

# 跨领域文本的可迁移情绪分析方法

张舒萌<sup>1</sup> 余增<sup>1</sup> 李天瑞<sup>1,2</sup>

1 西南交通大学计算机与人工智能学院 成都 611756

2 综合交通大数据应用技术国家工程实验室 成都 611756

(zzshumeng@163.com)

**摘要** 随着移动互联网的迅猛发展,社交网络平台充斥着大量带有情绪色彩的文本数据,对此类文本中的情绪进行分析研究不仅有助于了解网民的态度和情感,而且对科研机构和政府掌握社会的情绪变化及走向有着重要作用。传统的情感分析主要对情感倾向进行分析,无法精确、多维度地描述出文本的情绪,为了解决这个问题,文中对文本的情绪分析进行研究。首先针对不同领域文本数据集中情绪标签缺乏的问题,提出了一个基于深度学习的可迁移情绪分类的情感分析模型 FMRo-BLA,该模型对通用领域文本进行预训练,然后通过基于参数的迁移学习、特征融合和 FGM 对抗学习,将预训练模型应用于特定领域的下游情感分析任务中,最后在微博的公开数据集上进行对比实验。结果表明,该方法相比于目前性能最好的 RoBERTa 预训练语言模型,在目标领域数据集上 F1 值有 5.93% 的提升,进一步加入迁移学习后 F1 值有 12.38% 的提升。

**关键词:** 情绪分析;深度学习;特征融合;迁移学习

**中图法分类号** TP391

## Transferable Emotion Analysis Method for Cross-domain Text

ZHANG Shu-meng<sup>1</sup>, YU Zeng<sup>1</sup> and LI Tian-rui<sup>1,2</sup>

1 School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, China

2 National Engineering Laboratory of Integrated Transportation Big Data Application Technology, Chengdu 611756, China

**Abstract** With the rapid development of mobile internet, social network platform is full of a large number of text data with emotional color. Mining the emotion in such text not only helps to understand the attitude and emotion of internet users, but also plays an important role in scientific research institutions and the government to grasp the emotional changes and trends of society. Traditional sentiment analysis mainly focuses on the analysis of sentiment tendency, which can not accurately and multi-dimensionally describe the emotion of the text. In order to solve this problem, this paper studies the emotion analysis of the text. Firstly, aiming at the lack of fine-grained sentiment tags in text data sets of different fields, a deep learning based emotion analysis model, FMRo-BLA, is proposed. The model pre-trains the general domain text, and then applies the pre-trained model to the downstream situation of specific domain through parameter based migration learning, feature fusion and FGM Adversarial training. Compared with the best performance of RoBERTa pre-trained language model, the F1 value of the proposed method is improved by 5.93% on the target domain dataset, and it achieves 12.38% by further adding transfer learning.

**Keywords** Emotion analysis, Deep learning, Feature fusion, Transfer learning

## 1 引言

根据以往研究,文本的情感分析主要指用自然语言处理及计算机语言学等技术来对文本所蕴含的主观情感进行分析和挖掘。

近年来,伴随着移动互联网的不断发展,社交媒体逐渐成为人们表达自己情绪和对事物态度的平台,利用情感分析可以从这些交互数据中提取到用户的情绪和观点。而早期的工作主要集中在情感二分类(正向、负向)的研究中,因此,传统

的粗粒度情感分析无法精确、多维度地描述出文本的情感。例如,在实际场景中,“害怕”和“心痛”两个词语虽然都是负向情感,但是表达的情绪是不同的,更准确的情绪分类应该为恐惧和悲伤。情绪分析的研究还处于起步阶段,基于情绪词典和规则的方法或基于机器学习的方法都无法捕捉深层语义的特征,无法准确分析民众复杂的情绪表达,从而推动了情绪分析研究。

基于参数的迁移学习(transfer learning)<sup>[1]</sup>指把已经训练好的模型参数迁移到新的模型中来帮助新模型进行训练。

到稿日期:2021-04-02 返修日期:2021-07-28

基金项目:国家重点研发课题(2020AAA0105101)

This work was supported by the National Key R&D Program of China(2020AAA0105101).

