

文章编号: 1003-0077(2022)03-0036-09

## 融入主题特征的中越跨语言情感分类模型

施忆雪<sup>1,2</sup>, 余正涛<sup>1,2</sup>, 相艳<sup>1,2</sup>, 张亚飞<sup>1,2</sup>

(1. 昆明理工大学 信息工程与自动化学院, 云南 昆明 650500;

2. 昆明理工大学 云南省人工智能重点实验室, 云南 昆明 650500)

**摘要:** 越南语网络评论的情感分类是越南语事件观点分析的基础。越南语资源匮乏, 标注困难, 可借助中文标注语料进行跨语言情感分类, 实现越南语评论的情感极性预测。但现有的跨语言情感分类模型忽略了主题信息对加强情感表征学习、减小语言差异的作用。为此, 该文提出了一种融入主题特征的中越跨语言情感分类模型。将中文<sup>①</sup>和越南语的主题词分布作为外部知识引入模型, 利用门控机制将主题表征与语义表征进行融合编码, 并通过对抗过程使模型学习到语言分布差异最小的表征, 最终完成情感分类任务。实验结果表明, 该模型能更快拟合出语言分布差异, 其宏  $F_1$  值较多个基线模型均有明显提高。

**关键词:** 跨语言情感分析; 主题模型; 社交媒体评论; 对抗学习

中图分类号: TP391

文献标识码: A

### A Chinese and Vietnamese Cross-lingual Sentiment Classification Model Incorporating Topic Features

SHI Yixue<sup>1,2</sup>, YU Zhengtao<sup>1,2</sup>, XIANG Yan<sup>1,2</sup>, ZHANG Yafei<sup>1,2</sup>

(1. Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, Yunnan 650500, China;

2. Yunnan Key Laboratory of Artificial Intelligence, Kunming University of Science and Technology, Kunming, Yunnan 650500, China)

**Abstract:** Sentiment classification of Vietnamese online comments is the viald for the opinion analysis of Vietnamese event. As Vietnamese is a low-resource language, the cross-lingual sentiment classification can be performed with the help of Chinese annotated corpus to help the sentiment polarity prediction of Vietnamese. In this paper, a cross-lingual sentiment classification model of Chinese and Vietnamese comments incorporating confrontational topic features is proposed. The topic distributions of Chinese and Vietnamese are introduced into the model as external knowledge, and a gate layer is used to encode representation from topic representations with semantic representations. The model is optimized to learn the representations with the smallest differences in language distributions through the adversarial learning to finally complete the sentiment classification task. The experimental results show that the proposed model can significantly improve marco F1 values compared with several baseline models.

**Keywords:** cross-lingual sentiment classification; topic modeling; social media comments; adversarial learning

## 0 引言

互联网的快速发展推动了中越两国交流, 对越

南语的网络评论进行情感极性分析有助于了解民众对特定事件的态度, 是越南语事件观点分析的基础。越南语情感词典、标注数据等资源的稀缺阻碍了其情感分类研究, 可通过跨语言情感分类 (Cross-

收稿日期: 2021-09-15 定稿日期: 2021-10-30

基金项目: 云南省重大科技专项计划项目(202002AD080001); 国家重点研发计划(2018YFC0830105, 2018YFC0830100); 国家自然科学基金(61762056, 61472168, 61972186)

① 本文中的“中文”特指“汉语”。

Lingual Sentiment Classification, CLSC)方法,利用中文作为源语言的情感资源,解决越南语作为目标语言的情感分类问题。目前常用的跨语言情感分类方法有三类:一是基于双语对齐资源的方法,利用平行语料、情感词典等,将源语言中的情感资源直接迁移到目标语言中,弥补语言之间的语义鸿沟<sup>[1-4]</sup>。二是基于机器翻译技术,将两种语言翻译至同一语言,最终将跨语言情感分类转变为单语任务<sup>[5-8]</sup>。三是基于特征迁移<sup>[9-13]</sup>,将源语言与目标语言的特定

语言特征和语言无关特征映射到统一空间中,使得两种语言在语义空间下共享同一表示,以实现情感分类任务。

目前基于特征迁移的方法主要解决的是两种语言的语义表征对齐问题,而未关注文本所蕴含的主题对于情感分类的作用。实际上,无论是中文还是越南语,用户通常都会针对特定主题发表评论,主题信息将有助于推断情感状态<sup>[14]</sup>。以下将以图1和图2所示评论为例,分析主题在评论中的作用。

	评论句	主题	观点表述	情感极性
1	要 <b>相信</b> 科学的力量,我们的疫苗在安全性和时效性上都有 <b>保障</b> 的!	疫苗接种	相信, 保障	积极
2	<b>理解</b> 大家积攒半年的疲惫情绪,但只有 <b>遵守</b> 社交距离和戴口罩,才有希望早日 <b>恢复</b> 正常生活。	疫情防控	理解, 遵守, 恢复	积极

