

一种基于提示学习的话题讽刺识别方法

胡松华,唐家伟,马汝虎

合肥大学 人工智能与大数据学院,合肥 230601

摘要:目的 面向话题的讽刺识别旨在通过引入话题,将话题视作对象来判断评论是否为讽刺表达,但是现有基于提示学习的方法存在模板依赖人工构造的问题,存在复杂性与不稳定性。方法 针对此问题,根据提示学习的模板工程与映射器工程提出了新的设计方法,将连续与离散相结合的混合模板与映射器进行交互,通过对模板进行预训练减少人工选取模板引起的复杂性与不稳定性问题,再将设计好的模板与原始输入拼接输入到预训练模型 BERT-base-Chinese 中对标记的数据进行分类;在此基础上,提出面向话题讽刺识别任务的模型 MTPrompt。结果 该模型在 ToScarcasm 数据集上的实验表明,在全批量训练数据场景下,与现有手工模板与手工映射器交互模型 TOSPrompt 相比较,准确率提升了 2.78%、精确率提升了 1.89%、召回率提升了 5.26%、 F_1 值提升了 3.54%,同时在小批量训练数据场景下,准确率与 F_1 值均高于其他基线模型。结论 在话题讽刺识别应用场景中,混合模板的方法可以有效优化人工构造模板方法的复杂性与不稳定性,并且达到更佳识别性能。

关键词:自然语言处理;提示学习;面向话题的讽刺识别;情感分析

中图分类号:TP391.1;TP18 文献标识码:A doi:10.16055/j.issn.1672-058X.2026.0003.013

A Topic-Oriented Sarcasm Detection Method Based on Prompt Learning

HU Songhua, TANG Jiawei, MA Ruhu

School of Artificial Intelligence and Big Data, Hefei University, Hefei 230601, China

Abstract: Objective Topic-oriented sarcasm detection aims to judge whether a comment is a sarcastic expression by introducing topics and regarding them as objects. Nevertheless, the existing methods based on prompt learning have the problem of relying on manually constructed templates, which brings complexity and instability. **Methods** To tackle this problem, a novel design approach was put forward based on template engineering and mapper engineering of prompt learning. A hybrid template integrating continuous and discrete elements interacted with the mapper. By pre-training the template, the complexity and instability resulting from manual template selection were mitigated. Subsequently, the well-designed template was concatenated with the original input and fed into the pre-trained model BERT-base-Chinese for classifying the labeled data. On this basis, a model named MTPrompt for topic-oriented sarcasm detection was proposed.

Results Experiments on the ToScarcasm dataset show that, under the scenario of full-batch training data, when compared with the existing model TOSPrompt that involves the interaction between manual templates and manual mappers, the proposed model, MTPrompt, achieves significant improvements. Specifically, its accuracy was increased by 2.78%, precision by 1.89%, recall by 5.26%, and the F_1 score by 3.54%. Meanwhile, under the scenario of small-batch

收稿日期:2024-05-13 修回日期:2024-08-10 文章编号:1672-058X(2026)03-0108-08

基金项目:2021 年安徽省高校自然科学重点研究项目资助(KJ2021A0990);合肥学院教学研究项目资助(2021HFUHHKC18, 2022HFUJYZD04)。

作者简介:胡松华(1976—),男,博士,副教授,从事知识图谱、自然语言处理研究。

通信作者:唐家伟(1998—),男,硕士研究生,从事自然语言处理研究。Email:hfuutang@163.com。

引用格式:胡松华,唐家伟,马汝虎.一种基于提示学习的话题讽刺识别方法[J].重庆工商大学学报(自然科学版),2026,43(3):108-115.

HU Songhua, TANG Jiawei, MA Ruhu. A topic-oriented sarcasm detection method based on prompt learning[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2026, 43(3): 108-115.

training data, both the accuracy and F_1 score of this model were higher than those of other baseline models.

Conclusion In the application scenario of topic-oriented sarcasm detection, the hybrid template method can effectively alleviate the complexity and instability of manually constructed templates and attain better detection performance.