通信作者:李天瑞(trli@swjtu.edu.cn)

迁移学习被应用在多个领域,如图像识别领域<sup>[2]</sup>、文本分类任务<sup>[3]</sup>等。当前不同领域的文本之间存在情感类别分布不同的问题,对于特定领域的文本分析,通用的语料性能有所下降,模型对于领域有一定的依赖性,且情绪词典数据集比较匮乏,对每个领域的语料都进行收集和标注会耗费大量人力物力,因此通过迁移学习来提升通用模型在特定领域文本上的情感分析性能将具有重要的研究意义。本文的主要贡献有:

(1)由于情感丰富的表现方式,预训练模型 RoBERTa (A Robustly Optimized BERT Pretraining Approach) 中不同网络层提取的语言学特征不尽相同,本文将 RoBERTa 模型浅层、中层和最终层的特征同时提取出来,并对每层设置不同的权重因子后进行融合,能充分学习到每个编码层提取到的不同特征,提高模型的可迁移性。

(2)为了解决情绪标签数据缺乏的问题,本文提出了一种 FMRO-BLA (Fast Gradient Method with RoBERTa and Attention-based Bi-directional Long Short-Term Memory) 模型。该模型将预训练模型应用于目标领域的下游情感分析任务,并通过基于参数的迁移学习和 FGM (Fast Gradient Method)<sup>[4]</sup> 对抗学习,提高了模型的泛化性。

(3)在 SMP2020-EWWCT 大赛<sup>1)</sup> 提供的数据集上验证了所提模型的优越性,其 F1 值相比于目前性能较好的 RoBERTa 预训练语言模型有 5.93% 的提升,进一步加入迁移学习后 F1 值有 12.38% 的提升。

## 2 相关工作

由于情绪分析的重要性和在各领域广阔的应用前景,国内外许多学者都对其进行了广泛而深入的研究。最初,许多研究者从语言学角度出发,通过对文本中的词语构建情感词典来进行情绪分析<sup>[5-6]</sup>。除了使用情绪词典外,Aman 等<sup>[7]</sup> 和 Li 等<sup>[8]</sup> 还在词典的基础上,对情感词加入了倾向性和倾向强度来提高准确率。

传统的情绪分类方法主要基于统计手段,没有考虑到引起情绪的触发事件,而通过结合文本语法规则和情绪事件的成因<sup>[9-11]</sup>,可以从事件的结果和对象的性质中分析产生情感的原因事件,并从中挖掘出复杂的情绪关联规则,从而对情感进行情绪分类。基于规则或词典的方法可以在较短时间内获得分类结果,而随着数据量的增大,规则的维护也会更加复杂,不易扩展。在大数据时代的背景下,人们提出了基于机器学习的方法对情绪情感进行分析。MoodLens 模型<sup>[12]</sup> 是一个朴素贝叶斯分类器,该方法首先将 95 个表情符号分成 4 类,分别是快乐、悲伤、厌恶、愤怒,然后利用 350 多万条标记语句作为训练语料进行模型构建。Kang 等<sup>[13]</sup> 从空间角度进行分析,将词汇的矩阵空间通过内积的方式构成句子、篇章等更高级的情感矩阵,最后利用 SVM 进行中文文本的情绪分类。此后,Zhang 等<sup>[14]</sup> 将情绪和社会领域相关知识结合,提出了一种基于因子图算法 (factor graph) 的情绪分类模型,通过观察数据集,总结归纳出了情绪相关性和社会相关性两个影响情绪的主要因素并将其作为特征,该模型

取得了较好的分类效果。

随着海量数据的涌现,数据的采集已不再是难题,标记数据成为有监督学习的瓶颈。针对大量的未标记数据,半监督学习可以用来改善分类器的性能。Sun 等<sup>[15]</sup> 提出了一种半监督的情绪分类方法,首先通过表情符号对未标记数据进行初始化,然后将语句中提取出的一元和二元特征相结合,最后利用 SVM 与朴素贝叶斯分类器进行分类。Sintsova 等<sup>[16]</sup> 通过改进朴素贝叶斯,提出了一种多项贝叶斯分类器进行半监督学习,该方法首先选取文本最突出的标签进行标注,然后利用 n-grams 从文本中抽出特征,并将无关信息过滤,最后通过改进后的平衡伪标记数据和多项贝叶斯分类器对语料进行情绪分类。在此基础上,Suttles 等<sup>[17]</sup> 和 Jiang 等<sup>[18]</sup> 分别提出了一种基于半监督的分类方法,分别是基于离散二进制和表情符号空间的模型。