图1 中文评论句样例

	评论句	主题	观点表述	情感极性
1	Bắt đầu một ngày theo cách <b>tốt nhất</b> : Mẹ tôi <b>nhận được</b> liều vắc-xin coronavirus đầu tiên Khoa học trực tiếp (以 <b>最好的</b> 方式开始新的一天: 我妈妈 <b>接种</b> 了她的第一剂冠状病毒疫苗)	(Tiêm phòng) 疫苗接种	<b>tốt nhất</b> (最好的), <b>nhận được</b> (接种)	积极
2	<b>Việc kiểm soát</b> vi rút và phòng chống dịch bệnh phải được <b>tăng cường</b> , và những <b>nỗ lực</b> của con người để chống lại dịch bệnh cuối cùng sẽ mang lại chiến thắng toàn cầu trong cuộc chiến chống lại dịch bệnh. (对于病毒 <b>控制</b> 及防疫更要 <b>加强</b> , 人类为抗击疫情付出的 <b>努力</b> 终会迎来全球的抗疫胜利。)	(Phòng chống dịch) 疫情防控	Việc kiểm soát (控制), <b>tăng cường</b> (加强), <b>nỗ lực</b> (努力)	积极

图2 越南语评论句样例

图1所示的中文新冠疫情数据集中,有关于“疫苗接种”和“疫情防控”两种不同主题的评论句。针对“疫苗接种”主题的评论句中,常用的观点词为“相信”,“最好的”,“保障”等,表示对接种疫苗的支持和鼓励;而针对“疫情防控”主题的评论句中,观点词常用“理解”,“遵守”,“控制”等,表达疫情防控的必要性。可以看出评论在针对某一主题表达观点时,所使用的观点词具有特定性,而不同主题的观点表述存在差异。这一特性也体现在图2所示的越南语新冠疫情数据集上。因此,不论是中文还是越南语,评论句所使用的观点词通常是与特定主题高度相关

的,情感极性也会与主题信息相关联。

根据上述分析,本文认为在进行跨语言情感分类任务中,加入主题信息能够更准确地判断用户表达的情感状态,提高分类性能。因此,本文提出一种融入主题信息的对抗学习模型,实现对越南语评论的情感分类。本文贡献如下:

(1) 提出将中文和越南语的主题词分布作为外部知识引入模型,利用主题信息更好地构建双语评论在同一特征空间下的对应关系。

(2) 提出选择门控机制将主题表征与语义表征进行融合,并利用对抗学习使模型语言分布差异最

小,从而可以利用中文情感极性标签,实现对越南语情感极性的判别。

(3) 在所构建的中越评论数据集和公共的英中数据集上进行了实验,相比已有的基准模型,本文模型的性能都有较大提升。

## 1 相关工作

跨语言情感分析旨在利用源语言资源丰富的情感知识解决目标语言资源稀缺的问题,按策略可分为以下三类方法:

(1) **基于双语对齐资源的方法** 利用双语情感词典等双语资源,实现词粒度信息的对齐,以此判断两种语言的文本是否包含相同情感极性。例如,Wan<sup>[1-2]</sup>提出协同训练的方法,应用双语资源来提升分类器的性能;Balamurali等人<sup>[3]</sup>将多语言知识库中对齐连接词的语义信息作为监督信号,训练情感分类器;Barnes等人<sup>[4]</sup>利用双语词典,将源语言和目标语言的语义相似度及情感信息映射到同一空间中,通过联合学习预测目标语言的情感极性。

(2) **基于机器翻译的方法** 这是跨语言情感分类中常用的方法,该方法主要利用翻译引擎将两种语言翻译至同一语言,最终将跨语言情感分类任务转变为单语情感分类分析任务。例如,文献<sup>[5~8]</sup>利用机器翻译工具获得双语平行语料,建立跨语言文本间语义的对应关系;Meng等人<sup>[8]</sup>通过捕获跨语言的语义相似性构建平行句的共享表示;Li等人<sup>[6]</sup>利用机器翻译工具获取双语间的枢轴特征,并利用单词的分布式表示构建枢轴特征的一对多映射;Zhou等人<sup>[7]</sup>利用机器翻译获取目标语言的标签,并利用语义与情感的相关性完成同一嵌入空间下的双语信息映射。

(3) **基于特征迁移的方法** 其主要思想是将双语中特定语言特征和语言无关特征映射到统一的特征空间中。Chandar等人<sup>[9]</sup>利用平行句对,提出了一种用于学习共享表示的预测性自动编码器;Hermann和Blunsom等人<sup>[10]</sup>提出的方法能学习到一种组合的分布式语义;Zhou等人<sup>[11]</sup>则提出一种双语嵌入,用于跨语言情感分析。上述方法学习的特征难以解释,并且算法时间复杂度高。对此,Chen等人<sup>[12]</sup>提出基于特征编码器和语言判别器的对抗训练模型,利用源语言标签训练情感分类器,最后应用在目标语言上进行情感分类;Chen等人<sup>[13]</sup>

将语言通用的表情符号作为跨语言知识迁移的“桥梁”,提出了基于表情符号的跨语言情感分析模型,提高了目标语言情感分析的性能。

与前期工作不同的是,本文不仅关注跨语言语义差距的问题,同时也关注不同主题下观点表达差异的问题。因此本文提出一种融入主题分布的特征编码器,通过选择门控机制进行特征融合,利用对抗学习实现两种语言特征在特征空间下的对齐。