Keywords: natural language processing; prompt learning; topic-oriented sarcasm detection; sentiment analysis

讽刺是一种特殊的修辞手法,它通过语言的反转、夸张和隐喻等手段,以一种戏谑而尖刻的方式来表达话题背后的讽刺意图,近些年在心理学、语言学、社会学等学科领域都得到了广泛的关注。传统的文本讽刺检测^[1]主要专注于识别句子中的讽刺,但往往忽略讽刺语句与其对象之间的关系。近年来,由于诸多网络媒体如微博、贴吧论坛等互联网应用的广泛普及,众多用户针对网络话题发表的评论中包含了大量具有讽刺意味的语句表达。并且由于评论的针对性,讽刺语句与话题具有相当高的关联。因此,仅仅从评论语句来识别其中的讽刺意图,不足以精准、全方位地识别出用户在评论的语句中所包含的讽刺内容,给针对此类社交媒体的情感分析、舆论监测任务带来了诸多挑战。

梁斌等^[2]将话题语句引入对评论的讽刺识别任务中,通过关联评论与话题间的语义关系识别讽刺意图,构建出面向话题的讽刺识别任务。此前的研究者们针对传统讽刺识别任务提出一系列基于深度学习的方法,如孙晓等^[3]提出了一种结合卷积神经网络与长短时记忆网络的多特征融合模型,该模型通过深层模型学习深层语义特征和深层结构特征完成讽刺识别。Razali等^[4]提出使用深度学习提取的特征与上下文手工特征相结合来检测推文中的讽刺。但是传统的深度学习方法在面向话题的讽刺识别任务中无法更好地挖掘话题与评论之间的语义关系与事实知识。梁斌等^[2]提出基于手工模板的提示学习^[5](Topic Oriented Sarcasm Prompt Learning, TOSPrompt)模型,该模型通过对讽刺识别任务采取手工构造模板的方式,利用提示学习来获取预训练语言模型中已经学习到的知识与语义关系来理解话题与评论的关联性,从而识别句子是否为讽刺表达。虽然此方法相较于传统的预训练模型叠加分类器等方法取得了更优的效果,但是依然存在手工模板构造带来的效果不稳定性与构造模板工作的复杂性问题。

为了更好地解决上述问题,本文提出一种采用混合模板(Mixed Template)与手工映射器(Manual Verbalizer)相结合的基于提示学习(Prompt Learning)的

模型来完成讽刺识别任务。相较于现有的手工构建模板与映射器模型,本文模型无需人工比较最佳模板,解决了手工模板方差大、不稳定的问题。模型通过将模板转化为可以进行优化的连续向量,针对不同的话题讽刺任务,数据可以自适应地在语义空间中寻找合适的向量来代表模板中的每一个词,充分利用大规模预训练语言模型(Pre-trained Language Model)中所学习到的语义信息。本文模型在开源Tosarcasm讽刺数据集上的实验结果取得了最优性能,同时在小批量样本实验中同样也取得了更优的性能,进一步证明了其在特定话题讽刺检测任务上的实用性。

1 相关工作

1.1 讽刺识别

传统文本讽刺识别任务为句子级的讽刺识别,给定一条文本,识别该文本是否针对上下文表达了讽刺意图。目前,研究讽刺识别的方法主要分为3种类型:一是依赖规则的方法,二是采用统计学的方法,三是利用神经网络和深度学习技术的方法^[1]。早期,Riloff等^[6]通过构建从文本中区分正向动词和负向动词的分类器来识别讽刺;Hiremath等^[7]通过捕获3种数据模式,即语音、文本和时间等面部特征辅助完成讽刺识别。近年来,采用神经网络与深度学习的方法在完成讽刺识别任务方面越来越被广泛应用。Du等^[8]提出一种双通道卷积神经网络,不仅分析目标文本的语义学,还分析其情感上下文用于识别讽刺;Zhang等^[9]采用双向递归神经网络(Bidirectional RNNs)捕捉文本内容与语义信息,从而识别文本中的讽刺内容;Bharti等^[10]将文本和音频特征结合在一起,检测会话数据中的讽刺,弥补了仅靠文本数据的不足;张庆林等^[11]将对抗学习方法应用到讽刺识别当中,通过领域迁移方式优化在训练数据不充足情况下模型识别性能低下的问题。

近年来预训练语言模型(Pre-trained Language Model)在自然语言处理的诸多领域都取得了卓越的成果。在讽刺识别的任务领域当中,Babanejad等^[12]提出结合情感特征信息与上下文特征信息来扩展预训练模

型的架构,在识别讽刺的任务中取得了显著的优势。樊小超等^[13]针对文本讽刺识别中缺乏语境信息和语义表达等问题提出了基于多语义融合的讽刺识别方法,通过利用预训练模型 ELMo(Embeddings from Language Models)融合词性和语义表示,在句子级的讽刺识别任务中获得了较好的性能。