随着深度学习研究的深入,许多研究者开始将深度学习应用于文本情绪分类中。通过构建多级隐层模型,能够将更深层的句子级特征提取出来,从而提高了文本分类的准确率。Ouyang 等<sup>[19]</sup> 提出了一种基于卷积神经网络 (Convolutional Neural Network, CNN) 的情绪分类架构。该架构利用 Word2vec 方法将词向量从文本中提取出来,使文本中的句子转换为向量,从而能够被 CNN 理解。将其作为 CNN 的输入,通过 3 层卷积层、池化层的神经网络架构对语料进行分类。Santos 等<sup>[20]</sup> 在此基础上提出了一种基于字符到句子的卷积神经网络情绪分类方法 (Character to Sentence Convolutional Neural Network, CharSCNN),该模型通过双卷积层,分别从字符、词语和句子级别的文本中抽取特征。一些研究<sup>[21-22]</sup> 表明,CNN 能够有效捕获并提取文本的局部相关信息。Chen 等<sup>[23]</sup> 利用一种多通道卷积神经网络,通过调整滑动窗口的大小改变输入值,并在池化层后将不同通道提取到的局部特征进行合并,以此来学习句子中多维度的局部信息。此外,循环神经网络 (Recursive Neural Network, RNN) 也被部分研究者引入情绪分类,由于普通的 RNN 缺乏层级表示能力,因此 Irsoy 等<sup>[24]</sup> 提出将 3 个递归层进行叠加,构建了一种深度递归神经网络,并利用非线性的递归信息,对每个节点的贡献值进行计算。Zhu 等<sup>[25]</sup> 使用 LSTM 来解决评论语句情感长距离依赖的问题。此外,Brahma<sup>[26]</sup> 提出了一种基于后级双向 LSTM 模型来对句子建模进行改进。

上述研究均针对特定的领域进行分类。多数跨领域的情感分析研究<sup>[27-28]</sup> 仅进行了跨领域的粗粒度情感分类研究。传统的方法主要包括结构一致性学习<sup>[29]</sup>、谱特征对齐和最大均差值差异等方法<sup>[30]</sup>。近年来,一些深度学习方法<sup>[31]</sup> 也显露出良好的效果,如堆栈降噪自编码器和领域对抗神经网络。在此基础上,Borgwardt 等<sup>[32]</sup> 提出了层次注意力网络,该网络可以通过多层注意力得到更加精准的权重。Peng 等<sup>[33]</sup> 引入一些目标领域的文本数据学习特定领域的信息,能够有效利用特定领域的已标记数据。Yang 等<sup>[34]</sup> 为了充分利用无监督概率生成模型加强表征学习,提出了一个 NAACL 注意力模型,用于跨领域方面的层次情感分类。Zhang 等<sup>[35]</sup> 提出了

<sup>1)</sup> <https://smp2020.aconf.cn/smp.html#4>

一种分层注意力生成对抗网络,该网络能够将单词和句子的上下文信息编码到文档表示中,并在4种数据集上达到平均81.56%的准确率。Ji等<sup>[36]</sup>设计了一个Bifurcated-LSTM网络,该网络首先使用了增强数据集的方法,然后使用一个基于注意力机制的LSTM进行分类,其能够从原领域提取到固定的情感特征,最后在不同的目标领域文本上进行情感分析。

通过对现有研究工作的梳理发现,现有的研究中鲜有针对情绪分类的跨领域迁移方法,因此针对跨领域的情绪分类模型和方法亟待研究。

### 3 一种跨领域的情绪分析方法

本文提出了情绪分类模型FMRo-BLA,并通过迁移学习的方法将此模型应用于特定领域文本。FMRo-BLA模型指首先将RoBERTa模型中的浅层、中层和最终层的特征进行融合作为词向量,然后在此基础上,使用BiLSTM(Bi-directional LSTM)模型并融合注意力机制。该模型的结构图如图1所示。