## 2 模型介绍

融入主题特征的中越跨语言情感分类的任务定义如下:给定中文评论句 $S_s = \{s_1, s_2, \dots, s_{n_s}\}$ ,中文情感标签 $y_s$ ,以及越南语评论句 $S_t = \{s_1, s_2, \dots, s_{n_t}\}$ ,模型的目标是预测中文和越南语评论句的情感极性 $y \in \{0, 1\}$ ,其中, $s_i$ 表示词序列中一个词,0表示消极,1表示积极。

### 2.1 模型整体结构

模型主要分为评论特征编码、对抗学习和情感分类三个部分,具体结构包括语义特征编码器 $F$ 、主题特征编码器 $T$ 、选择门控机制 $G$ 、语言鉴别器 $Q$ 、情感分类器 $P$ ,如图3所示。

首先,利用语义特征编码器 $F$ 对评论文本进行语义表征,表示为 $h_c$ ,利用主题特征编码器 $T$ 进行主题表征,表示为 $h_t$ ,二者通过门控机制 $G$ 进行融合,得到评论的综合表征 $h_f$ ;然后对语言鉴别器 $Q$ 进行迭代的对抗训练,使双语评论综合表征在情感特征空间下得到对齐;最后,利用中文评论的情感极性标签训练情感分类器 $P$ ,用来预测越南语评论的情感极性。

### 2.2 评论综合表征

#### 2.2.1 语义表征

语义表征由语义特征编码器 $F$ 获得,其具体操作为:

(1) 对于给定的评论数据,使用双语词嵌入将其转换为序列嵌入。评论句表示为一个词序列 $S = \{s_1, s_2, \dots, s_n\}$ ,其中 $n$ 表示输入文本对应的词数。单个词嵌入为 $s_i \in \mathbf{R}^d$ ,则输入序列嵌入为 $S \in \mathbf{R}^{d \times n}$ 。

(2) 采用三种不同宽度的核,其窗口大小为 $l_k$ 。通过卷积操作后得到向量 $f_k \in \mathbf{R}^{l \times (n-l_k+1)}$ 。

(3) 采用最大池化操作,获取每个卷积特征中

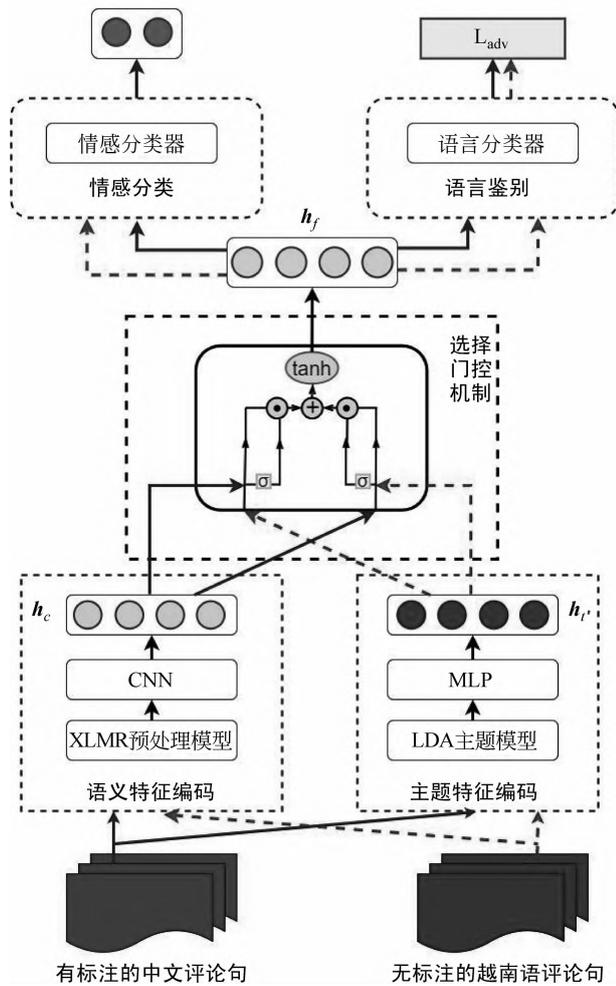


图3 融入主题信息的跨语言情感分类模型

的最大值,经过拼接和线性转换,得到卷积后的语义表征  $h_c \in \mathbf{R}^m$ ,其中  $m$  为线性变换后的隐层维度。

### 2.2.2 主题表征

本文将主题词分布引入到跨语言情感分类任务中,利用主题特征编码器  $T$  进行主题表征,其具体操作为:

(1) 使用 LDA 主题模型预训练源语言和目标语言数据集,得到每个评论在其最大主题概率下的前  $K$  个主题词。

(2) 利用双语词嵌入,得到  $K$  个主题词的嵌入序列  $h_t \in \mathbf{R}^{K \times d}$ ,其中  $d$  是嵌入维度。