1.2 提示学习

伴随着大规模预训练语言模型(Pre-trained Language Model)的发展,各类预训练语言模型的参数呈指数级增长,对预训练语言模型进行微调(Fine-Tuning)所需的资源成本急剧上升。基于提示学习^[15](Prompt Learning)的研究受到了研究者们广泛的关注。相较于传统针对预训练语言模型进行参数微调的方式,提示学习的核心思想是对下游任务进行调整,使得下游任务的结构、输入等更加贴合已训练好的预训练语言模型,从而提高了预训练语言模型参数的有效性与利用率,以更好地完成任务。Schick 等^[14]提出 PET(Pattern Exploiting Training)方法,将输入样本转换为完形填空类型,可帮助预训练语言模型更好地理解任务,以及应用在全监督与半监督场景。其中 PET 更是详细设计了提示学习中的重要组件 PVP(Pattern-Verbalizer-Pair)。在后续的提示学习研究中,如何构建模板(Template)与映射器(Verbalizer)成为热点之一。Shin 等^[15]提出一种 AUTOPROMPT 方法,基于梯度搜索策略,结合原始任务输入和触发词(Trigger Tokens)集合,生成适用于所有输入的提示模板。Han 等^[16]提出利用启发式规则定义若干子模板,通过组合若干子模板的方式来得到最终模板。Gao 等^[17]提出基于 T5 模型(Text-to-Text Transformer)生成的方法构建模板,在确认相应的模板后,再以基于排序搜索的方法得到映射器。Liu 等^[18]提出使用可训练连续空间内模板嵌入的方法 P-tuning,相较于离散模板,可以避免局部最优的情况,同时在离散和连续模板占位符混合时有助于模板的优化。实验表明,GPT3 模型^[19]运用 P-tuning 方法在小样本(Few-shot)训练场景下取得了非常不错的效果。

1.3 面向话题的讽刺识别任务

由于传统的句子级讽刺识别存在局限性,梁斌等^[2]提出一种面向话题的讽刺识别任务。此任务把评论语句所对应的话题融入任务本身中去,通过将话题

作为评论语句的目标或者背景,评估语句是否具有讽刺的含义。其形式化定义为给定一条输入 $P=(Q,S)$,其中 Q 为话题上下文, S 为目标评论语句,任务旨在从目标语句 S 中挖掘出针对话题 Q 是否存在讽刺意图信息,最终判断目标语句 S 的类别是属于“讽刺”类别还是“非讽刺”类别。

具体例子如表 1 所示。针对同样一个语句“做得好!”,样例 I 中没有指明话题信息,因而无法从上下文语义信息中判断语句是否包含讽刺意图。样例 II 中给出了“某国将公然排放污水”的话题,因此结合话题信息可以得出语句包含强烈的讽刺意图。在样例 III 中给出的“我国反对该国行为,将禁止进口该国水产”话题中,同样可以判断出语句并非讽刺意图。因此在该任务当中,同样的一个语句在面向不同的话题时,可能蕴含了不同的意图,即不一样的讽刺标签。

表 1 面向话题的讽刺识别样例

Table 1 Examples of topic-oriented sarcasm detection

样 例	话 题	语 句	标 签
I	无	做得好!	无
II	某国将公然排放污水	做得好!	讽刺
III	我国反对该国行为, 将禁止进口该国水产	做得好!	非讽刺

从例子中可以发现,面向话题的讽刺识别任务相比传统的句子级讽刺识别任务更加具体,更加完整,更加适合现实应用场景。新闻评论、商品评价、电影评论等诸多现实应用场景均可使用到本任务当中。灵活的运用方式和广泛的使用场景赋予了本任务相当具有前景的研究意义。

2 模型设计与建立

本文提出一种用于识别话题讽刺性的提示学习模型(MTPrompt)。该模型的主要部分包括:提示模板构造、标签词映射器和预训练语言模型。在模板构造阶段,通过将原始话题和评论与模板部分拼接成一个整体,再进行处理得到词嵌入表示并输入到预训练模型当中。在预训练模型阶段,经过注意力机制获取输入中的语义信息从而得到掩码标识符[MASK]的隐藏层向量。在标签词映射阶段,通过将[MASK]的隐藏层向量映射至预先定义的标签词当中,每个标签词对应至相应的类别从而获得预测的结果。模型的整体结构如图 1 所示。

