逻辑学研究　2023 年第 2 期，49–65
文章编号：1674-3202(2023)-02-0049-17

论辩文本立场检测
——基于提示模型的小样本研究

鲜于波　　黄伟鑫

摘　要：立场检测研究旨在提取文本相对于对特定话题所持的立场倾向。本文采用自然语言处理领域的基于提示学习的方法，针对论辩性文本，根据提示学习两大主要工程方法（模板工程和表达器工程）提出了新的设计方法：掩码位置导向的手工模板 MPOT 和语义相似度加权表达器 SSWV。据此，本文提出了一个基于提示的论辩文本立场检测模型 $P-RoBERTa_{\mathrm{MPOT}+\mathrm{SSWV}}$，并和自动模板方法 P-tuning 进行比较。该模型在中山大学网络文本论辩语料库上和 NLPCC 中文微博立场检测数据集都达到了较好的准确率。本文实验显示在小样本学习条件下，$P-RoBERTa_{\mathrm{MPOT}+\mathrm{SSWV}}$ 模型超过了使用预训练模型 + 微调方法的效果。本文研究表明，提示学习的模型设计方法有助于文本论辩立场检测任务，并在小样本的条件下也依然能取得十分可观的性能。

关键词：立场检测；提示模型；论辩；小样本

中图分类号：B81　　**文献标识码：**A

1　引言

按照经典论辩理论，论辩是单个或多个主体之间通过单个命题或命题组合来证明自身观点、表达自身立场、反驳他方观点，以达到说服对方、消除争议、谋求共识的理性行为。（[24]）

由于互联网的高速发展，用户在网上产生了海量的论辩性文本。随着人工智能的发展，尤其是自然语言处理技术的迅速发展，越来越多的研究开始借助机器学习等方法对这些文本进行计算分析。在这些工作中，立场检测旨在检测论辩者针对特定话题所发表的观点的立场倾向，研究如何从非结构化自然语言文本中识别其立场。无论是从其应用前景还是学术价值来看，立场检测都是非常重要的内容，因此立场检测已经成为一个新兴的热门研究领域。

收稿日期：2022-03-10

作者信息：鲜于波　中山大学逻辑与认知研究所
中山大学哲学系
xianyubo@mail.sysu.edu.cn
黄伟鑫　中山大学逻辑与认知研究所
中山大学哲学系
562718834@qq.com

基金项目：国家社科重大项目（18ZDA033）。

从现有的研究来看，大部分的立场检测都集中在法庭辩论（[23]）、或者在线论坛中的讨论（[26]）和微博（[32]）上，相关研究也取得了一定的进展。但是在中文界，论辩挖掘的研究还没有受到充分的重视，现有大部分立场检测主要分析社交媒体。从研究的现状来看，由于问题本身的挑战性，立场识别的准确性和模型的通用性都还有较大的改进空间。

近几年来，预训练大模型的兴起（[12]）标志自然语言处理范式发生了很大的变化。然而人们逐渐发现对大模型进行微调的要求也越来越高，复杂多样的下游任务也使得预训练和微调架构的设计变得十分困难繁琐。因此，研究者们迫切希望能有一种更加高效、轻量级别的少样本学习方法，而提示学习就代表了这方面的最新研究范式。

在中文自然语言处理领域，有关提示学习的研究方兴未艾，但是使用提示模型进行论辩研究的很少，特别是关于论辩文本立场检测的相关工作也不多。此外，作为论辩挖掘研究基础的论辩语料库的构建也是一件很耗费人工的工作。不过提示学习的少样本学习的特点和优势非常有利于论辩领域的研究，可以在一定程度上解决数据稀缺、模型微调难和泛化能力差等问题。因此，将提示方法应用到论辩文本立场检测就有良好的理论和实践意义。

本文通过开展论辩文本立场检测的研究，在新构建的论辩数据库上采用新的提示学习方法进行模型构建和小样本实验，表明提示学习的模型设计方法有助于文本论辩立场检测任务，并在小样本的条件下也依然能取得十分可观的性能。

主要创新在于针对论辩文本立场检测任务下实现了提示学习的方法。本文设计了两种新型的适用于立场检测任务方法：掩码位置导向手工模板和语义相似度加权表达器，还实现了提示学习领域自动模板生成方法（[13]）在立场检测上的设计，并与手工模板进行了实验对比。实验结果显示本文基于提示学习的中文立场检测模型 $P-RoBERTa_{\mathrm{MPOT+SSWV}}$ 达到了较好的效果：通过使用少量数据训练，该模型就能达到与主流大模型微调模型相近的效果，特别是在中文论辩数据集上实现了较大的性能改进。

本文的内容安排如下：首先分析论辩文本立场检测的发展与现状，并对基于提示学习的立场检测模型的设计思路和方法进行介绍，然后在一般提示学习模型的主要框架上，引入了 $P-RoBERTa_{\mathrm{MPOT+SSWV}}$ 中的模板工程和表达器工程，随后进行模型的实验评估与分析，最后是分析和展望。

2 相关工作分析

文献中关于立场检测任务主要有三种主流定义：通用立场检测（[14]），谣言立场检测（[30]）以及假新闻立场检测（[8]）等。通用立场检测又可以分为多目标立场检测（[19]）和跨目标立场检测（[3]）。目前文献中数量最多、最常见的关于立场检测的定义是单目标立场检测。从关注的问题来看，论辩文本的立场检测

