two types of stance features to boost stance detection ability.In this paper,we develop an effective approach to distinguish the types of target-related stance expressions to better

* 通讯作者

learn transferable stance features.To be specific,inspired by self-supervised learning,we frame the stance-feature-type identification as a pretext task in ZSSD.We apply prompt learning to predict changing relationship between stance polarity labels and topic information in pretext task.This essentially allows the model to learn transferable stance features. Extensive experiments on two benchmark datasets show that the proposed method obtains an improved performance than the baseline in ZSSD.

Keywords: Zero-shot Stance Detection , Prompt Learning , Pretext Task

1 引言

立场检测(Stance Detection)是自然语言处理(Natural Language Processing)领域(Kachuee et al., 2021)的一个重要任务。立场检测的目的是识别说话者面向特定目标、主题和主张时表达的立场与态度(Somasundaran et al., 2010; Augenstein et al., 2016; Mohammad et al., 2016)。在以往研究聚焦的目标集合内部的立场检测任务中(Gunel et al., 2020)，训练集和测试集共享可见的目标集合。然而在现实生活中，存在大量目标相对于现有立场检测模型是未知的样例(Allaway et al., 2020)。为了解决这个问题进而出现了零样本立场检测(Zero-shot Stance Detection)任务，旨在面向未知目标进行立场检测。

现有方法引进了注意力机制(Allaway et al., 2020)和额外知识在已知目标和未知目标间捕捉可迁移立场特征。但在实践中，这些方法的捕捉能力不强，对外部能力依赖大，缺少对数据集本身信息的充分利用。一般认为，一句话的立场表达是与所讨论的目标和主题紧密联系的，本文将说话者的立场表达划分为两种类型：一类在说话者面向不同的主题和讨论目标时表达相同的立场态度，本文称之为目标无关的表达；另一类在说话者面向特定主题和讨论目标时才表达相应的立场态度，本文称之为目标依赖的表达。在立场检测任务中，区分说话者表达的立场是目标依赖还是目标无关是十分重要的。前者随着面向目标的改变有可能改变立场，而后者则不会。这一思路仅从数据本身挖掘信息，能够有效加强模型的可迁移学习能力，且无需依赖外部知识或特殊网络结构。为了说明本文提出的方法，表1给出了一些目标依赖和目标无关的立场表达的样例。在目标及其相关主题词被隐藏后，对于目标无关的立场表达，样例的立场态度并未发生改变；然而对于目标依赖的立场表达，失去与目标间的联系后立场态度发生了改变。因此，在零样本立场检测任务中，当寻找可迁移特征时，模型需要强化对目标无关的立场表达特征的学习并排除目标相关的立场表达特征的干扰。

目标无关的立场表达
目标：Feminist Movement 立场极性：Against 样例句：feminist only want the same benefiting right as men not those harmful ones 隐藏目标和主题词后的句子：$[MASK]$ only want the same benefiting $[MASK]$ as $[MASK]$ not those harmful ones 隐藏信息后样例句的立场表达：Against
目标依赖的立场表达
目标：Donald Trump 立场极性：Against 样例句：white terrorism is alive and well 隐藏目标和主题词后的句子：$[MASK]$ is alive and well 隐藏信息后样例句的立场表达：Favor

Table 1: 立场检测中立场表达类型样例

本文的主要贡献概括如下：

- 本文基于提示学习，从一种新的角度探讨了零样本立场检测问题。该方法通过主题词的保留和掩盖，以提示学习来自动学习立场表达是否依赖目标，进而将目标无关的立场表达特征用于未知目标的立场检测任务。
- 本文提出了一种创新的特征生成方法。该方法通掩盖训练样本的目标词及其相关主题词来生成辅助训练样本，并使用提示学习，通过自监督学习的方式利用预训练语言模型的先验知识判断立场表达特征与目标的关联性。
- 在两个公开数据集上的实验结果表明，本文提出的方法在零样本立场检测任务中取得了比基线模型更优的性能。