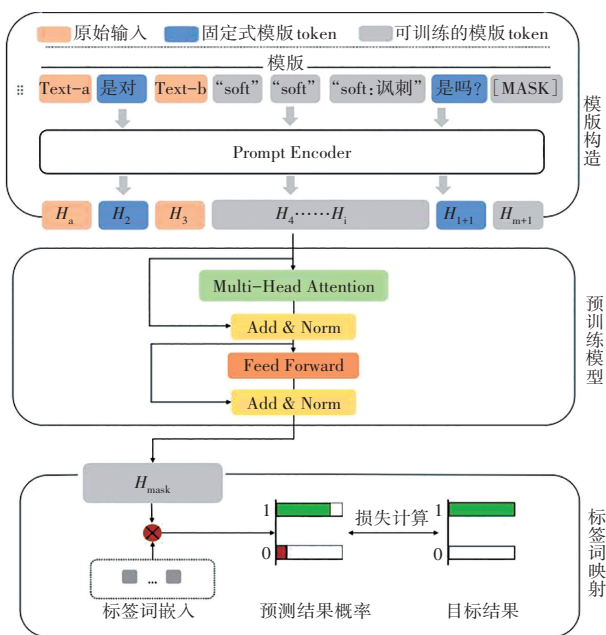


图 1 MTPrompt 模型结构

Fig. 1 MTPrompt model structure

2.1 提示模板构造

在提示学习^[5]方法中,为了充分利用预训练语言模型所包含的丰富语义信息,更精准地完成讽刺识别任务,需要为其构造适宜本任务的模板。在目前的应用中,通常在输入文本句中或者句末添加模板,将任务转化为完形填空式预测任务或者生成式任务。在本任务中,需要模型理解话题语句与评论语句之间的语义关系,因此选择在句中或句末插入模板。

在构造符合任务的表达模板时,本文采用混合模板 (Mixed Template) 形式,这种方法受 Liu 等^[18]在提示学习模型中提出的 P-tuning 方法启发,并针对输入样本进行了调整。此外,借鉴预训练语言模型中遮蔽语言模型^[15](Masked Language Model, MLM)的方法,用以填补

对[MASK]标签位置的空白。其目的在于利用遮蔽语言模型中非遮蔽部分的语义特征为遮蔽区域[MASK]标签进行预测,进而得到符合条件的类别标签。本文构造的混合模板定义如下:

$$X_{prompt} = s \text{ 在对 } t \{ \text{"soft"} \} \{ \text{"soft"} \} \{ \text{"soft"} : \text{"讽刺"} \} \text{是吗? [MASK]} \quad (1)$$

其中,“soft”表示连续提示标记。在该模板中,连续提示标记将以两种形式进行参数初始化:占位符{“soft”}表示此处的占位符参数将随机初始化,占位符{“soft”:“讽刺”}即为占位符添加中文文本标记“讽刺”,参数将由该文本标记初始化。同时,Liu^[18]等发现在离散和连续模板混合时,通过显式地插入领域相关离散占位符,将有助于模板的优化。因此本文中的模板采用随机初始化占位符,由文本标记初始化占位符与离散占位符相结合的方式。在确定好模板的构成后,再输入到预训练语言模型中,对[MASK]位置完成标签词映射,得到最终的识别结果。

2.2 标签词映射器

标签词映射器旨在通过预训练模型预测得到的标签词对应到预设的分类类别当中,可以是一对一或是一对多的映射关系。本文任务定义为二分类任务,即目标文本是否为讽刺语句,故标签词采用手工构建和预设的方式,指定标签词 x 赋值“1”或“0”来分别对应各类别 Y 的赋值“是”或“否”。该映射过程以类别 $Y()$ 来表示,即

$$Y(x) = \begin{cases} \text{是}, & x = 1 \\ \text{否}, & x = 0 \end{cases} \quad (2)$$

通过标签词映射器得到输出后,再通过最小化交叉熵损失对模型中需要训练的参数进行优化,具体流程如图 2 所示。



图 2 MTPrompt 样例流程图

Fig. 2 MTPrompt sample flowchart

3 实验与结果分析

3.1 数据集

目前,权威的讽刺识别中文数据集稀少,本文实验

的数据集采用哈尔滨工业大学研究组开源的面向话题的讽刺识别数据集 Tosarcasm。其中,实验对于样本进行随机划分,训练集、验证机和测试集的数量比例为

6 : 2 : 2, 具体数量统计信息如表 2 所示。

表 2 数据集样本数量统计

Table 2 Statistics of the number of samples in the dataset

标 签	训练集样本	验证集样本	测试集样本
讽刺	1 464	486	486
非讽刺	1 461	487	487
合计	2 925	973	973

3.2 对比模型

在本实验中选取 BERT^[20]、Bi-LSTM^[21]、MIARN^[22]、ADGCN^[23]、PET^[14]、TOSPrompt^[2] 共 6 个模型作为基线模型进行对比实验,以验证本文方法的有效性。