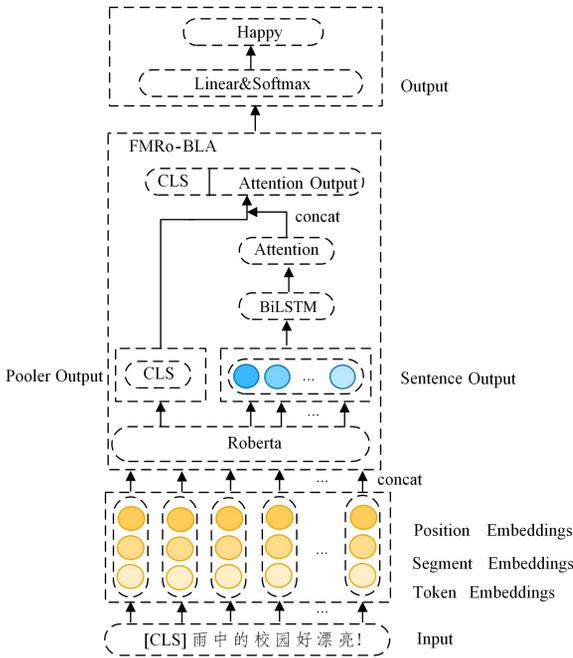


图1 FMRo-BLA的结构图

Fig. 1 Structure diagram of FMRo-BLA

#### 3.1 预训练模型特征融合

考虑到网络模型深层特征具有特异性,只能完成特定任务,而浅层提取到的特征更具有泛化性,为提高模型的可迁移性,将网络层中深层和浅层的特征进行融合。同时,情绪的表现方式具有多样性,有些情绪表达相对内敛,需要理解文本中的具体内容来提取情感信息,而有些直接的表达则可以在原始文本中提取具有感情色彩的词语。

FMRo-BLA模型在提取特征之前,首先需要经过RoBERTa模型的3个嵌入层将输入文本转化为向量表示,分别为字向量(Token Embeddings)、段向量(Segment Embeddings)和位置向量(Positional Embeddings),将这3层嵌入作为RoBERTa的输入。然后采用RoBERTa模型提取文本特征。RoBERTa包含12个隐藏层,每一个隐藏层都包含相应

的self-attention结构,用来得到不同位置的特征。这样可以在模型中同时提取浅层特征和中层特征,与最终层特征进行拼接得到上下文嵌入矩阵后输入到后续网络。

Jawahar等<sup>[37]</sup>通过实验得出,BERT(Bidirectional Encoder Representation from Transformer)的浅层提取的特征是表层信息,中间层为句法信息,深层为语义信息。基于此,为了补充丢失的浅层和中层特征,既能够避免学习到冗余的浅层特征,又能够学习到不同层的特征,本文除了最终输出12层的特征,还将解码器隐藏层中的特征融合部分的第5层、第8层和第10层的特征同时提取出来后进行编码。首先对模型进行动态权重融合,对RoBERTa需要提取的层的编码生成赋予一个权重,权重的初始化如式(1)所示:

$$w_l = \text{Dense}_{\text{init}=1}(\tilde{h}_l) \quad (1)$$

然后通过训练来确定权重值,并将提取出的每一层生成的表示加权平均,再将提取出的层特征从768维通过一层全连接层映射为512维,如式(2)所示:

$$h_i = \text{Dense}_{\text{init}=512}(\sum_{l=s} \tilde{w}_l \text{RoBERTa}(\tilde{h}_l)) \quad (2)$$

其中, $\tilde{h}_l$ 为编码器隐藏层第*l*层状态, $h_i$ 是序列中第*i*个字经过模型的编码向量,*s*为选取的RoBERTa的层序号,为第*l*层的权重因子。完成编码后,将编码序列拼接得到矩阵并输入到下一层中,如式(3)所示:

$$E = [h_1, h_2, \dots, h_n] \quad (3)$$

其中,*n*为输入语句的长度。

最后得到的模型即为经过特征融合的RoBERTa模型,取名为FM-RoBERTa。

#### 3.2 融合注意力机制的高维度特征提取

由于CNN模型会丢失词汇顺序和位置信息,LSTM在文本的时序信息建模方面更具优势,因此本文选取双向LSTM来提取更高维度的特征。将通过预训练模型得到的状态信息输入到更深层的网络模型,并将矩阵E输入到双向LSTM网络,用以提取更深层的语义信息,随后生成带时序信息的输出。该网络包含两个子网络,分别用于左序列上下文和右序列上下文,表述为前向传递和后向传递。第*i*个字的输出如式(4)所示:

$$m_i = [\tilde{m}_i, \bar{m}_i] \quad (4)$$

双向LSTM的输出向量 $[m_1, m_2, \dots, m_n]$ 构成矩阵M。通过全连接层得到矩阵u,如式(5)所示:

$$u_i = \tanh(W^T M + b_w) \quad (5)$$

其中, $M \in \mathbb{R}^{d^w \times T}$ , $d^w$ 是字向量的维度, $w$ 是训练参数向量, $W^T$ 是转置矩阵, $b_w$ 为偏置矩阵。