2 相关工作

早期对零样本立场检测方法的研究多集中在目标集合内部的立场检测，也就是训练集和测试集共享目标的检测任务(Du et al., 2017; Sun et al., 2018; Li et al., 2019; Siddiqua et al., 2019; Kawintiranon et al., 2021)。跨目标的立场检测是一种和零样本立场检测相似的任务，该任务基于一个已知目标训练分类器对一个未知目标的数据进行立场预测(Xu et al., 2018; Wei et al., 2019; Zhang et al., 2020; Liang et al., 2021)。现存跨目标的立场检测研究通常使用了基于注意力机制(Xu et al., 2018; Wei et al., 2019)或图网络(Zhang et al., 2020; Liang et al., 2021)的模型，根据训练集的目标学习目标关联特征，然后用于与目标数据集相近的测试集的预测。不同于跨目标的立场检测任务，零样本立场检测希望能够自动判断各种未知目标数据的立场结果。在这一任务要求下，Conforti等(Conforti et al., 2020)搭建了一个大范围专家标注的立场检测数据集，其中测试集的目标相对于训练集是不可见的。Allaway等(Allaway et al., 2020)搭建了一个零样本立场检测数据集，该数据集拥有大量的主题，相关话题类别十分广阔。Allaway等(Allaway et al., 2020)还在该数据集的基础上提出了一个主题分组的注意力模型来捕捉目标和通用主题表示间的关系。在另一项研究中，Allway等(Allaway et al., 2021)将一个用于目标内部立场检测的数据集(Mohammad et al., 2016)应用到零样本立场检测中，并使用了对抗学习来提取样本无关的可迁移特征。此外，Liu等(Liu et al., 2021)同时从结构层面和语义层面引入相关的外部知识，提出了一种基于BERT(Devlin et al., 2019)的常识增强图模型来解决零样本立场检测任务。

在特征学习的监督信号是从数据自动生成的情况下，自监督学习有着良好的表现。近年来自监督学习有一种发展方向就来自自动设计的预测任务，或者常被称作代理任务(Zhang et al., 2016; Gidaris et al., 2018; Chen et al., 2020)。很多现有的计算机视觉研究领域的方法，在包括拼图问题(Noroozi et al., 2016)、旋转预测(Gidaris et al., 2018)等任务上都设计了启发式的无注释代理任务，以便为目标问题提供特征学习的替代监督信号(Zhang et al., 2016; Gidaris et al., 2018; Chen et al., 2020; Larsson et al., 2016; Simard et al., 2021)。由此，受现有的自监督方法的启发，本文设计了一个代理任务来挖掘立场表达是否关联目标这一重要特征并用于可迁移学习任务。

通常定义下，提示学习是一种通过在文本输入部分增加提示信息，将下游学习任务转化为文本生成任务的一种学习方法。Fabio等(Petroni et al., 2019)提出了LAMA，一种将关系抽取任务转化为填空任务的提示学习方法，并取得了比基于外部知识的方法更好的性能表现。Shin等(Shin et al., 2020)将提示学习方法应用到文本分类和文本蕴含识别任务上，在没有改变预训练语言模型的情况下完成了这些问题的良好预测。Timo等(Schick et al., 2021)提出了PET，一种通用的半监督训练模式，适用于一系列自然语言处理问题。PET尤其在自动标注数据和扩充训练集的任务上取得了优异表现。受这些方法的启发，本文设置了一个运用提示学习方法自动标注文本立场极性的代理任务，以有效判断立场主题信息改变是否会同样改变立场表达。

3 模型方法

本章将详细描述本文提出的一种基于主题提示学习的零样本立场检测方法（Data Augmentation for **S**tance **T**opic features via **P**rompt **L**earning Pretext Task，ST-PL），其总体结构如图1所示。本方法的核心是在模型训练阶段之前设置一个代理任务，该代理任务应用提示学习在自监督立场主题特征这一维度上进行了数据增强，以帮助训练阶段更好地学习可迁移的立场特