(1) Bi-LSTM。结合前向和后向两个 LSTM 网络,通过对话题和评论所包含的隐式表征进行分别学习,再将隐式表征拼接作为最终分类识别的依据。

(2) MIARN。将注意力机制融入神经网络模型当中,通过注意力机制突出句子中矛盾的“词对关系”,最后分别将前后不一致的隐式表征进行拼接作为最终的表征完成讽刺识别分类任务。

(3) ADGCN。通过结合外部情感知识的图神经网络分析情感的不一致性来识别上下文的讽刺表达。

(4) BERT。基于中文预训练模型 BERT base Chinese,将文本嵌入为词向量,再送入分类器中完成讽刺识别。

(5) PET。基于中文预训练模型 BERT base Chinese,使用“[CLS]s[SEP]t[SEP],这是[MASK]”构建模板,再输入到基于中文训练的预训练模型中,通过对标签词进行映射完成讽刺识别。

(6) TOSPrompt。基于中文预训练模型 BERT base Chinese,使用“s 是对 t 的讽刺吗? [MASK]”构建模板,再输入到预训练模型中,对[MASK]位置进行预测,最终映射为“是”或者“否”完成分类任务。

3.3 参数设置与评估指标

本文实验分为全量训练样本实验和小批量训练样本实验,提出的模型实验参数设置如下:训练轮数选取迭代次数为 6,训练集的批大小设置为 32,学习率设置为 0.000 02,权重衰减系数设置为 0.002,Dropout 数值设置为 0.1,正则化系数设置为 0.000 01,参数优化器采用 Adam 算法。

为了综合评估模型在区分“讽刺”与“非讽刺”表达这一面向话题讽刺识别任务上的性能,本文采用 4 个主要的评估指标:准确率 (Accuracy),其值用 $\varphi_{Accuracy}$ 表示;召回率 (Recall),其值用 φ_{Recall} 表示;精确率

(Precision),其值用 $\varphi_{Precision}$ 表示; F_1 值 (F_1 -score)。本次任务被设计为一个二分类问题,其中模型需要区分的两个类别是“讽刺”(定义为正例,Positive)和“非讽刺”(定义为负例,Negative)。式(3)一式(6)分别描述了这些评估指标的具体计算方法。

$$\varphi_{Accuracy} = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{TN} + N_{FP} + N_{FN}} \quad (3)$$

$$\varphi_{Precision} = \frac{N_{TP}}{N_{TP} + N_{FP}} \quad (4)$$

$$\varphi_{Recall} = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad (5)$$

$$F_1 = \frac{2 \times \varphi_{Precision} \times \varphi_{Recall}}{\varphi_{Precision} + \varphi_{Recall}} \quad (6)$$

识别结果分别定义为 4 种情况:当预测类别和实际类别都是“讽刺”时,这称为真正例 (TP, True Positive),其值用 N_{TP} 表示;如果预测结果标记为“讽刺”,但实际上是“非讽刺”,则这种情况被称为假正例 (FP, False Positive),其值用 N_{FP} 表示;相反,当一个实际为“讽刺”的样本被错误地预测为“非讽刺”,在此称之为假负例 (FN, False Negative),其值用 N_{FN} 表示;最后,当预测结果和实际类别都指示“非讽刺”时,则此情形归类为真负例 (TN, True Negative),其值用 N_{TN} 表示。

3.4 全量训练样本实验结果

为了全面检验本研究所提出的改良型混合模板提示学习模型在讽刺识别任务中的实际效果,本文选择 Tosarcasm 数据集完成实验,并与其他 6 种基准模型进行深入比较,对比实验结果如表 3 所示。通过对这些实验数据的分析,可以发现:本文所提出的改进混合模板提示学习模型在全评价指标上都获得了最好的性能,显示了该模型在目标任务中的有效性。

表 3 实验模型在面向话题的讽刺识别任务中的性能对比

Table 3 Performance comparison of experimental models on topic-oriented sarcasm detection tasks /%

模 型	准确率	精确率	召回率	F_1 值
Bi-LSTM	64.01	61.38	73.25	66.79
MIARN	65.19	63.20	74.82	68.52
ADGCN	66.23	63.56	76.82	69.56
BERT	70.32	67.89	75.15	71.33
PET	70.82	69.72	75.86	72.48
TOSPrompt	70.90	75.87	76.12	75.99
MTPrompt	73.18	77.76	81.38	79.53