属于一种特殊的立场检测类型，但是由于论辩文本的特点，论辩立场的检测也自有特殊之处。

立场检测的早期工作中，一个典型的方法来自 Somasundaran 等（[20]），他们把重点放在了在线辩论上。其他的一些研究则增加了对讨论线程之间互动性的研究，例如 Stede 等（[22]）强调了反驳（rebuttals）在讨论当中的重要性。Hasan 等（[9]）对意识形态论辩进行了研究。对于在线辩论，Sridhar 等（[21]）的研究将语言学特征与基于网络结构的特征进行了结合。还有一些立场检测研究分析了学生论文，如 Faulkner 等（[6, 32]）。

在现有文献中，现有的立场检测方法大致可以分为两大类：基于特征的统计机器学习方法和深度学习方法。（[33]）

在现有的研究中，基于特征的传统统计机器机器算法如支持向量机（[9]）、决策树和随机森林（[1]）、隐马尔可夫模型 HMM 和条件随机场 CRF（[9]），K 近邻算法（[18]），对数线性模型（[5]），最大熵（[10]）等方法都得到广泛的应用。

近年来，随着人工智能的发展，深度学习（如 RNN 及其变体和 CNN 等）被大量应用于立场检测的研究中，如双向 LSTM（[3]）、卷积神经网络 CNN （[36]）和加入注意力机制 Attention 的神经网络方法（[29, 32]）等。

深度学习虽然功能强大，但存在模型体积大、训练时间长等特点，并且不同的下游任务需要训练和保存不同的模型，这要消耗大量的资源空间。因此随着预训练和大规模语言模型尤其是 GPT（[16]）的发展，提示学习成为自然语言处理领域的一种新范式。大模型如 bert （[25]）、GPT-3（[16]）等带来了一种新处理下游任务的方法：通过使用自然语言提示信息和任务示例作为上下文，因此只需少量样本甚至零样本。该方法能更好地利用预训练模型中的知识，能获得更好的性能和增强模型的泛化性能。

提示学习方法在许多自然语言处理任务上均可获得不错的效果。随着这一新范式的提出，许多研究者开始针对不同任务手工设计提示模板（[4, 15, 27]）等。但是手工设计提示模板比较耗时耗力，因此最近也有一些针对自动化模板的方法研究。例如 Hambardzumyan（[7]）等提出了一种自动提示生成方法，简化了提示模板的设计。

在提示学习中，一项十分重要的工作是设计表达器或映射（verbalizer），如 Schick 等（[17]）手工制作的表达器等。但手工设计受先验知识影响较大，还需要足够的数据集来进行优化调整。有鉴于此，一些研究提出了提示学习的自动表达器的构造方法，如 Wei 等（[28]），还有一些其他工作尝试从外部知识库中选择相关词等（[11]）。

总的来看，提示学习新范式具有很多的优势。使用提示模板，使得模型可以在少样本学习条件下也能达到较高性能。提示学习通过将下游任务融入提示，设计模板的工作量和资源耗费要大大小于微调预训练模型。此外，提示学习中手工

模板是由自然语言构成的，这意味着人们可以比较方便地借助自己对下游任务相关的领域知识来设计模板，提高模型的可解释性。

从上面的分析也可以看到，现有研究在论辩文本立场的研究中采用提示学习方法的研究很少，尤其是中文世界在论辩挖掘方面的研究还处于一个起步阶段。因此，在中文文本包括网络文本分析领域，采用提示学习和少样本学习的方法对论辩文本立场进行分析是非常有必要的和有意义的工作。

3 立场检模型设计

3.1 提示学习框架

本文立场检测模型基于提示学习，这里提示学习主要框架包括提示模板工程、预训练模型以及表达器工程三个大模块。（[12]）

3.1.1 提示模板工程

在提示学习方法中，第一步是提示模板工程。本文先给出提示学习在文本分类任务中的一般定义。给定一个输出 $x = \{x_1, \ldots, x_n\}$，其真实标签 $y \in Y$（Y 表示所有类别的标签集），标签词集 V_y 定义为 $\{v_1, \ldots, v_n\}$，其中 $V_y \in V$（V 表示模型的整个词汇表）被映射到标签为 y 的类别中。在预训练语言模型 M 中，V_y 的每个词 v_i 被填充到掩码（[MASK]）的概率可以表示为：$P([\text{MASK}]= v \in V_y|x_p)$

由此，立场检测的任务可以转换为标签词的概率计算问题，计算公式如公式 (1):

$$P(y \in Y|x) = P([\text{MASK}]= v \in V_y|x_p) \tag{1}$$

3.1.2 预训练模型

关于预训练模型，翁沫豪（[34]）最近的研究对比了各类预训练模型在论辩立场检测任务上的性能，他的实验证明 *RoBERTa*（[35]）在立场分类任务上表现最好。*RoBERTa* 在其他场合也得到广泛的应用。因此本文将采用 $\textit{RoBERTa}_{\text{base}}$ 进行实验。

3.1.3 表达器工程

表达器是一类用来将标签词映射到类别词的函数以缓解文本与标签空间的差异。一种最常见的做法是用与标签 y 相关的词对 V_y 进行扩展，例如在话题识别任务中 $V = \{$“体育”$\}$ 可以通过知识图谱、语义关联等方法进行扩展，如：

$V = \{$“体育”,“篮球”,“解说”,“射门”,“分”$\}$

在立场检测任务下，必须注意的是，立场检测任务的文本所涉及的领域并不是某一特定领域，标签之间的领域存在交叉性，这在模型设计中应予以考虑。

3.2　立场检测模板设计

模板工程的方法可以分为手工设计模板和自动模板生成两大类。手工设计的模板采用自然语言的形式，由词汇表中离散的词或 token 组成。由于手工模板是较为耗时费力的任务，因此开始学术界提出了自动化的模板生成方法。（[13]）