征。本方法主要分为四步：1）对每一个目标使用主题模型挖掘需要隐藏的相应目标和主题信息；2）使用训练后的提示学习模型对隐藏主题信息后的训练集数据判断立场极性，判断结果与原立场极性不同的数据则认为包含目标依赖的特征；3）通过移除训练集数据里包含目标依赖特征的数据进行数据增强；4）在增强后数据上训练立场检测主任务模型，以加强模型对零样本立场检测主任务的预测能力。经过本文提出的ST-PL方法处理后，后续零样本立场检测模型的训练和预测阶段的表现得到了提升。

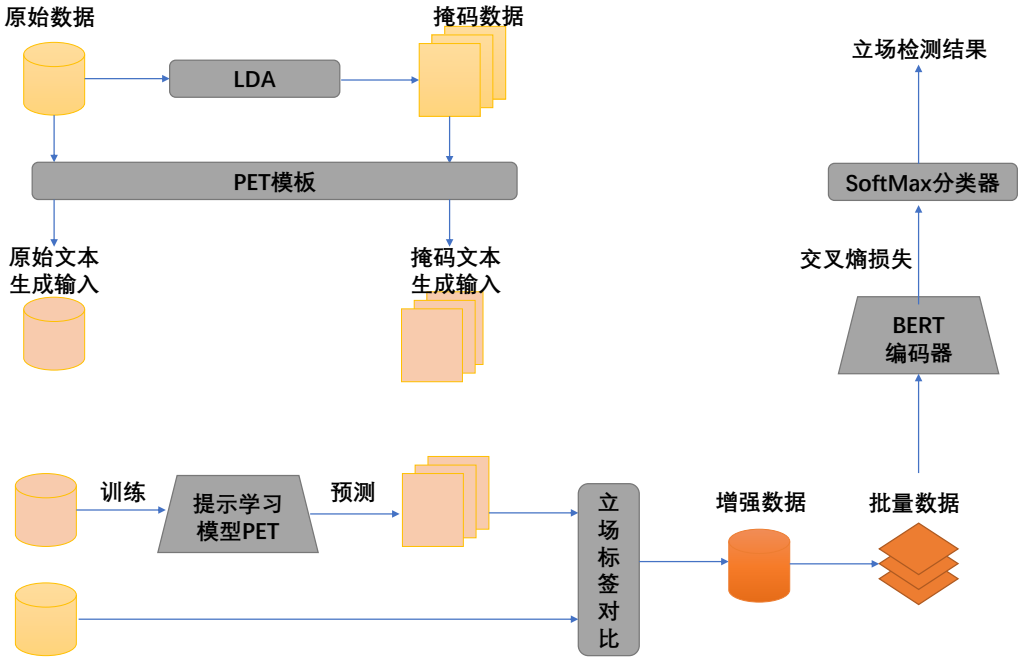

Figure 1: ST-PL方法的总体架构

3.1 任务定义

为不失一般性，假设有一个面向已知目标的已标注实例集合$\mathcal{D}_s = \{(r_s^i, t_s^i, y_s^i)\}_{i=1}^{N_s}$ 和一个面向未知目标的未标注实例集合$\mathcal{D}_d = \{(r_d^i, t_d^i)\}_{i=1}^{N_d}$。其中$y_s^i$ 是一个面向已知目标t_s^i的已标注实例的立场标签。N_s 和N_d 是已知目标数据集和未知目标数据集的数据量。$\mathcal{D}_s$ 和$\mathcal{D}_d$间没有重合的立场目标。零样本立场检测的目标是对每个来自数据集$\mathcal{D}_s$ 的面向已知目标t_s^i 的句子r_s^i，训练一个模型能够在来自数据集$\mathcal{D}_d$ 的面向未知目标t_d^i的句子r_d^i上具有泛用的预测能力。