从实验结果来看,基于预训练的模型相比基于传统词向量的模型总体上性能更优,在预训练模型的基

础上能更好地完成讽刺识别任务。同时,通过对比普通模板与针对特定任务优化模板几种模型(PET、TOSPrompt、MTPrompt)的实验结果可以看出,在提示学习中不同模板对目标任务的完成结果影响很大,通过对模板的针对性改进可以使模型在目标任务中取得更优的效果。本文提出的改进混合模板在目标任务中相较于此前的提示学习模型,所有评价指标均得到了性能提升,与此前最优的 TOSPrompt 模型比较,准确率提升了 2.78%、精确率提升了 1.89%、召回率提升了 5.26%、 F_1 值提升了 3.54%。

3.5 小批量训练样本实验结果分析

为了检验不同类型模板在小批量训练数据样本中对识别性能产生的影响,本文针对 MTPrompt、TOSPrompt、PET 3 种模型进行了小批量训练样本对比实验。训练数据样本均随机抽取自原始训练数据样本库,比例分别为 20%、40%、60%,其中 3 种模型采用手工标签词映射器,实验结果如图 3 所示。与同类型基于模板构造的模型相比,本文提出的 MTPrompt 模型在小批量训练数据中亦取得了最优的性能。特别是在 20% 的训练数据时, MTPrompt 在 F_1 值评估标准上大幅度领先其他模型,说明本文提出的混合模板在小批量训练数据中能够更准确地识别出“讽刺”类别样例,在缺少训练数据时,混合模板具备更高的潜力挖掘出原始文本中的语义信息,获得相对最佳的预测性能。

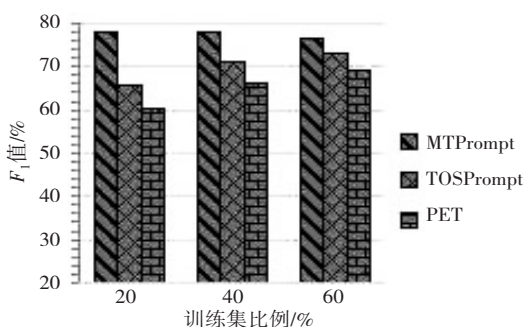
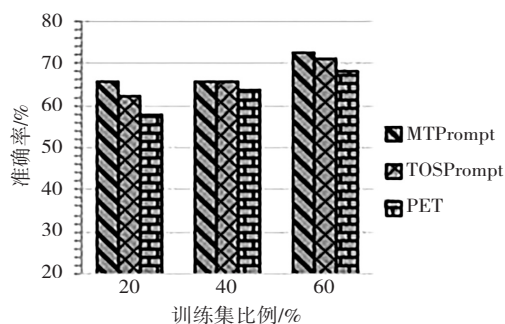


图 3 不同批量训练集的实验结果

Fig. 3 Experimental results for training sets of different batches

3.6 不同类型模板与映射器组合的影响

为了评估不同类型模板和映射器组合的有效性,本文将各类型模板和映射器进行组合搭配实验。其中,模板分为手工模板、软模板、混合模板 3 种。手工模板采用 TOSPrompt 中所设计的模板,软模板使用如表 4 所示预定义位置和长度的占位符模板,混合模板即本文提出的 MTPrompt 模型中使用的模板。所有模板均搭配采用“是”、“否”的二分类手工标签词映射器和软映射器,如表 4 所示。

表 4 不同类型模板示例

Table 4 Examples of different types of templates

类 型	示 例
手工模板	{“text_b”} 是对 {“text_a”} 的讽刺吗? {“mask”} {“text_b”} {“soft”} {“soft”}
软模板	{“text_a”} {“soft”} {“soft”} {“soft”} {“soft”} {“mask”}
混合模板	{“text_b”} 在对 {“text_a”} {“soft”} {“soft”} {“soft”}:“讽刺”是吗? {“mask”}

实验结果如表 5 所示,混合模板和手工映射器在 F_1 值与准确率中均取得了最好的性能,相较于同样采用手工映射器的手工模板准确率提升了 2.28%、 F_1 值提升了 3.54%。这证明在面向话题的讽刺识别任务中,通过对模板改进优化能取得显著的效果。值得一提的是,将同样采用手工映射器的软模板和手工模板相比较,软模板获得了更高的 F_1 值但准确率更低,这说明软模板识别正例的能力较强但是识别负例的能力并不稳定。通过增加文本标记初始化作为锚点,可以进一步提升模型识别负例的能力。