文本中每个词对于判断文本情感都有不同的贡献,因此在双向LSTM网络后叠加注意力机制,可以对不同文本分别设置权重,用来增强重要词语的权重。通过计算其与上下文向量 $u_w$ 的相似度,然后经过softmax函数进行转换得到概率分布,如式(6)所示:

$$\alpha_i = \frac{\exp(u_i^T u_w)}{\sum_j \exp(u_j^T u_w)} \quad (6)$$

将训练模型与注意力机制的高维度特征进行拼接后,句子的表示r由这些输出向量的加权和构成(见式(7)),最终

送入多层感知机中进行分类。

$$\mathbf{r} = \mathbf{H}\mathbf{\alpha}^T \quad (7)$$

其中,  $\mathbf{\alpha}, \mathbf{r}$  的维度分别是  $T, d^w$ 。

### 3.3 激活函数和损失函数

常见的激活函数有 Sigmoid, Tanh 和 ReLU。Sigmoid 函数变化敏感区间较窄, 在  $(-1, 1)$  之间较为敏感, 当 Sigmoid 函数的输出值接近或超出区间时会处于饱和状态, 影响神经网络的精确度。ReLU 函数容易造成梯度爆炸, 并且有某些神经元可能永远不会被激活, 导致相应的参数存在永远不能被更新的风险。本文 FMRo-BLA 方法选取的激活函数为双曲正切函数, 即 Tanh, 如式(8)所示, 情感特征会随着神经网络的不断循环而更加明显。

$$\tanh(x) = 2\text{Sigmoid}(2x) - 1 \quad (8)$$

本文损失函数选取了传统的多分类交叉熵函数作为基础, 如式(9)所示:

$$L = -\frac{1}{N} \sum_{i=0}^{N-1} \sum_{k=0}^{K-1} y_{\text{true}}^{i,k} \log y_{\text{pred}}^{i,k} \quad (9)$$

其中,  $i$  表示第  $i$  个样本,  $K$  为标签的数量。  $y_{\text{true}}$  表示实际得到的结果,  $y_{\text{pred}}$  表示对每个样本的所有标签的预测值。而当情感倾向不明显时, 损失函数分类结果不稳定, 为此我们引入了一种单位阶跃函数, 如式(10)所示:

$$\theta(x) = \begin{cases} 0, & x < 0 \\ 1/2, & x = 0 \\ 1, & x > 0 \end{cases} \quad (10)$$

将  $\theta(x)$  引入传统的交叉熵函数, 如式(11)所示:

$$L = -\frac{1}{N} \sum_{i=0}^{N-1} \sum_{k=0}^{K-1} y_{\text{true}}^{i,k} \lambda(y_{\text{true}}, y_{\text{pred}}) \log y_{\text{pred}}^{i,k} \quad (11)$$

其中,  $\lambda$  函数表示为:

$$\lambda(y_{\text{true}}, y_{\text{pred}}) = 1 - \theta(y_{\text{true}} - m) \theta(y_{\text{pred}} - m) - \theta(1 - y_{\text{pred}} - m) \theta(1 - y_{\text{true}} - m) \quad (12)$$

### 3.4 FGM 对抗学习

本文使用 FGM 对抗学习的方法, 通过对输入文本嵌入层中的字向量 (token embeddings) 添加扰动来增强模型的鲁棒性, 并提高泛化能力。主要思想为根据梯度生成扰动, 在原本的梯度上累加对抗训练的梯度。

直接对嵌入层的参数矩阵生成干扰样本, 对每个文本中的同一个字嵌入使用相同的扰动, 具体流程为计算每个词的前向损失、反向传播得到的梯度, 然后根据嵌入矩阵的梯度计算出  $\mathbf{r}_{\text{adv}}$  并加入当前的矩阵, 再次计算出对抗梯度进行累加, 最后对参数进行更新, 如式(13)、式(14)所示:

$$\mathbf{r}_{\text{adv}} = \epsilon \cdot \mathbf{g} / \|\mathbf{g}\|_2 \quad (13)$$

$$\mathbf{g} = \nabla_x L(\theta, \mathbf{x}, \mathbf{y}) \quad (14)$$

其中,  $\epsilon$  为 1,  $\mathbf{g}$  为字向量的梯度。

两个数据集在标签上有一致性, 对于通用的情感词, 如“哈哈”, 无论在哪个数据集中, 都更倾向于分类到“积极”标签, 基于这样的客观事实, 可以在训练好的适用范围更广的通用模型的基础上, 加入目标领域的已标注数据对模型进行适应性训练, 相比于在数据条数不足一万的数据集上训练得到的模型, 这种方式大大增加了模型的泛化能力。