3.2 基于代理任务的数据增强方法

一句话的立场表达是与所讨论的目标和主题紧密联系的。一部分立场表达是较为通用的，可以出现在各种所讨论主题不同的场合而不改变其立场态度，而另一部分则大多局限于某些特定目标和议题。因此本文将立场表达划分为目标无关的和目标依赖的两类。受现有方法(Zhang et al., 2016; Gidaris et al., 2018; Chen et al., 2020; Simard et al., 2021; Schick et al., 2021)启发，为了区分这两种立场特征类型以更好地学习到可用于在零样本立场检测任务中进行迁移学习的立场数据，本文探索了一个结合自监督代理任务与主题提示学习的数据增强策略，用于为下游任务学习提供新的监督信号。

3.3 基于主题提示学习的代理任务框架

对于每一个由句子r_s^i和目标t_s^i组成的实例，本文使用提示学习方法PET(Schick et al., 2021)训练立场检测模型$\mathcal{M}$。PET通过预训练语言模型直接预测被转化为文本生成问题的立场检测任务。这里PET使用的训练模板如表3所示。随后，本文对于每一个目标所对应的训练实例，隐藏这些实例中的目标词和相关主题词，由此得到了隐去部分信息的候选实例。这种候选实例生成方式的目的是运用后续训练的模型学习得到立场表达和目标是否具有关联性。在这

里，提取目标相关主题词所用的是LDA（Latent Dirichlet Allocation），一种基于隐含狄利克雷分布的主题模型。相关样例由表2所示。然后，使用训练后的$\mathcal{M}$预测被隐去目标信息和相关主题词的句子所组成的新训练集的立场极性，记录预测立场极性与未隐去信息的原数据的立场极性的异同。此处使用提示学习这一并非直接适用于立场检测的文本分类任务的方法，其目的在于考虑到提示学习使用文本提示信息帮助模型学习的特性，可以通过对主题词信息的提示，有效挖掘样例中的主题词和其他词的关联，能够捕捉隐藏信息前后对句子带来的变化，以提高本文设置的代理任务场景对立场目标相关表达的学习效果。

目标	主题词
Donald Trump	right,vote,obama,president,america
Hillary Cliton	women,president,right,campaign,wakeupamerica

Table 2: 主题词样例

目标无关的立场表达
目标：Feminist Movement 立场极性-生成用词：Against-No 样例句：Feminist only want the same benefiting right as men not those harmful ones. 提示学习模板：[目标]，[生成用词]。[样例句]。 模板生成句：Feminist Movement , [?]. Feminist only want the same benefiting right as men not those harmful ones. 模板预测结果：Feminist Movement , **No**. Feminist only want the same benefiting right as men not those harmful ones.

Table 3: 提示学习方法PET的训练模板

3.4 生成增强数据

为了区分目标依赖和目标无关的立场特征，以便更好地学习零样本立场检测中的可迁移立场特征，本文通过自监督学习设计的代理任务自动为训练数据生成辅助监督信号。首先，用一个特殊标记$[MASK]$替代被隐去的目标词和相关主题词，如表1所示。然后，将被隐去信息的句子输入PET模型$\mathcal{M}$以重新预测该实例的立场极性标签。此处参照表3样例，$\mathcal{M}$所预测的文本生成任务句形式为：$[MASK]$, [?]. $[MASK]$ only want the same benefiting right as $[MASK]$ not those harmful ones。[?]处为文本生成任务需要预测的生成用词，对应立场检测任务的立场极性。如果重新预测的标签正确，说明该立场表达不依赖于目标，该实例的立场表达是目标无关的，被标记为标签“*target-invariant*”；否则该实例的立场表达是目标依赖的，被标记为标签“*target-specific*”。训练集中目标依赖的数据将会被剔除以生成增强数据。数据增强后，训练集可以从形式上表示为$\mathcal{D}_s = \{(r_s^i, t_s^i, y_s^i, p_s^i)\}_{i=1}^{N_s}$。