本文分别在手工与自动模板上进行了立场检测模型的设计。在手工模板设计上使用掩码位置导向的手工模板，在自动模板生成方法上使用 P-tuning 方法实现立场检测任务。

3.2.1　掩码位置导向的手工模板

如采用手工方式设计一个立场检测模板，首先一个问题是如何将立场检测的任务特点与模型的架构结合起来。本文采用的 *RoBERTa* 是一种掩码语言模型，那么我们需要设计的模板就应该是通过 MASK 遮蔽某些词，然后让模型预测出它的值，然后将预测出的值映射到数据的真实标签上。

一个做法是将立场检测任务转换成一个完形填空问题。但与一般的文本分类任务不同的是，立场检测任务有两个输入，一个是话题，一个是该话题下的一段文本。上述两个输入关联性很强，如何将这两个输入关联起来，就是立场检测任务进行手工模板设计的主要挑战。

本文从 MASK 出发，以 MASK 在模板中的位置关系作为导向，设计了掩码位置导向的手工模板– MPOT（Mask Position oriented Template）。为了尽可能的对比各类模板以体现方法的可靠性和一般性，本文从模板中出现的输入和掩码的位置关系出发，穷举了所有可能并对每种可能的位置关系设计了一个手工模板。具体方法如下。用 X 表示输入的话题，T 表示输入的文本，M 表示掩码 MASK。根据三者在模板中的位置关系，总共有 XTM、TXM、XMT、TMX、MXT、MTX 六种情况，分别设计了如下六种手工模板。

P_1 =: X 话题中，观点 T 持 M 态度
P_2 =: 在 X 话题中持 M 态度
P_3 =: 以下观点在 X 话题中持 M 态度：T
P_4 =: 观点 T，M 话题 X
P_5 =: M 话题 X：T
P_6 =: 一个 M 观点：T，话题：X

根据掩码 M 所在的位置，模板可以分为三类：前置掩码模板：P_1, P_2，中置掩码模板：P_3, P_4，后置掩码模板：P_5, P_6。同时，由于立场检测任务的分类标签

是支持、反对以及中立，均为两个字组成的词。*RoBERTa* 的中文分词器是以字为单位进行分割的，因此我们的模板在实际实验时需要添加两个空位，如对模板 P_1 来说，实际输入时应该为“在 X 话题中，观点 T 持 MM 态度”。

3.2.2 自动提示模板

本文对立场检测模板如何自动化构建进行了探究。自动模板设计的一项新方法来自 Liu 等（[13]）提出的一种名为 P-tuning 的方法，该方法的特点是可以自动搜索连续空间中的提示，使用梯度下降方法将模板的构建转化成了连续参数优化问题。

据此，本文设计了立场检测自动提示模板（Auto P-tuning for Stance Detection），其主要结构如下：

令 V 表示预训练模型 M 的词汇表，$[P_i]$ 表示模板 P 中的第 i 个 token。X 表示话题，T 表示文本，Y 表示标签。P-tuning 将 $[P_i]$ 视为伪 token 并将模板 P_{auto} 映射为：

$$P_{\text{auto}} = \{h_0, \ldots . h_i, e(X), h_{i+1}, \ldots, h_j, e(T), h_(j+1), \ldots, h_m, e(Y)\} \tag{2}$$

其中，$h_i \in V(0 \leq i \leq m)$ 是可学习的嵌入向量（Embedding vector），将其定义为新的输入嵌入层作为模型编码器 Encoder 的一部分。在训练的过程中，保持预训练模型参数不变，仅更新构造的嵌入层参数，最后，根据本任务的目标函数 L，对连续提示进行优化。

$$h_{i:m} = \arg_h \min\ L(M(X, T, Y)) \tag{3}$$

本文采用分类问题常用的交叉熵作为目标函数，可以通过最小化交叉熵来得到目标函数分布的近似分布，它的表达式为：

$$L(D, M) = -\sum_{i=1}^{n} D(x_i) \log M(x_i) \tag{4}$$

其中 D 为数据真实结果的概率分布，M 为模型预测结果的概率分布。

与 [13] 提出的模式类似，图 1 是掩码位置导向的手工模板 MPOT 以及实现了自动模板生成方法 Auto P-tuning 在立场检测任务上的概念示意图。

其中离散手工提示以手动模板作为例子。黄色框代表模板中固定的词；蓝色框中 X 代表输入的话题，T 代表输入的文本；红色框 [MASK] 代表掩码的所在位置（模型需要预测的位置）。在图 1 左图离散手工提示中，MPOT 模板生成器生成手工模板后，模板固定不再改变，通过预训练模型和表达器来实现预测。在图 1 右图连续自动提示中，通过伪提示词和提示编码器通过反向传播机制可以不断更新与优化。

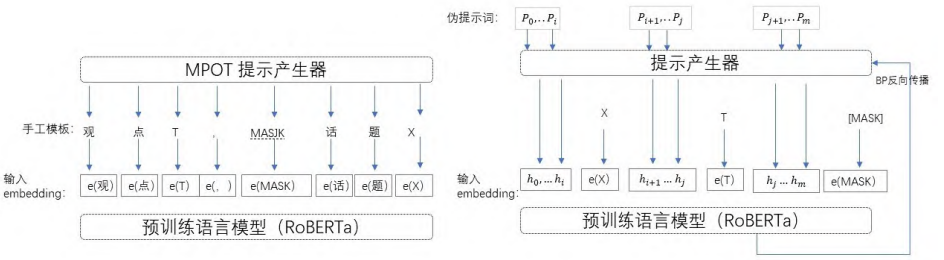

图 1: 离散手工提示和连续自动提示的示意图

3.3　语义相似度加权表达器

表达器的主要目的是构建标签词到类标签的映射。本文经过标签词集的构建与扩展、表达器的使用两个步骤构造了一个语义相似度加权表达器 SSWV（Semantic Similarity Weighting Verbalizer）。