(3) 将评论的嵌入序列经过平均池化操作,并通过多层感知机 MLP 进行维度转换,得到其主题向量  $h_t \in \mathbf{R}^m$ , $m$  是变换后的维度。

$$h_t = \text{MLP}(\text{avg\_pooling}(h_t)) \quad (1)$$

### 2.2.3 选择门控机制

在获取到语义表征  $h_c$  和主题表征  $h_t$  后,由于两者的贡献程度并不相同,因此设计一种选择门控

机制  $G$ ,利用一个语义特征门和一个主题门对两个表征进行融合,得到综合表征。门控机制的计算如式(2)~式(4)所示。

$$g_t = \sigma(\mathbf{W}^{(g_t)} h_c + \mathbf{U}^{(g_t)} h_t + \mathbf{b}^{(g_t)}) \quad (2)$$

$$g_c = \sigma(\mathbf{W}^{(g_c)} h_c + \mathbf{U}^{(g_c)} h_t + \mathbf{b}^{(g_c)}) \quad (3)$$

$$h_f = \tanh(h_c \odot g_c + h_t \odot g_t) \quad (4)$$

其中,  $g_t$  是主题门,  $g_c$  是语义特征门,  $\mathbf{W}^{(g_t)}$ ,  $\mathbf{U}^{(g_t)}$ ,  $\mathbf{b}^{(g_t)}$  是可学习的参数,  $h_c$ ,  $h_t$ ,  $h_f$  维度均为  $\mathbf{R}^m$ 。门控机制允许网络自适应地学习语义表征和主题表征的重要性,组成评论的综合表征  $h_f \in \mathbf{R}^m$ 。

### 2.2.4 对抗学习

使用对抗学习对两种语言融入主题信息后的特征进行空间下的对齐。对抗学习过程包括生成器和鉴别器两部分,其中生成器是由评论综合表征中的语义特征编码器  $F$ 、主题特征编码器  $T$  和选择门控机制  $G$  融合构成的,语言鉴别器  $Q$  由多层感知机构成,输出近似为语言分布。本文采用 Chen<sup>[13]</sup> 等人提出的 Wasserstein 距离方法减小源语言和目标语言分布距离,能够用于更稳定的超参数选择。鉴别器  $Q$  的损失函数表示如式(5)所示。

$$\mathcal{L}_{adv} = \min_{\theta_q} (-E_{h_f \sim \text{src}} [Q(h_f)] + E_{h_f \sim \text{tgt}} [Q(h_f)]) \quad (5)$$

其中,  $h_f$  表示源语言的综合表征,  $h_{f'}$  表示目标语言的综合表征,  $\theta_q$  表示语言鉴别器的参数。

### 2.2.5 情感分类

在得到双语对齐特征后,利用源语言情感标签对情感分类器  $P$  进行训练。分类器将源语言综合表征  $h_f$  输入到一个线性层,转换为一个标签向量  $v_s \in \mathbf{R}^e$ ,其中  $e$  表示标签数量。最后将标签向量  $v_s$  送入 softmax 层,得到预测类别  $\hat{y}_s \in \{0, 1\}$ 。

$$\hat{y}_s = \text{softmax}(\mathbf{W}^{(l)} v_s + \mathbf{b}^{(l)}) \quad (6)$$

其中,  $\mathbf{W}^{(l)} \in \mathbf{R}^e$ ,  $\mathbf{b} \in \mathbf{R}^e$  是最后一个线性层的待训练参数。

情感分类器  $P$  的训练损失采用交叉熵损失  $L_p(\hat{y}_s, y_s)$ ,其中,  $y_s$  为源语言的真实标签。

最终模型的整体目标函数由鉴别器损失和情感分类器损失构成如式(7)所示。

$$\mathcal{L}_f = \min_{\theta_f} [L_p(\theta_f) + \lambda \mathcal{L}_{adv}(\theta_f)] \quad (7)$$

其中,  $\lambda$  是平衡情感分类器  $P$  和鉴别器  $Q$  的超参数,  $\theta_f$  表示语义特征编码器  $F$ , 主题特征编码器  $T$  和选择门控机制  $G$  的参数,由  $\theta_f$  参数化的  $F$ 、 $T$  和  $Q$  力求最小化鉴别器损失和情感分类器损失。

### 2.2.6 模型训练

之前的研究发现,对抗训练时生成器和鉴别器的训练可能不完全同步<sup>[15]</sup>,因此训练时先对鉴别器进行迭代训练,用来协调生成器和鉴别器的训练效果。训练鉴别器时利用梯度反转层<sup>[16]</sup>将其与生成器相连接,在反向传播的过程中利用超参数 $\lambda$ 平衡 $P$ 和 $Q$ 对 $F$ 、 $G$ 和 $T$ 的影响,让整个网络使用标准的反向传播进行完整训练。因此首先训练鉴别器 $Q$ ,再对 $F$ 、 $T$ 、 $G$ 和 $P$ 进行训练,训练的完整伪代码如算法1所示。