3.5 训练架构

3.5.1 编码器模块

给定一个词语序列$r = \{w_i\}_{i=1}^{n}$ 和对应的目标t，n是文本r的长度。这里使用r和t来表示训练实例的句子和目标。在输入实例时，模型将忽略具有目标依赖标签的实例，以便在训练过程中偏好训练可迁移的立场特征。然后，本方法采用预训练的BERT (Devlin et al., 2019)作为编码器模块并将“ $[CLS]r[SEP]t[SEP]$”作为输入以获取每个输入样例的标记$[CLS]$ 的d_m维隐藏表示$\boldsymbol{h} \in \mathbb{R}^{d_m}$:

$$\boldsymbol{H} = \text{BERT}([CLS]r[SEP]t[SEP]), \boldsymbol{h} = \boldsymbol{H}_{[CLS]} \tag{1}$$

对于一个批量的数据用例，用例的隐藏表示可以定义为$\mathcal{B} = \{\boldsymbol{h}_i\}_{i=1}^{N_b}$，其中$N_b$ 是批量的大小。

数据集	目标	Favor	Against	Neutral	Unrelated
SEM16	DT	148	299	260	-
	HC	163	565	256	-
	FM	268	511	170	-
	LA	167	544	222	-
	A	124	464	145	-
	CC	335	26	203	-
WT-WT	CA	2469	518	5520	3115
	CE	773	253	947	554
	AC	970	1969	3098	5007
	AH	1038	1106	2804	2949

Table 4: SEM16和WT-WT数据集的数据分布

3.5.2 立场分类器

该模块将批量中数据的隐藏向量$\mathcal{B} = \{\boldsymbol{h}_i\}_{i=1}^{N_b}$输入到softmax函数的分类器中，以生成立场的预测分布：

$$\hat{\boldsymbol{y}}_i = \text{softmax}(\boldsymbol{W}\boldsymbol{h}_i + \boldsymbol{b}) \tag{2}$$

其中$\hat{\boldsymbol{y}}_i \in \mathbb{R}^{d_y}$是输入实例$x_i$的立场预测概率，$d_y$是立场标签的维度$\boldsymbol{W} \in \mathbb{R}^{d_y \times d_m}$和$\boldsymbol{b} \in \mathbb{R}^{d_y}$是可训练的参数。基于立场预测概率，我们采用实例$\boldsymbol{y}_i$的预测分布和真实分布$x_i$间的交叉熵损失来训练分类器：

$$\mathcal{L}_{class} = -\sum_{i=1}^{N_b}\sum_{j=1}^{d_p} y_i^j \log \hat{y}_i^j \tag{3}$$

3.6 模型训练与预测

在此阶段，本文提出的模型的学习目标是在增强训练集上优化立场检测的监督损失函数$\mathcal{L}_{class}$。总的损失函数$\mathcal{L}$通过将监督损失函数和正则项相加得到：

$$\mathcal{L} = \gamma_c \mathcal{L}_{class} + \lambda ||\Theta||^2 \tag{4}$$

其中γ_c是可调整的超参数，Θ表示模型所有可训练参数，λ表示L_2正则化的系数。

4 实验

4.1 实验数据集

本文在以下两个数据集上进行实验：

- SemEval2016(Mohammad et al., 2016)。该数据集包含在多领域的六个预定义的目标：Donald Trump(DT)，Hillary Clinton (HC)，Feminist Movement(FM)，Legalization of Abortion (LA)，Atheism (A)，Climate Change(CC)。参考Allaway等(Allaway et al., 2021)的做法，本文将其中一个目标的数据作为测试集，其余目标数据作为训练集，并随机选择训练集中15%的数据作为验证集来调整超参。数据集的详情在表4中显示。

- WT-WT(Conforti et al., 2020)。该数据集主要讨论公司之间的并购业务，包含4个目标：CVS_AET (CA)，CI_ESRX (CE)，ANTM_CI (AC)，AET_HUM (AH)。每个实例可能包含如下立场标签：Support（即Favor），Refute（即Against），Comment（即Neutral），Unrelated。参考Conforti等(Conforti et al., 2020)的做法，本文将每个目标都作为测试集并在其余三个目标数据作为训练集，并随机选取15%训练集数据作为验证集。数据集详情在表4中显示。