## 4 实验验证

为了验证本文所提方法的优越性, 本节设置了多组对比实验, 通过比较本文模型与其他同类模型在源领域及目标领域的迁移情绪分析时的性能来说明本文方法的有效性。

### 4.1 实验数据

本文使用 SMP2020-EWWCT 大赛<sup>1)</sup>提供的数据进行对比实验。数据集的信息如表 1 所列。

表 1 数据集统计信息

Table 1 Dataset statistics

(单位: 条)

Topic	training set	test set
General (source domain)	27768	7000
Epidemic situation (target area)	10606	3000

数据集共分为两部分, 一个是通用领域微博数据, 一个是新冠疫情微博数据。通用微博数据不针对特定领域, 涵盖的内容较广, 疫情微博数据是通过“疫情”“新冠”等疫情期间的关键词获得的相关微博内容。为了使两部分数据集的训练集与测试集维持同样的比例, 本文对数据集重新进行了划分。

### 4.2 参数设置

将 RoBERTa 的每层设置 12 个自注意力, 每个头的大小是 64, 隐层维度均为 768。特征融合部分选取第 5 层、第 8 层、第 10 层和第 12 层的特征进行融合。

对比模型的学习率均设为 0.001, 训练迭代次数设为 20。使用预训练模型将初始学习率设为  $1 \times 10^{-5}$ , 训练迭代次数为 5。

### 4.3 评价指标

本文中使用了精确率、召回率、F1 值 3 个评价指标对模型进行多角度的评价。精确率为分类正确的正样本个数和所有标注为正样本个数的比值, 召回率为标注正确的正样本个数和所有的正样本个数的比值, 精确率和召回率的计算方法分别如式(15)、式(16)所示:

$$P = \frac{TP}{TP + FP} \quad (15)$$

$$R = \frac{TP}{TP + FN} \quad (16)$$

其中,  $TP$  表示标注正确的正样本个数,  $FP$  表示未被标注正确的正样本个数,  $FN$  表示标注错误的负样本个数。精确率和召回率评价模型无法进行平衡, 因此使用二者的调和平均值 F1 值进行评价, F1 值的计算方法如式(17)所示:

$$F1 = \frac{2PR}{P+R} = \frac{2TP}{2TP + FP + FN} \quad (17)$$

### 4.4 对比模型

本文选取的对比模型一方面基于 BERT 及其改进模型, 另一方面基于深度学习的传统文本分类模型。下面介绍本文选取的对比模型。

(1)Text-CNN。利用卷积神经网络提取文本的局部特征, 将进行卷积运算后的特征进行拼接, 并送入分类器进行分类。

<sup>1)</sup> <https://smp2020.aconf.cn/smp.html#4>

(2)RCNN。循环神经网络(RNN)是一种多层的反馈神经网络,每个处理单元之间都具有内部正向和逆向反馈过程。RNN网络与前馈神经网络相比,并处理的数据规模更大,速度更快、更灵活。然而,如果网络中的参数权重过大,则梯度特征相关指数又会成倍地增长,造成无法收敛而发生梯度爆炸的现象。结合RNN和CNN模型,使用循环结构,通过调整滑动窗口的大小,能够捕捉更广泛的上下文信息。

(3)BiLSTM。双向循环神经网络<sup>[38]</sup>由两个LSTM组成,前向后向是两个上下叠加的LSTM网络,并且这两个RNN网络都连接一个输出层,神经网络的输出由两个LSTM共同决定,从而能够提供给输出层输入序列中每一个字全部的上下文信息。

(4)BiLSTM-Attention。将双向长短时记忆网络与注意力机制结合,将文本中不同的词语进行权重区分。

除上述模型外,本文还基于预训练语言模型进行特征提取后进行情绪分析。下面介绍文本所选取的预训练语言模型。

(1)BERT<sup>[39]</sup>。利用BERT预训练模型提取文本的情感特征后,直接通过全连接层输入到多层感知机中进行分类。BERT为一个典型的双向编码模型,建立在transformer之上,拥有强大的特征提取能力和语言表征能力。