3.3.1　标签词集的构建与扩展

对于本文的立场检测问题，真实类标签集为 $Y = \{$ 支持 , 反对 , 中立 $\}$。那么最简单、直观、基础的标签集可以构建为 $V_{\text{支持}} = \{\text{支持}\}$、$V_{\text{反对}} = \{\text{反对}\}$ 以及 $V_{\text{中立}} = \{\text{中立}\}$ 。

有了这个基础标签集后，如何对其进行扩展有不同的做法，而对于立场检测任务，一般的方法比较难以适用。任何领域、任何话题下的文本都可能存在立场，立场可以用支持、反对、中立来被划分，从这三类标签不难看出，很难找到一个领域的知识能与支持或反对或中立相匹配，且不产生知识交叉。一种方法是通过外部知识融入的方法来扩展标签词集（[11]），另外很多研究者采用了不同的分类标签命名来完成立场检测任务，例如 Gorrell 等（[11]）使用了评论（comment），支持（support），疑惑（query），否认（deny）作为立场分类标签。不难发现，这其中很多标签所表达的含义和指向的数据是相同的，比如赞成、同意、支持等，都是表达正面一方的立场。因此本文采用寻找与标签语义相近的方法对标签词集进行扩展，以支持、反对、中立三个标签作为语义中心词，寻找词向量空间中与之距离相近的词作为扩展词。

本文选择词表容量最大的中文近义词工具包 Synonyms 来进行语义相似度计算。我们利用 Synonyms 对三个标签分别进行近义词语义相似度搜索，只选择其中长度为 2 的词，并从每个标签的候选词中选择了相似度最高的 10 个词作为扩展词。扩展后的标签词集见表 1：

表 1: 立场检测扩展标签词集

	支持	反对	中立
1	支持：1.0	反对：1.0	中立：1.0
2	鼓励：0.662021	谴责：0.786821	保守：0.665451
3	拥护：0.647356	抵制：0.773777	独立：0.663212
4	支援：0.636656	指责：0.718662	模糊：0.659456
5	扶持：0.582030	抨击：0.708622	和平：0.656437
6	赞成：0.561304	抗议：0.687414	中等：0.645767
7	倡议：0.535234	杯葛：0.661550	不惧：0.634558
8	扶持：0.528623	责难：0.65755	无此：0.613412
9	扶助：0.523059	敌视：0.644686	联合：0.612123
10	默许：0.490322	阻扰：0.638986	唯有：0.604446

3.3.2 表达器的使用——基于加权平均算法

根据上小节，给定一个立场标签，其对应的标签词集已经被扩展到包含 10 个标签词的集合。下面对这些标签词在表达器中如何恰当使用进行设计。一些论文采用了平均法对标签词集的概率值进行计算，本文中，每个标签词的重要性明显是不同的，因此采用加权平均算法，将各标签词集的相似度进行归一化后作为加权权重，以区分个标签词的重要性。

具体计算公式如下：类标签 $y \in Y$ 所对应的标签词集 $V_y = \{v_1, \ldots, v_n\}$，它所对应的语义相似度集为：$S_{V_y} = \{S_{V_1}, \ldots, S_{V_n}\}$。那么根据公式 (4) 可得权重向量 $W_{V_y} = \{w_{V_1}, \ldots, w_{V_n}\}$

$$w_{V_i} = \frac{S_{V_i}}{\sum_{i=1}^{n} S_{V_1}} \left(S_{V_i} \in S_{V_y}\right) \tag{5}$$

那么，模型的输出标签 $y^{\sim}$ 的计算公式如下：

$$y^{\sim} = \arg\max_{y \in x} \frac{\exp(h(y|x_p)}{\sum_{y'} \exp(h(y'|x_p)} \tag{6}$$

其中 $(h(y|x_p) = \sum_{v \in V_y} w_v \log P_M([\text{MASK}] = v \in V_y|x_p)$

即给定一个输入 x_p，通过模板工程和预训练模型后获得某类标签 y 所对应的标签词集所有标签词的对数概率后，根据权重向量加权求和，得到该标签的条件概率值，并由此输出立场预测。

4 实验评估与分析

本文实验的第一阶段是模板筛选，对本文设计的 6 个 MPOT 手工模板和自动模板方法 P-tuning 进行对比实验，从而筛选出其中最佳的模板方式，得到本文的最

佳模型。第二阶段是少样本学习实验。目的是检验第一阶段得到的 $RoBERTa_{\text{MPOT}+\text{SSWV}}$ 在少样本场景下的性能。分别设置了 0、32、64、128 四个数量级别的训练样本数，还对原始数据集的每个话题进行分层抽样以提高样本的代表性。

4.1　数据集

所有实验在两个中文数据集上进行，分别为 NLPCC 2016 年共享任务中所发布的“中文微博立场检测”数据集和中山大学网络文本论辩语料库，前者在中文立场检测领域是最为常用的数据集。

“中文微博立场检测数据集”共包括 4000 条微博数据，被分成五个话题类，每个数据都有一个立场标签，为代表支持的“FAVOR”、代表反对的“AGAINST”以及代表中立的“NONE”中的一个。