#### 算法1 训练过程

输入: 有标注的源语言语料  $X_{src}$ ; 无标注的目标语言语料  $X_{tgt}$ ;  
超参数: 设置平衡参数  $\lambda > 0$ , 设置鉴别器  $Q$  迭代次数  $k \in N$ ;  
输出: 源语言评论句和目标语言评论句的情感倾向  $\hat{y}_s$ ;  
1: for epoch in Epoch do  
2: 鉴别器  $Q$  训练迭代过程:  
3: for qiter = 1 to k do  
4: 采样无标注的源语言批次数据  $x_{src} \sim X_{src}$   
5:  $h_{f^{src}} = G(F(x_{src}), T(x_{src}))$   
6:  $h_{f^{tgt}} = G(F(x_{tgt}), T(x_{tgt}))$ , 得到源语言和目标语言语料的特征向量  
7:  $loss_q = -Q(h_{f^{src}}) + Q(h_{f^{tgt}})$ , 通过更新  $Q$  的参数来减小  $loss_q$   
8: end  
9: 采样有标注的源语言批次数据  $(x_{src}, y_s) \sim X_{src}$   
10: 采样无标注的目标语言语料批次数据  $x_{tgt} \sim X_{tgt}$   
11:  $h_{f^{src}} = G(F(x_{src}), T(x_{src}))$   
12:  $h_{f^{tgt}} = G(F(x_{tgt}), T(x_{tgt}))$ , 得到源语言和目标语言语料的特征向量  
13:  $\hat{y}_s = p(h_{f^{src}})$ , 得到源语言的情感预测标签  
14:  $loss = L_p(\hat{y}_s, y_s) + \lambda loss_q$ ,  $y_s$  为源语言的真实情感标签。通过更新  $F$ 、 $T$ 、 $G$  和  $P$  的参数来减小  $loss$   
15: end

## 3 实验设置

### 3.1 数据集

为了证明实验的有效性,本文在中越实验数据集上进行实验,验证情感分类的结果,通过爬虫技术在 twitter<sup>①</sup> 和新浪微博<sup>②</sup> 上爬取新冠疫情相关评论作为实验数据。本文对收集到的中文评论数据集进行划分,选取 5 000 条标注评论用于训练情感分类器,选取 1 000 条标注评论进行验证。对收集到的越南语评论数据集进行划分,选取 4 600 条评论用于训练,选取 1 200 条标注评论用于验证和测试,越南语的训练数据用于对抗学习,不需要进行标注。

数据集的具体信息如表 1 所示。

表 1 中越实验数据集

语言		训练语料 /条	验证语料 /条	测试语料 /条
源语言	中文	5 000	1 000	无
目标语言	越南语	4 600	1 200	1 200

为了验证融入主题信息模型的泛化能力,本文同时在英中公开数据集上进行了实验验证,该实验以英文作为源语言,中文作为目标语言,其中英文数据集由 Datafiniti<sup>③</sup> 1 000 家酒店的列表以及评论,使用评论数据并将评论星级 4、5 映射为情感极性积极,将星级 1、2 映射为情感极性消极。中文数据集使用酒店评论<sup>[17]</sup>。将英文作为源语言,中文作为目标语言,在目标语言上选择相同规模的数据进行验证和测试,该数据集的具体信息如表 2 所示。

表 2 英中实验数据集

语言		训练语料 /条	验证语料 /条	测试语料 /条
源语言	英文	20 000	8 000	无
目标语言	中文	9 600	1 200	1 200

### 3.2 评价指标

与其他情感分类任务类似,本文的实验评价标准使用测试数据集上准确度  $A$  (Accuracy)、精确度  $P$  (Precision)、召回率  $R$  (Recall) 和  $F_1$  值的宏平均值结果作为评价指标,从而衡量模型的性能,如式(8)~式(11)所示。

$$A = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

$$\text{macro}P = \frac{TP/(TP + FP) + TN/(TN + FN)}{2} \quad (9)$$

$$\text{macro}R = \frac{TP/(TP + FN) + TN/(TN + FP)}{2} \quad (10)$$

$$\text{macro}F1 = \frac{2 \times \text{macro}P \times \text{macro}R}{(\text{macro}P + \text{macro}R)} \quad (11)$$

其中,TP 表示正类被正确预测,FP 表示负类被错误预测,FN 表示正类被错误预测,TN 表示负类被

① <https://twitter.com/home>

② <https://weibo.com>

③ <https://datafiniti.co/products/business-data/>

正确预测。

### 3.3 对比模型

本文选取了以下 6 种模型作为基准模型,所有基准模型的训练集、验证集和测试集划分均与本文提出的模型一致。

**LR(Logistic Regression)**: 是监督学习中经典的分类方法,以线性回归为理论支持,通过 Sigmoid()函数引入了非线性因素,解决分类任务。使用源语言中文训练的基线模型,并仅依靠双语词嵌入对目标进行分类。

**CNN(Convolutional Neural Networks)**<sup>[18]</sup>: 采用 textCNN 模型,使用源语言中文训练的基准模型,并仅依靠双语词嵌入对目标进行分类。设置卷积核大小为{3,4,5}。

**LR+MT**: 使用机器翻译的方式,首先将目标语言文本翻译成源语言,再利用逻辑回归模型对目标语言进行情感分类。

**CNN+MT**: 同样使用机器翻译的方式,将目标语言文本翻译成源语言,再利用 CNN 对目标语言进行情感分类。设置卷积核大小为{3,4,5}。

**mSDA**<sup>[19]</sup>: 采用基于 Auto Encoder 的方法,通过非线性降维,使得源域和目标域的数据分布在同一个子空间里面。设置去噪自编码器层数为 3。

**ADAN (Adversarial Deep Averaging Networks)**<sup>[13]</sup>: 使用 CNN 作为特征编码器,利用对抗学习来弥合源语言和目标语言的差异,设置特征编码中使用的卷积核大小为{3,4,5},平衡超参  $\lambda$  为 0.1,对抗迭代次数  $k=5$ 。