4.2 训练设置

本文使用预训练好的768维词嵌入的uncased BERT-base(Devlin et al., 2019)作为编码器。[0]，其中学习率设置为0.000005；根据Xu等(Xu et al., 2018)的做法将L_2正则化系数λ设置为0.00001；优化器使用Adam；批量大小设置为16。本文使用gensim库中的LDA获取每个目标的关联主题词，根据实际数据分布设置T=10，K=10，并去除了重复的主题词。本文训练过程中设置在超过5轮迭代仍未能提升性能时模型将提前停止。下文记录的实验数据均是10次运行试验结果的平均值，以获得统计学上的稳定结果。

4.3 评价指标

对于SemEval2016数据集本文参考Allaway等(Allaway et al., 2021)的做法使用F_{avg}衡量性能，也就是在Favor和Agianst两种立场标签上的F1的平均值。对于WT-WT数据集本文参考Conforti等(Conforti et al., 2020)的做法使用每个目标的Macro F1衡量性能。

4.4 主要结果

在表5中记录了在两个基准数据集上进行零样本立场检测的主要实验结果。BiCond方法来自 (Augenstein et al., 2016),CrossNet方法来自 (Xu et al., 2018)，BERT方法来自 (Devlin et al., 2019)。可以观察到本文提出的ST-PL方法在两个数据集上的性能大都优于传统模型，也优于直接使用提示学习对主任务进行学习，这验证了本文提出的方法在零样本立场检测中的有效性。这表明使用提示学习进行自监督学习代理任务以获取目标无关的特征的监督信号，对于学习相对模型不可见的目标的可迁移立场特征是有效的，并可凭借此特征提高零样本立场检测的性能。

Model	SEM16 (%)						WT-WT (%)			
	DT	HC	FM	LA	A	CC	CA	CE	AC	AH
BiCond	30.5‡	32.7‡	40.6‡	34.4‡	31.0‡	15.0‡	56.5♯	52.5♯	64.9♯	63.0♯
CrossNet	35.6	38.3	41.7	38.5	39.7	22.8	59.1♯	54.5♯	65.1♯	62.3♯
BERT	40.1‡	49.6‡	41.9‡	44.8‡	**55.2**‡	37.3‡	56.0♭	60.5♭	67.1♭	67.3♭
PET	45.6	50.9	49.3	46.7	45.8	32.3	68.6	63.7	70.7	71.5
ST-PL (ours)	**48.4**	**53.7**	**51.2**	**48.1**	52.2	**35.2**	**71.2**	**68.6**	**73.5**	**75.7**

Table 5: ST-PL在两个零样本立场检测数据集上的实验结果。带‡ 符号的结果取自文献 (Allaway et al., 2021)，带♯ 符号的结果取自文献 (Conforti et al., 2020)，带♭ 符号的结果取自 (Liang et al., 2021)

.

Model	SEM16 (%)						WT-WT (%)			
	DT	HC	FM	LA	A	CC	CA	CE	AC	AH
ST-PL (ours)	**48.4**	**53.7**	**51.2**	**48.1**	**52.2**	**35.2**	**71.2**	**68.6**	**73.5**	**75.7**
PET	45.6	50.9	49.3	46.7	45.8	32.3	68.6	63.7	70.7	71.5
w/o Topic Information	38.4	45.0	38.6	41.0	46.4	33.3	67.2	63.2	68.6	71.5
w/o Prompt Learning	46.2	51.6	47.8	46.6	54.4	37.8	69.9	67.1	74.2	73.0
w/o Pretext Task	40.1	49.6	41.9	44.8	55.2	37.3	56.0	60.5	67.1	67.3

Table 6: 消融实验结果

.