中山大学网络文本论辩语料库是 2022 年由中山大学逻辑与认知研究所建立的一个中文论辩挖掘语料库，文本来源于国内最大的问答社区知乎，包含的话题总计有 26 个，涵盖了经济、社会等多个领域的争议性话题。该数据库也有三个标签，分别为代表支持，反对和中立的的“pro”，“con”和“non”。

表 2: 数据集大小统计表

数据集	数据量	平均字数	最大字数	最小字数
中文微博立场检测数据集	4000	85	319	11
中山大学网络文本论辩语料库	9126	55	332	13

4.2　实验设置与对比模型

4.2.1　模板筛选实验

对于两个数据集均使用 70% 的数据作为训练数据，15% 的数据作为测试数据为验证集数据，15% 的数据作为测试数据。根据表 2，数据的长度平均在与 100 以内，最大的也仅为 332，数量也很少，因此本文实验的 maxlen 统一设置为 256 以尽可能的保留更多的文本信息，并对不足长度的数据进行填充。

对于微博数据集，在每个模板上均训练了 30 个轮次，对于中山大学网络文本论辩数据集，由于数据集更大，对每个模板训练了 50 个轮次。优化后的实验参数设置见表 3。

4.2.2　少样本学习实验

该部分实验设置了 0、32、64、28 四个数量的样本训练集。由于数据量不同，模型最终收敛的轮次也不同，这里每个实验均取过度拟合前最大的指标值。

表 3: 模板筛选实验参数表

模型	数据集	轮次	优化器	学习率	权重衰减	maxlen	批大小
MPOT/AUTO P-Tunning+ *RoBERTa*+SSWV	中文微博立场检测数据集	30	Adam	1*e*-5	0.1	256	16
	中山大学网络文本论辩语料库	50					

实验对比的主要模型有：

微博数据集的对比模型：奠雨洁等（[31]）中文微博立场检测模型以及赵姝颖（[35]）。另外国外相关研究也进行了参考 [2]。论辩语料库的对比模型：翁沫豪（[34]）对所构建的论辩语料库进行了立场检测任务，在 $RoBERTa_{\text{base}}$ 上达到了 70.6% 的准确率。

本文对 $RoBERTa_{\text{base}}$ + Fine tuning 在少样本实验上保持翁（[34]）参数设置，学习率为 $2e-5$，训练批次 3 轮。由于奠雨洁等（[31]）并没有公开其模型的具体实验代码与参数，无法对其进行少样本学习实验。本文主要与其完整训练的模型结果进行对比。

4.3 实验结果与分析

4.3.1 模板筛选实验

表 4: 模板筛选实验结果

模型	NLPCC 中文微博检测数据集				中山大学网络文本论辩语料库			
	Prec	Rec	F1	Acc	Prec	Rec	F1	Acc
P_1-*RoBERTa*-SSVW	0.66	0.65	0.66	0.6756	0.84	0.81	0.8231	0.8316
P_2-*RoBERTa*-SSVW	0.68	0.65	0.66	0.6826	0.84	0.83	0.8328	0.8427
P_3-*RoBERTa*-SSVW	0.7	0.66	0.67	0.7052	0.87	0.84	0.8586	0.8777
P_4-*RoBERTa*-SSVW	0.66	0.63	0.65	0.6585	0.85	0.84	0.8419	0.8674
P_5-*RoBERTa*-SSVW	0.67	0.64	0.65	0.6836	0.87	0.85	0.8674	0.8823
P_6-*RoBERTa*-SSVW	0.69	0.67	0.68	0.7017	0.89	0.84	0.8725	0.8918
P-tuning-*RoBERTa*-SSVW	0.53	0.63	0.58	0.6173	0.53	0.68	0.61	0.6673

在两个数据集上的实验结果都记录在了表 4 中，从实验结果可以看出：

(1) 手工模板 P_6 在两个数据集上均取得了最高的 F1-Score，分别为 0.68 和 0.8725，并且在论辩语料库中取得了最高的准确率 0.8918。虽然在微博数据集上

P_6 的准确率低于 P_3，但只相差不到 0.003。因此，根据实验结果可得 P_6 是当中最佳的立场检测手工模板。

(2) 前置掩码的模板要比中置、后置掩码的模板效果要好（前置掩码模板：P_1, P_2，中置掩码模板：P_3, P_4，后置掩码模板：P_5, P_6）。其可能原因有以下两点：一是前置的掩码，其位置在长度不断变化的数据加入后，掩码的 Mask 的位置比较固定。例如对比模板 P_5 和模板 P_1，在文本数据输入并生成模板后，P_5 中的 MASK 位置始终保持在首位（除模型固定的特殊符号 [CLS] 外）；而对于 P_1 模板，其位置则会随着输入的观点和话题的长度大小而改变。相对固定的 MASK 位置会给与模型一定的监督信号，并在不断的训练过程中叠加增强。另一个可能的原因是与模型学习的难度有关，可能数据前置位的规律更容易被识别，相对与后置位更容易学习。

(3) 对比手工模板和自动模板方法，可看出 P-tuning 在两个数据集上的表现不尽人意，而且 P-tuning 方法在收敛速度较慢。

(4) 对比数据集之间的性能差别，中山大学网络文本论辩语料库（以下简称论辩语料库）比中文微博立场检测数据集性能高。这说明了论辩语料库中的文本立场能更准确被模型所检测，这体现了论辨性越高的文本，立场就越容易识别。