表 5 不同模板与映射器组合的性能

Table 5 Performance of combinations of different

templates and mappers			/%
模 板	映 射 器	F_1 值	准 确 率
手工模板	手工映射器	75.99	70.90
	软映射器	77.64	71.53
软模板	手工映射器	78.74	69.58
	软映射器	78.07	64.03
混合模板	手工映射器	79.53	73.18
	软映射器	76.66	70.40

3.7 错误案例分析

通过表 3 的实验结果可知,虽然本文模型取得了

最优的效果,但是各指标依然有很大的提升空间,反映出本任务的难度性较高。本节将对部分预测错误的案例进行分析,同时将预测错误案例分为以下 3 种类别:

(1) 由于评论使用流行语或别称,此类词语或语句包含原本词语或语句所不具备的语义,导致模型无法准确分析评论所表达的内容。

(2) 由于中文语义本身的复杂性与话题产生的时效性,部分语句本身存在歧义,对数据标注的环节也产生了较高的要求,存在部分误差。

(3) 由于缺乏最新的背景知识或评论与话题的语义关联跳跃性过大,模型无法从语境中判断出评论所包含的讽刺意图。

对于表 6 中类型 i 案例,评论中“真香警告”为网络流行用语,具有讽刺话题中角色前后行为不一致的意味,但是模型无法捕捉语义,这需要预训练语言模型针对最新语料进行更新学习,才能得出正确的结果。类型 ii 案例评论中“没有附上购物链接”,从背景上来看,正值网络购物直播风靡时期,是否具有讽刺意图需要结合评论的背景时效判断,因此在数据标注阶段如何标注难以定性,模型需要结合话题的时效性来判别,但是话题中缺乏此类信息,故存在一定的难度。类型 iii 案例中,话题主题为“GDP 增速”,评论却跳跃至“房贷”,模型较难捕捉二者的语义关联是否存在讽刺意图,此外“房贷”与“GDP 增速”的关系依然需要考虑时效性来辨别二者之间的关联性从而判断讽刺意图,即类型 ii 中话题的时效性问题,因此此类案例对模型的判定性能存在很大的挑战。

表 6 错误案例分析

Table 6 Analysis of error cases

类型	案例
i	话题:新西兰总理阿德恩今天访华,欲就华为 5G 问题与中方对话。 评论:真香警告。
ii	话题:太空探索新时代到来,NASA 授权宇航员首度公开全过程。 评论:没有附上购物链接,还是有些腼腆啊。
iii	话题:曹和平表示双组合促增长,GDP 增速有望回升至 6.6 以上。 评论:房贷呢?

4 结论与展望

本文提出一种基于提示学习和混合提示模板的用于面向话题讽刺任务的模型 MTPrompt,通过设计无须人工比较的混合模板与标签词映射器相结合,充分挖掘语言预训练模型的语义理解能力,从而识别话题评论语义中所包含的讽刺表达。实验结果证明:在全批量训练数据场景下,MTPrompt 与近年来一系列相关研究的基线模型相比,全指标均取得了最佳性能;在小批量训练数据场景下,与同样采用提示学习方法的其他模型相比,在 F_1 值与准确率指标上取得大幅领先,证明了 MTPrompt 在讽刺识别任务当中取得了更好的效果。

改进的模板配合预训练模型虽然获得了更佳的性能,并改善了人工比较模板过程的复杂性,但是在面对部分场景时依然难以准确完成识别任务。在下一步工作中,将考虑通过融入外部知识库,以解决预训练模型在面对最新背景知识下语义信息缺乏的问题;同时,考虑扩充多跳提示模板,解决话题与评论之间跨度过大导致模型无法准确推理其相关性的问题。

参考文献(References):