### 3.4 参数设置

利用 Adam 优化器对  $F$ 、 $P$  和  $T$ 、 $G$  进行联合优化,用相同学习率的 Adam 优化器训练  $Q$ 。实验使用 xlmr<sup>[20]</sup> 预训练模型获得两种语言的跨语言词嵌入向量,向量维度为 1 024,使用 dropout 防止过拟合。模型训练 30 个批次后,在验证集上选择最佳模型。具体信息如表 3 所示。

表 3 参数设置

参数	值
batch_size	50
dropout	0.5
learning_rate	0.000 5

续表

参数	值
t_learning_rate	0.000 1
q_learning_rate	0.000 5
topic	15

## 4 实验结果分析

### 4.1 基线模型实验结果

表 4 列出了本文模型与基准模型在中越数据集上的对比结果。

表 4 中越跨语言情感分类方法性能对比

(单位: %)

方法	Accuracy	macroP	macroR	macroF1
LR	43.1	42.12	76.07	54.22
CNN	64.5	62.68	39.05	50.80
LR+MT	44.80	43.99	90.06	59.11
CNN+MT	59.5	52.90	78.10	63.08
mSDA	55.7	51.13	50.10	50.22
ADAN	65.1	58.58	72.46	65.78
our_model	<b>66.4</b>	<b>66.34</b>	71.58	<b>68.86</b>

从表 4 实验结果可以看出,本文模型与其他基准模型相比有较大优势,具体分析如下:

(1) 将本文模型与 LR 和 CNN 进行对比,可以看出本文模型结果的 macroF1 提升了大约 18%,原因在于只使用双语词嵌入的方法尚不具备将中文评论中所含有的情感知识转移到越南语评论中的能力。

(2) 对比 LR、CNN 和 LR+MT、CNN+MT 时,也可以看出 LR+MT 的结果要优于 LR, macroF1 提高了 5%,原因是通过机器翻译的方式能够减小语义鸿沟。而在对比本文模型与 LR+MT、CNN+MT 时,可以看出本文模型要优于基于机器翻译的方法,macroF1 值提高了 7%,原因可能是翻译过程中丢失特定语言包含的语义特征。本文采用的对抗学习策略可以在不丢失特定语义特征情况下,实现两种语言情感表征在同一特征空间下的对齐。

(3) mSDA 的结果在所有对比实验中效果最

差,macroF1 仅为 50.22%,推测原因是这种基于词袋特征的方式不适用于跨语言的任务,两种语言具有完全不同的词汇表,使语言间的差异过大。

(4) 分析本文模型与 ADAN 的结果,本文模型的 macroF1 提高了 3.1%。通过分析认为,利用对抗学习能够缩小两种语言之间的语义差距,而对于描述同一事件的两种不同语言评论,本文方法能更好地学习到二者的对应关系,从而可以使分类性能得以提升。

为了验证模型的泛化能力,本文还进行了英中跨语言情感分类实验,结果如表 5 所示。

表 5 英中跨语言情感分析方法性能对比  
(单位: %)

方法	Accuracy	macroP	macroR	macroF1
LR	55.00	52.98	96.70	68.46
CNN	78.08	77.89	79.04	78.46
LR+MT	58.58	56.86	87.96	69.07
CNN+MT	76.5	81.33	71.79	76.26
mSDA	48.84	48.12	49.87	48.98
ADAN	79.42	81.43	76.73	79.01
our_model	<b>82.00</b>	<b>82.28</b>	82.01	<b>82.15</b>

根据表 5 的实验结果,可以验证本文模型在英文和中文数据集上同样适用,这表明本文模型在英中跨语言情感分类上也有较好的表现。

#### 4.2 不同组件对情感分类结果的影响

为了验证本文模型组成结构中主题特征、门控机制和对抗学习三种不同组件对模型最终性能的影响,针对这三个部分进行实验验证。实验结果如表 6 所示。其中,our\_model 表示本文提出的模型;-topic module 表示没有融入主题特征的模型;-gate module 表示模型融入主题特征后,没有利用选择门控机制作为融合策略,而是将语义特征和主

表 6 不同组件对情感分类结果的影响  
(单位: %)

方法	Accuracy	macroP	macroR	macroF1
-topic module	65.1	58.58	72.46	65.78
-gate module	64.5	60.51	69.16	65.90
-adversarial module	63.5	58.55	68.27	64.40
our_model	<b>66.4</b>	<b>66.34</b>	71.58	<b>68.86</b>