4.5 消融研究

在表6中记录了本方法的消融实验结果，其中PET表示直接使用提示学习进行零样本立场检测，w/o Topic Information表示在代理任务中不对主题词进行隐藏，w/o Prompt learning表

[0]由于BERT-base也是本文用于比较的基线模型，故本方法也是基于BERT-base进行构建以保证公平的比较。

示在代理任务中不使用提示学习仅使用普通过拟合模型BERT对隐藏主题词后的样例进行立场极性预测，w/o Pretext task表示不使用代理任务直接使用主模型进行零样本文本立场检测。可以看出，ST-PL方法构建代理任务学习主题相关的立场可迁移信息是有效的，因为直接使用PET进行零样本文本检测并不能取得比ST-PL的代理任务+主任务架构更好的成果。该结果也验证了提示学习的提出动机，即提示学习更适合对样例中的主题词和其他词的关联进行学习，并且能够捕捉隐藏信息前后对句子带来的变化，而非直接应用于立场检测这一本质文本分类的任务。此外，对主题词隐藏这一步骤的消融带来了最明显的性能下滑，验证了本方法挖掘主题词信息以关联目标信息这一动机；而其他步骤的任何一个缺失也都会带来性能下降，说明这些设置对零样本文本立场检测的性能提升都是必要的。

4.6 案例分析

在表7中，样例1在本文基于提示学习的代理任务下其可迁移与否的标签均判断正确。可以观察到方法在短句中表现良好，因为短句的立场表达词少且语法结构简单。在样例1中，提示学习模型预测隐去信息后数据的立场极性与原样例的立场极性相同，意味着该样例的立场表达是零样本立场检测中的一个可迁移表达，所以模型为这个样例生成一个目标无关标签。然而，在样例2中，模型对其可迁移与否的标签均判断错误，提示学习模型预测了与原样例不同的立场极性，判定该样例的倾向是随着目标信息的变化而会发生改变的，因此模型为这个样例生成一个目标依赖标签。根据类似样例2的复杂句数据的预测准确度相对偏低现象，可以推测出本课题所提出的方法在处理具有复杂语法结构、相异立场态度词表达和复杂语言现象（如讽刺、俗语等）的长句子时，代理任务由于当前模型立场极性预测能力的客观瓶颈产生了错误的可迁移表达预测，这一问题待未来研究进行进一步改善。

样例1
目标：Donald Trump 立场极性：Favor 样例句：one of the key promblems today is that policis is such a disagrace, good people don't go to government 隐藏目标和主题词后的句子：one of the key promblems today is that policis is such a $[MASK]$,$[MASK]$ people don't go to government 正确的可迁移与否标签：目标无关的 模型预测的可迁移与否标签：目标无关的
样例2
目标：Donald Trump 立场极性：Against 样例句：donald trump did not apply to immigrants one of the trade basis , win to win . ignorance can not be excuse 隐藏目标和主题词后的句子：$[MASK]$ did not apply to immigrants one of the trade basis,win to win . $[MASK]$ can not be excuse 正确的可迁移与否标签：目标无关的 模型预测的可迁移与否标签：目标依赖的

Table 7: 两个典型样例

5 结论

本文提出了一种基于主题提示学习的零样本立场检测方法(ST-PL)，该方法通过结合提示学习和代理任务生成增强数据，能有效帮助主任务模型完成零样本立场检测。具体地，ST-PL方法利用了面向目标不变和依赖于目标的立场特征之间的差异来学习可迁移的立场特征，从而能显式地将可迁移的立场特征用于零样本立场检测中，并有效提升零样本立场检测的性能。最终在两个零样本立场检测基准数据集上进行的实验结果表明，本文提出的ST-PL方法性能表现全面优于基线模型。具体的案例分析表明，该方法在语法结构相对简单、立场词态度表达一

致的数据上表现优异，对于复杂语法、相异立场态度词表达与复杂语言现象句的可迁移预测受限于现有模型的预测瓶颈表现一般，有待未来改进。

参考文献

Allaway E, Mckeown K. 2020. *Zero-Shot Stance Detection: A Dataset and Model using Generalized Topic Representations. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*:8913–8931.

Allaway E, Srikanth M, Mckeown K. 2021. *Adversarial Learning for Zero-Shot Stance Detection on Social Media. Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*:4756–4767.