经过以上对第一阶段实验结果分析，本文通过掩码位置导向的手工模板方法 MPOT 选择出的最佳模板为手工模板 P_6，结合所使用的 *RoBERTa* 预训练模型和语义相似度加权表达器，得到提示学习的立场检测模型：$P-RoBERTa_{\text{MPOT +SSWV}}$。

4.4　少样本实验

表 5: 少样本学习实验结果

Molde	Shot	微博数据集				论辩语料库			
		Proc	Rec	F1	Acc	Prec	Rec	F1	Acc
P−*RoBERTa* (MPOT+ +SSWV)	0	0.39	0.48	0.4365	0.4959	0.48	0.57	0.5265	0.5859
	32	0.43	0.51	0.4837	0.5244	0.52	0.6	0.5737	0.6144
	64	0.51	0.55	0.5461	0.5869	0.606	0.646	0.6421	0.6629
	128	0.58	0.6	0.5931	0.6075	0.646	0.676	0.6591	0.6835
	A11	0.69	0.67	0.68	0.7017	0.89	0.84	0.8725	0.8918
RoBERTa +FT	0	0.18	0.22	0.1912	0.2177	0.232	0.272	0.2432	0.2697
	32	0.19	0.23	0.2024	0.2384	0.242	0.282	0.2544	0.2904
	64	0.2	0.25	0.2262	0.2546	0.252	0.302	0.2782	0.3066
	128	0.23	0.26	0.2551	0.2744	0.282	0.312	0.3071	0.3264
	A11				0.6483				0.706
SVM+ 随机森林	A11		0.7106						

少样本学习实验结果如表 5，模型 $P-RoBERTa_{\text{MPOT}+\text{SSWV}}$ 仅用 128 个数据训练的条件下在两个数据集上分别达到了 60.75% 和 68.35% 的准确率。从实验结果可见：

(1) $P-RoBERTa_{\text{MPOT}+\text{SSWV}}$ 在零样本学习下均取得了 50% 左右的准确率。如果直接使用 *RoBERTa* 的预训练任务进行零样本预测，其准确率在两个数据集上均不到 30%。这反映了本文设计的 MPOT 和 SSWV 确实能够提高 *RoBERTa* 的预测能力。

(2) 随着实验从 0-shot 到 128-shot 的变化，所有方法的性能都有所提升。这表明增加有标签的数据数量可以一定程度上增加少样本学习的效果。

(3) 针对微博数据集而言，$P-RoBERTa_{\text{MPOT}+\text{SSWV}}$ 在与使用 128 个数据的少样本学习条件下达到了 60.75%，与使用预训练模型 *RoBERTa* 在所有数据的条件下进行微调的结果 64.83% 只相差了 4 个百分点，而两者所使用的数据量相差大约 23 倍之多，训练时间相差 15 倍。由此可见，基于提示学习的 $P-RoBERTa_{\text{MPOT}+\text{SSWV}}$ 模型在少样本学习下，几乎可以媲美预训练模型 + 微调的方法，这将节省大量的数据标注、模型训练的人力和时间。

(4) 针对论辩语料库而言，$P-RoBERTa_{\text{MPOT}+\text{SSWV}}$ 在少样本学习下所达到的效果几乎与方法持平（在 128-shot 下仅相差 2 个百分点）。而如使用所有数据，更是大幅超越了 *RoBERTa* + FT 的方法，提升了 18.58%。

总的来说，模型 $P-RoBERTa_{\text{MPOT}+\text{SSWV}}$ 在大多数情况下达到了较好的表现，尤其是在少样本学习的场景下以及在论辩语料库上实现了较大幅度的性能提升，并且模型在具有一定差异性的数据集上均实现了稳定的性能，这也表明了该模型具有较好的泛化性能。

4.5 补充实验

为了更好地理解模型的内部结构以及模型对数据的处理，本文进行了三个补充实验，以观察所设计的 SSWV 表达器对模型的影响。

补充实验一：表达器 vs. SSWV 表达器

表 6: 表达器对比实验

模型	NLPCC 中文微博立场检测数据集	中山大学网络文本论辩语料库
P−RoBERTa (MPOT+SSWV)	0.7017	0.8918
P−RoBERTa (MPOT)	0.6215	0.7852

本实验保持其他模型架构不变，仅改变其中的表达器。以未进行任何知识扩展的表达器，只是将标签类名作为标签词的作为对比表达器，结果显示，本文设

计的 SSWV 在不同数据集上均使模型有较大幅度的提升（分别约为 8%、11%）。这体现出使用语义相似度对进行标签扩展的方法的有效性。

补充实验二：立场检测各类别的模型指标

本文的立场检测有三个类别，分别是支持、中立和反对。本实验对模型在这三个类别数据上的性能进行了统计，使用精确率、召回率和 F1-Score 作为评价指标。实验结果见表 7：

表 7: 类别指标对比实验

模型	Label	Prec	Rec	F1
P−RoBERTa (MPOT+SSWV)	支持	0.88	0.94	0.91
	中立	0.84	0.64	0.73
	反对	0.89	0.94	0.92

通过实验结果，可以看到本文模型在支持和反对类别的数据上均有较好的性能表现，F1-Score 均超过 0.9，而在中立类别的数据上表现较弱一些。这显示了模型对立场鲜明的数据有更好的预测能力，而对立场不明确的文本表现稍弱。