题特征直接拼接;-adversarial module 表示模型没有进行对抗训练,直接将所得特征用于训练情感分类器。

实验结果表明,-adversarial module 的性能效果在四个方法中最差,说明在本文提出的模型中对抗训练的重要性,对抗训练模块使双语评论表征在情感特征空间下得到对齐,利用中文情感极性标签,实现对越南语情感极性的判别;再对比融入对抗训练模块后的几组模型,-topic module 的性能要低于-gate module 的性能,说明主题特征融入的有效性,能够丰富双语评论的情感特征;-gate module 模型性能要低于本文提出的模型,说明选择门控机制进行特征融合,能够学习到更好的评论综合表征,最终提高情感分类的结果。

#### 4.3 主题特征对语言鉴别结果的影响

为了验证融入主题特征的有效性,本文针对主题特征部分做进一步的验证,实验结果如图 4 所示。其中,“-topic”、“+topic”分别表示不包含主题特征、包含主题分布特征。

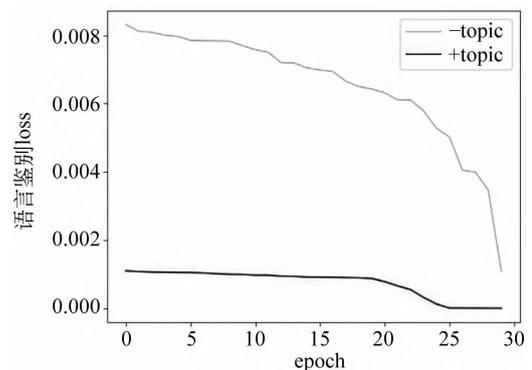


图 4 语言鉴别 loss 对比

语言鉴别器的作用是实现源语言和目标语言的特征在同一空间下的对齐。从“-topic”曲线可以看到,语言鉴别损失在前 25 个批次较大,到第 30 批次时才有所减小。从“+topic”曲线上可以看到,前 25 个批次语言鉴别损失比“-topic”小,并且在 25 到 30 批次时,语言鉴别损失再次减小并趋于零。因此可以得出以下结论:融入主题特征信息,可以更快缩小两种语言评论在特征空间下的距离。

#### 4.4 主题数对情感分类结果的影响

主题数的大小决定了主题划分的粒度。本文针对主题数进行实验,观察不同的主题数对模型在中越数据集上的分类效果影响,如图 5 所示。

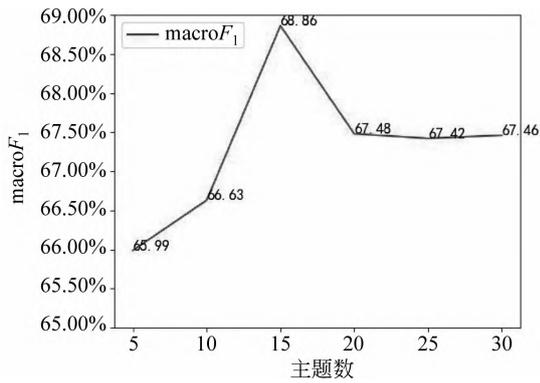


图5 不同主题数目下的性能比较

对折线图进行分析可以看出,当主题数小于15时,模型的 macroF<sub>1</sub> 变化处于上升趋势,在主题数为15时 macroF<sub>1</sub> 值达到最高,而当主题数大于15后,模型的 macroF<sub>1</sub> 值变化趋于稳定。因此根据实验分析的结果可以看出,主题的融入可以使得分类效果得到提升,并且当主题数为15时,模型在中越数据集上的性能最高。

#### 4.5 不同主题表征的性能比较

利用主题模型可以得到评论的主题概率分布和主题词分布两种不同的特征。将式(1)中所使用的主题词嵌入平均池化替换为主题概率分布  $\mathbf{h}_{t1} \in R^N$ , 其中,  $N$  表示主题数,然后通过多层感知机进行线性变换,得到另一种主题表征  $\mathbf{h}_{t'} \in R^m$ , 其中,  $m$  为隐层维度,如式(12)所示。

$$\mathbf{h}_{t'} = \text{MLP}(\mathbf{h}_{t1}) \quad (12)$$

本文测试了两种不同的主题表征方式对于模型性能的影响,实验结果如图6所示。

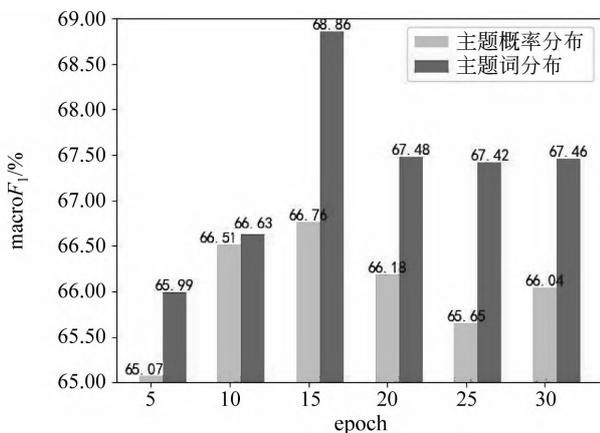


图6 不同主题表征的模型性能比较

对实验结果分析可得,利用主题概率分布和主题词分布均可以提升模型性能。对比两种主题表征

方式,主题词分布的性能在不同主题数目时均优于主题概率分布,表明主题词分布的表征方式能更好解释不同主题的观点表述差异。

## 5 结论

以往的跨语言情感分类工作更多地关注于解决不同语言语义差异的问题,而忽略主题信息对于跨语言表征的作用。针对这一问题,本文提出一种融入主题特征的跨语言情感分类模型,利用门控机制有效地将主题表征与语义表征进行融合,并通过对抗学习将不同语言更好地映射至同一空间。通过实验验证了所提模型在中越跨语言情感分类上的有效性,同时证明主题词作为融入信息能更快拟合出语言分布差异。未来工作中,我们将进一步研究如何利用双语评论之间的对应关系,如评论中的观点对象等辅助跨语言情感分类任务。