Augenstein I, Rocktäschel T, Vlachos A, et al. 2016. *Stance Detection with Bidirectional Conditional Encoding. Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*:876–885.

Chen T, Kornblith S, Norouzi M, et al. 2020. *A simple framework for contrastive learning of visual representations. International conference on machine learning. PMLR, 2020*:1597–1607.

Conforti C, Berndt J, Pilehvar M T, et al. 2020. *Will-They-Won't-They: A Very Large Dataset for Stance Detection on Twitter. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*:1715–1724.

Devlin J, Chang M W, Lee K, et al. 2019. *Bert: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*:4171–4186.

Du J, Xu R, He Y, et al. 2017. *Stance classification with target-specific neural attention networks. International Joint Conferences on Artificial Intelligence*:3988–3994.

Gidaris S, Singh P, Komodakis N. 2018. *Unsupervised Representation Learning by Predicting Image Rotations. International Conference on Learning Representations,2018.*

Gunel B, Du J, Conneau A, et al. 2021. *Supervised Contrastive Learning for Pre-trained Language Model Fine-tuning. International Conference on Learning Representations,2021.*

Kachuee M, Yuan H, Kim Y B, et al. 2021. *Self-Supervised Contrastive Learning for Efficient User Satisfaction Prediction in Conversational Agents. Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*:4053–4064.

Kawintiranon K, Singh L. 2021. *Knowledge enhanced masked language model for stance detection. roceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*:4725–4735.

Larsson G, Maire M, Shakhnarovich G. 2016. *Learning representations for automatic colorization. European conference on computer vision. Springer, Cham, 2016*:577–593.

Liang B, Fu Y, Gui L, et al. 2021. *Target-adaptive graph for cross-target stance detection. Proceedings of the Web Conference 2021*:3453–3464.

Liu R, Lin Z, Tan Y, et al. 2021. *Enhancing zero-shot and few-shot stance detection with commonsense knowledge graph.. Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*:3152–3157.

Li Y, Caragea C. 2019. *Multi-task stance detection with sentiment and stance lexicons. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*:6299–6305.

Mohammad S, Kiritchenko S, Sobhani P, et al. 2016. *Semeval-2016 task 6: Detecting stance in tweets. Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*:31–41.

Noroozi M, Favaro P. 2016. *Unsupervised learning of visual representations by solving jigsaw puzzles. European conference on computer vision. Springer, Cham, 2016*:69–84.

Petroni F, Rocktäschel T, Riedel S, et al. 2019. *Language Models as Knowledge Bases?. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*:2463–2473.

Schick T, Schütze H. 2021. *Exploiting Cloze-Questions for Few-Shot Text Classification and Natural Language Inference. Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics*,Main Volume 2021:255–269.

Shin T, Razeghi Y, Logan IV R L, et al. 2020. *AutoPrompt: Eliciting Knowledge from Language Models with Automatically Generated Prompts. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*:4222–4235.

Siddiqua U A, Chy A N, Aono M. 2019. *Tweet stance detection using an attention based neural ensemble model. 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):1868–1873.

Simard N, Lagrange G. 2021. *Improving Few-Shot Learning with Auxiliary Self-Supervised Pretext Tasks. arXiv preprint arXiv:2101.09825, 2021.*

Somasundaran S, Wiebe J. 2010. *Recognizing stances in ideological on-line debates. Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text*:116–124.

Sun Q, Wang Z, Zhu Q, et al. 2018. *Stance detection with hierarchical attention network. Proceedings of the 27th international conference on computational linguistics*:2399–2409.

Wei P, Mao W. 2019. *Modeling transferable topics for cross-target stance detection. Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*:1173–1176.

Xu C, Paris C, Nepal S, et al. 2018. *Cross-Target Stance Classification with Self-Attention Networks. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*:778–783.

Zhang B, Yang M, Li X, et al. 2020. *Enhancing cross-target stance detection with transferable semantic-emotion knowledge. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*:3188–3197.

Zhang R, Isola P, Efros A A. 2016. *Colorful image colorization. European conference on computer vision.Springer, Cham, 2016*:649–666.