- [1] 龚晓畅. 中文文本讽刺识别方法研究[D]. 哈尔滨: 哈尔滨工业大学, 2020: 1-19.
GONG Xiao-chang. Research on sarcasm recognition in Chinese text[D]. Harbin: Harbin Institute of Technology, 2020: 1-19.
- [2] 梁斌, 林子杰, 徐睿峰, 等. 面向话题的讽刺识别: 新任务、新数据和新方法[J]. 中文信息学报, 2023, 37(2): 138-147, 157.
LIANG Bin, LIN Zi-jie, XU Rui-feng, et al. Topic-oriented sarcasm detection: new task, new dataset and new method[J]. Journal of Chinese Information Processing, 2023, 37(2): 138-147, 157.
- [3] 孙晓, 何家劲, 任福继. 基于多特征融合的混合神经网络模型讽刺语用判别[J]. 中文信息学报, 2016, 30(6): 215-223.
SUN Xiao, HE Jia-jin, REN Fu-ji. Pragmatic analysis of irony based on hybrid neural network model with multi-feature[J]. Journal of Chinese Information Processing, 2016, 30(6): 215-223.
- [4] RAZALI M S, HALIN A A, YE L, et al. Sarcasm detection using deep learning with contextual features[J]. IEEE Access, 2021, 9: 68609-68618.
- [5] LIU P, YUAN W, FU J, 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.
- [6] RILOFF E, QADIR A, SURVE P, et al. Sarcasm as contrast between a positive sentiment and negative situation[C]// *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. Kerrville: Association for Computational Linguistics, 2013: 704-714.
- [7] HIREMATH B N, PATIL M M. Sarcasm detection using cognitive features of visual data by learning model[J]. *Expert Systems with Applications*, 2021, 184: 115476.
- [8] DU Y, LI T, PATHAN M S, et al. An effective sarcasm detection approach based on sentimental context and individual expression habits[J]. *Cognitive Computation*, 2022, 14(1): 78-90.
- [9] ZHANG M, ZHANG Y, FU G. Tweet sarcasm detection using deep neural network[C]// *Proceedings of International Conference on Computational Linguistics*. Kerrville: Association for Computational Linguistics, 2016: 2449-2460.
- [10] BHARTI S K, GUPTA R K, SHUKLA P K, et al. Multimodal sarcasm detection: A deep learning approach[J]. *Wireless Communications and Mobile Computing*, 2022, 2022: 1653696.
- [11] 张庆林, 杜嘉晨, 徐睿峰. 基于对抗学习的讽刺识别研究[J]. *北京大学学报(自然科学版)*, 2019, 55(1): 29-36.
ZHANG Qing-lin, DU Jia-chen, XU Rui-feng. Sarcasm detection based on adversarial learning[J]. *Acta Scientiarum Naturalium Universitatis Pekinensis*, 2019, 55(1): 29-36.
- [12] BABANEJAD N, DAVOUDI H, AN A, et al. Affective and contextual embedding for sarcasm detection[C]// *Proceedings of the 28th International Conference on Computational Linguistics*. Kerrville: Association for Computational Linguistics, 2020: 225-243.
- [13] 樊小超, 杨亮, 林鸿飞, 等. 基于多语义融合的反讽识别[J]. *中文信息学报*, 2021, 35(6): 103-111.
FAN Xiao-chao, YANG Liang, LIN Hong-fei, et al. Irony recognition based on multiple semantic fusion[J]. *Journal of Chinese Information Processing*, 2021, 35(6): 103-111.
- [14] 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: Main Volume*. Kerrville: Association for Computational Linguistics, 2021: 255-269.
- [15] SHIN T, RAZEGHI Y, LOGAN R L, et al. AutoPrompt: eliciting knowledge from language models with automatically generated prompts [C]// *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*. Kerrville: Association for Computational Linguistics, 2020: 4222-4235.
- [16] HAN X, ZHAO W, DING N, et al. PTR: Prompt tuning with rules for text classification[J]. *AI Open*, 2022, 3: 182-192.
- [17] GAO T, FISCH A, CHEN D. Making Pre-trained language models better few-shot learners[C]// *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*. Kerrville: Association for Computational Linguistics, 2021: 3816-3830.
- [18] LIU X, JI K, FU Y, 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*. Kerrville: Association for Computational Linguistics, 2022: 61-68.
- [19] BROWN T, MANN B, RYDER N, et al. Language models are few-shot learners[J]. *Advances in Neural Information Processing Systems*, 2020, 33: 1877-1901.
- [20] 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*. Kerrville: Association for Computational Linguistics, 2019: 4171-4186.
- [21] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. *Neural Computation*, 1997, 9(8): 1735-1780.
- [22] TAY Y, LUU A T, HUI S C, et al. Reasoning with sarcasm by reading In-between [C]// *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*. Kerrville: Association for Computational Linguistics, 2018: 1010-1020.
- [23] LOU C, LIANG B, GUI L, et al. Affective dependency graph for sarcasm detection[C]// *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York: ACM, 2021: 1844-1849.

责任编辑:李翠薇