## 参考文献

- [1] Wan X. Using bilingual knowledge and ensemble techniques for unsupervised Chinese sentiment analysis [C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing. Honolulu, Hawaii: EMNLP, 2008: 553-561.
- [2] Wan X. Co-training for cross-lingual sentiment classification[C]//Proceedings of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP. Suntec, Singapore: ACL, 2009: 235-243.
- [3] Balamurali A R, Joshi A, Bhattacharyya P. Cross-lingual sentiment analysis for Indian languages using linked wordnets [C]//Proceedings of the COLING 2012 Organizing Committee, 2012: 73-82.
- [4] Barnes J, Klinger R, Walde S S. Bilingual sentiment embeddings: Joint projection of sentiment across languages[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne, Australia: ACL, 2018: 2483-2493.
- [5] Banea C, Mihalcea R, Wiebe J. Multilingual subjectivity: Are more languages better? [C]//Proceedings of the 23rd International Conference on Computational Linguistics. Beijing, China: Coling 2010 Organizing Committee, 2010: 28-36.
- [6] Li N, Zhai S, Zhang Z, et al. Structural correspondence learning for cross-lingual sentiment classification with one-to-many mappings[C]//Proceedings of the

- Thirty-First AAAI Conference on Artificial Intelligence. Palo Alto, California: AAAI Press, 2017; 3490-3496.
- [7] Zhou X, Wan X, Xiao J. Cross-lingual sentiment classification with bilingual document representation learning[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin, Germany: ACL, 2016: 1403-1412.
- [8] Meng X, Wei F, Liu X, et al. Cross-lingual mixture model for sentiment classification[C]//Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics. Jeju Island, Korea: ACL, 2012: 572-581.
- [9] Chandar AP S, Lauly S, Larochelle H, et al. An autoencoder approach to learning bilingual word representations[J]. Advances in Neural Information Processing Systems, 2014, 27(c): 1853-1861.
- [10] Hermann K M, Blunsom P. Multilingual models for compositional distributed semantics[C]//Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. Baltimore, Maryland: ACL, 2014: 58-68.
- [11] Zhou H, Chen L, Shi F, et al. Learning bilingual sentiment word embeddings for cross-language sentiment classification[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. Beijing, China: ACL, 2015: 430-440.
- [12] Chen X, Sun Y, Athiwaratkun B, et al. Adversarial deep averaging networks for cross-lingual sentiment classification[J]. Transactions of the Association for Computational Linguistics, 2018, 6(c): 557-570.
- [13] Chen Z, Shen S, Hu Z, et al. Emoji-powered representation learning for cross-lingual sentiment classification[C]//Proceedings of the World Wide Web Conference. New York, USA: ACM, 2019: 251-262.
- [14] He D, Feng Z, Jin D, et al. Joint identification of network communities and semantics via integrative modeling of network topologies and node contents [C]//Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. Palo Alto, California: AAAI Press, 2017: 116-124.
- [15] Ganin Y, Lempitsky V. Unsupervised domain adaptation by backpropagation[C]//Proceedings of the 32nd International Conference on Machine Learning. Lille, France: PMLR, 2015: 1180-1189.
- [16] Ganin Y, Ustinova E, Ajakan H, et al. Domain-adversarial training of neural networks[J]. The Journal of Machine Learning Research, 2016, 17(1): 2096-2030.
- [17] Lin Y, Lei H, Wu J, et al. An empirical study on sentiment classification of Chinese review using word embedding[C]//Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation. Shanghai, China: ACL, 2015: 258-266.
- [18] Yoon Kim. Convolutional neural networks for sentence classification[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing. Doha, Qatar: ACL, 2014: 1746 - 1751.
- [19] Chen M, Xu Z, Weinberger K Q, et al. Marginalized denoising autoencoders for domain adaptation[C]// Proceedings of the 29th International Conference on International Conference on Machine Learning. Wisconsin, US: ACM, 2012: 1627-1634.
- [20] Conneau A, Khandelwal K, Goyal N, et al. Unsupervised cross-lingual representation learning at scale [C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online: ACL, 2020: 8440-8451.



施忆雪(1997—), 硕士研究生, 主要研究领域为自然语言处理、信息检索。  
E-mail: 1029169383@qq.com



相艳(1979—), 通信作者, 副教授, 博士研究生, 主要研究领域为文本挖掘、情感分析。  
E-mail: 50691012@qq.com



余正涛(1970—), 教授, 博士生导师, 主要研究领域为自然语言处理、信息检索、机器翻译。  
E-mail: ztyu@hotmail.com