补充实验三：语料库大小对自动模板生成方法 P-tuning 的性能影响

由前面的结果可见，自动模板生成方法 P-tuning 的性能表现欠佳，可能原因之一有数据样本过少。本文在中山大学网络文本论辩语料库的基础上，扩充该语料库，观察 P-tuning 的性能。我们采用了增加现有的语料库的方式。UKP Sentential Argument Mining Corpus 是一个英文文本论证语料库，含 8 个话题共 25492 个句子的数据，每个论证采用“话题 + 论证句 + 立场”的形式。本文通过使用谷歌翻译获得了相应的中文文本。然后将其与原语料库混合并随机打乱进行实验，结果显示在表 8 ：

表 8: 扩充语料库对自动模板生成实验

模型	语料库	数据量	准确率
P-tuning-*RoBERTa*-SSVW	中山大学论辩语料库	9126	0.6673
	中山大学论辩语料库＋UKPSentential Argument Mining	24618	0.7112

由此可见，扩充语料库对自动模板生成确实有一定的准确率的提升（5%），但对比手工模板仍然有一定的差距，这说明了单纯扩大语料库规模的方法还是有一定的局限性。

5 结语与展望

本文结合自然语言处理领域最新的提示学习方法对立场检测任务展开了研究，提出了一种新型的基于提示学习的立场检测模型 $P-RoBERTa_{\mathrm{MPOT+SSWV}}$，获得了较好的效果，并在少样本学习的场景下取得了与预训练 + 微调的方法相近的性能。实验显示，在零样本学习场景下，提示学习的方法确实能改进现有的结果，这显示了提示学习的有效性和发展前景。实验也表明论辩结构越明显的文本，其立场检测效果越高。这既体现出论辩文本的特点，也证明提示学习在论辩文本立场的分析上可以发挥更多的作用，值得我们进一步去探索。

本文也存在一些不足和有待改进之处，如在提示学习框架中，对不同体量和架构的预训练模型的对比研究或采用更新的预训练大模型也是十分有必要的。此外，如何更好地设计具体手工模板的语言是一个重要问题，如何改进自动模板生成方法以及提出更好的提示方法也有待将来进一步的研究。

参考文献

[1] A. Addawood, J. Schneider and M. Bashir, 2017, “Stance classification of twitter debates: The encryption debate as a use case”, *Proceedings of the 8th international conference on Social Media & Society*, pp. 1–10, http://dx.doi.org/10.1145/3097286.3097288.

[2] A. Aldayel and W. Magdy, 2021, “Stance detection on social media: state of the art and trends”, *Information Processing & Management*, **58(4)**: 102–597.

[3] I. Augenstein, T. Rocktäschel, A. Vlachos *et al.*, 2016, “Stance detection with bidirectional conditional encoding”, arXiv preprint, https://arxiv.org/pdf/1606.05464.pdf.

[4] J. Davison, J. Feldman and A. M. Rush, 2019, “Commonsense knowledge mining from pretrained models”, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing* (*EMNLP-IJCNLP*), pp. 1173–1178, https://arxiv.org/abs/1909.00505.

[5] J. Ebrahimi, D. Dou and D. Lowd, 2016, “Weakly supervised tweet stance classification by relational bootstrapping”, *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 1012–1017, http://ix.cs.uoregon.edu/~dou/research/papers/emnlp16.pdf.

[6] A. Faulkner, 2014, “Automated classification of stance in student essays: An approach using stance target information and the wikipedia link-based measure”, *Proceedings of the The 27th International Flairs Conference*.

[7] K. Hambardzumyan, H. Khachatrian and J. May, 2021, “Warp: word-level adversarial reprogramming”, arXiv preprint, http://arxiv.org/abs/2101.00121.

[8] A. Hanselowski, A. Pvs, B. Schiller *et al.*, 2018, “A retrospective analysis of the fake news challenge stance detection task”, arXiv preprint, http://arxiv.org/abs/1806.05180.

[9] K. S. Hasan and V. Ng, 2013, "Stance classification of ideological debates: data, models, features, and constraints", *Proceedings of the 6th International Joint Conference on Natural Language Processing*, pp. 1348–1356, http://aye.comp.nus.edu.sg/~antho/I/I13/I13-1191.pdf.

[10] T. Hercig, P. Krejzl, B. Hourová *et al.*, 2017, "Detecting stance in czech news commentaries", *ITAT*, **1885**: 176–180.

[11] S. Hu, N. Ding and H. Wang, 2021, "Knowledgeable prompt-tuning: incorporating knowledge into prompt verbalizer for text classification", arXiv preprint, http://arxiv.org/abs/2108.02035.

[12] P. Liu, W. Yuan, J. Fu *et al.*, 2021, "Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing", arXiv preprint, https://arxiv.org/abs/2107.13586.

[13] X. Liu, Y. Zheng, Z. Du *et al.*, 2021, "Gpt understands, too", arXiv preprint, https://arxiv.org/abs/2103.10385v1.

[14] S. Mohammad, S. Kiritchenko, P. Sobhani *et al.*, 2016, "Semeval-2016 task 6: detecting stance in tweets", *Proceedings of the 10th international workshop on semantic evaluation* (*SemEval-2016*), pp. 31–41, aclanthology.org/S16-1003.pdf.

[15] F. Petroni, T. Rocktäschel, P. Lewis *et al.*, 2019, "Language models as knowledge bases?", arXiv preprint, https://arxiv.org/abs/1909.01066.

[16] A. Radford, K. Narasimhan, T. Salimans *et al.*, 2018, "Improving language understanding by generative pre-training", https://www.cs.ubc.ca/~amuham01/LING530/papers/radford2018improving.pdf.

[17] T. Schick, H. Schmid and H. Schütze, 2020, "Automatically identifying words that can serve as labels for few-shot text classification", arXiv preprint, https://arxiv.org/abs/2010.13641v1.

[18] G. G. Shenoy, E. H. Dsouza and S. Kübler, 2017, "Performing stance detection on twitter data using computational linguistics techniques", arXiv preprint, https://arxiv.org/ftp/arxiv/papers/1703/1703.02019.pdf.

[19] P. Sobhani, D. Inkpen and X. Zhu, 2017, "A dataset for multi-target stance detection", *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Short Papers*, **Vol. 2**, pp. 551–557, https://aclanthology.org/E17-2088.pdf.

[20] S. Somasundaran and J. Wiebe, 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*, pp. 116–124, https://dl.acm.org/doi/10.5555/1860631.1860645.

[21] D. Sridhar, L. Getoor and M. Walker, 2014, "Collective stance classification of posts in online debate forums", *Proceedings of the Joint Workshop on Social Dynamics and Personal Attributes in Social Media*, pp. 109–117, https://aclanthology.org/W14-2715.pdf.

[22] M. Stede and J. Schneider, 2018, "Argumentation mining", *Synthesis Lectures on Human Language Technologies*, **11(2)**: 1–191.

[23] M. Thomas, B. Pang and L. Lee, 2006, “Get out the vote: Determining support or opposition from congressional floor-debate transcripts”, arXiv preprint, https://arxiv.org/pdf/cs/0607062.pdf.

[24] F. H. van Eemeren and R. Grootendorst, 2004, *A systematic theory of argumentation: The pragma-dialectical approach*, Cambridge: Cambridge University Press.

[25] A. Vaswani, N. Shazeer, N. Parmar *et al.*, 2017, “Attention is all you need”, *Proceedings of the 31st Conference on Neural Information Processing Systems* (*NIPS 2017*), https://arxiv.org/pdf/1706.03762.pdf.

[26] M. A. Walker, J. E. F. Tree, P. Anand *et al.*, 2012, “A corpus for research on deliberation and debate”, *Proceedings of the 8th International Conference on Language Resources and Evaluation*, pp. 812–817.

[27] A. Wang, Y. Pruksachatkun, N. Nangia *et al.*, 2019, “Superglue: A stickier benchmark for general-purpose language understanding systems”, *Proceedings of the 33rd Conference on Neural Information Processing Systems* (N*eurIPS 2019*), https://arxiv.org/pdf/1905.00537.pdf.

[28] Y. Wei, T. Mo, Y. Jiang *et al.*, 2022, “Eliciting knowledge from pretrained language models for prototypical prompt verbalizer”, in E. Pimenidis, P. Angelov *et al.* (eds.), *Artificial Neural Networksand Machine Learning–ICANN 2022*, pp. 222–233, Springer.

[29] T. Yue, S. Zhang, L. Yang *et al.*, 2019, “Stance detection method based on two-stage attention mechanism”, *Journal of Guangxi Normal University (Natural Science Edition)*, **37(1)**: 12–49.

[30] A. Zubiaga, A. Aker, K. Bontcheva *et al.*, 2018, “Detection and resolution of rumours in social media: A survey”, *ACM Computing Surveys* (*CSUR*), **51(2)**: 1–36.

[31] 奠雨洁、金琴、吴慧敏，“基于多文本特征融合的中文微博的立场检测”，计算机工程与应用，2017 年第 21 期，第 77–84 页。

[32] 李洋、孙宇晴、景维鹏，“文本立场检测综述”，计算机研究与发展，2021 年第 11 期，第 20 页。

[33] 刘玮、彭鑫、李超等，“立场分析研究综述”，中文信息学报，2020 年第 12 期，第 1–8 页。

[34] 翁沫豪，中文论辩搜索引擎的设计与实现，2021 年，中山大学硕士论文。

[35] 赵姝颖、肖宁、曾华圣等，“基于 RoBerta 的立场检测与趋势预测模型设计”，应用科技，2021 年第 3 期，第 27–33 页。

[36] 赵圆丽、梁志剑，“基于异核卷积双注意机制的立场检测研究”，计算机工程与应用，2021 年第 8 期，第 119–125 页。

（责任编辑：执子）

Argumentative Text Stance Detection

— A Small Sample Study Based on Prompt Model

Bo Xianyu　Weixin Huang

Abstract

Stance detection aims at detecting the position tendency of a text with respect to the views expressed on a specific topic. This paper adopts current prompt-based learning techniques in the field of natural language processing and proposes novel design methods among the two main engineering approaches for prompt learning (template engineering and expression engineering): the masked position-oriented manual template MPOT and the semantic similarity-weighted expression SSWV. Accordingly, the paper proposes a prompt-based stance detection model $P-RoBERTa_{\mathrm{MPOT}+\mathrm{SSWV}}$, and compares it with the automatic template method P-tuning. The model achieves excellent accuracy on the NLPCC Chinese microblog stance detection dataset and the online textual argument corpus of Sun Yat-sen University. The experiments in this paper show that the $P-RoBERTa_{\mathrm{MPOT}+\mathrm{SSWV}}$ model outperforms the model using the pre-training + fine-tuning approach under small sample learning conditions. The paper shows that the prompt learning-based model design approach is useful for text stance detection tasks and achieves very impressive performance even under small sample learning conditions.

Bo Xianyu　Institute of Logic and Cognition, Sun Yat-sen University
Department of Philosophy, Sun Yat-sen University
xianyubo@mail.sysu.edu.cn

Weixin Huang　Institute of Logic and Cognition, Sun Yat-sen University
Department of Philosophy, Sun Yat-sen University
562718834@qq.com