(2)RoBERTa<sup>[40]</sup>。这是BERT的优化方案,实质上包括对原始BERT模型以及数据和输入操作进行微调。其引入了动态掩蔽,并通过增强数据进一步提高计算能力。利用RoBERTa预训练模型提取文本的情感特征后,直接通过全连接层输入到多层感知机中进行分类。

(3)FM-RoBERTa。RoBERTa包含12个隐藏层,每一个隐藏层都包含相应的self-attention结构,用来得到不同位置的特征。本文将解码器中第5层、第8层、第10层和最终的输出(12)层的特征取出,进行浅层和深层特征融合,以便于在模型中同时提取浅层特征和深层特征,进行拼接得到上下文嵌入矩阵后输入到后续网络,用于最后的分类。

#### 4.5 结果分析

各模型在源领域数据集上的实验结果如表2所列。

表2 源领域数据集实验结果

Table 2 Experimental results on source domain dataset

(单位:%)			
Algorithm	Precision	Recall	F1
Text-CNN	68.87	64.34	66.53
Text-RCNN	66.36	66.24	66.30
TextRNN-Att	70.00	69.00	69.50
BiLSTM	69.59	71.87	70.71
BiLSTM-Att	74.88	73.06	73.96
BERT	72.65	74.62	73.62
RoBERTa	76.56	72.04	74.23
FM-RoBERTa	77.28	76.68	76.98
FMRo-BLA	80.85	76.13	78.42

从实验结果可以看出,卷积神经网络表现优于循环神经网络,主要原因在于针对微博这类短文本,卷积神经网络相比循环神经网络能够更好地抓取到关键信息。

从BiLSTM-Att比BiLSTM效果更优,TextRNN-Att比Text-RCNN效果更优可以看出,加入注意力机制的模型性能

均有所提升,说明注意力机制突出了重点特征,可以将不重要的特征有区别地过滤。Text-CNN模型是通过使用卷积神经网络配合最大池化来实现的,表现也较好,但相比TextRNN-Att,最大池化只能将该模型认为不重要的特征直接过滤掉。

同时,在对预训练模型进行特征融合后,FM-RoBERTa相比RoBERTa进一步提高了文本情感的分类效果。本文提出的模型在源领域数据集上的F1值相比目前性能较好的RoBERTa预训练语言模型有5.64%的提升。

上述模型与FMRo-BLA模型经过迁移学习后的模型,在目标领域数据集上的结果如表3所列。

表3 目标领域数据集上的实验结果

Table 3 Experimental results on target field dataset

(单位:%)

Algorithm	Precision	Recall	F1
Text-CNN	72.35	56.43	63.41
Text-RCNN	70.35	54.02	61.11
TextRNN-Att	67.55	54.42	60.28
BiLSTM	65.00	49.52	56.21
BiLSTM-Att	69.28	56.39	62.17
BERT	74.65	60.23	66.67
RoBERTa	71.22	64.95	67.94
FM-RoBERTa	70.18	66.05	68.05
FMRo-BLA	72.75	71.20	71.97
FMRo-BLA-FGM	74.23	68.6	71.30
FMRo-BLA-FGM-transferencoder	74.56	78.23	76.35
FMRo-BLA-FGM-transferall	65.15	70.89	67.90

FMRo-BLA-FGM模型为本文模型在文本嵌入层中加入FGM扰动,FMRo-BLA-FGM-transfer模型还在此基础上加入了迁移学习,使用训练好的通用领域模型的编码器参数初始化目标领域的模型。可以看到,加入迁移学习后效果有所提升,这是因为更大规模的通用数据与目标领域的数据存在内在相似性,使得初始化参数能够迁移使用。

FMRo-BLA-FGM-transferall模型尝试使用完整的通用领域模型参数来初始化疫情数据的模型,但性能反而有所下降,说明通用领域的模型的参数对于解析深层语义并不适用。

浅层特征能够提取到更多短语信息,而微博内容的特点就是篇幅短小,结构单一,其中多数情感表达直接,因此本文提出的FMRo-BLA模型在RoBERTa中将浅层特征引入,这对情感的表达非常有利;并且深层特征具有特异性,浅层特征与之相比,提取到的特征更具有泛化性,因此在跨领域文本中会体现出更好的效果。同时,RoBERTa预训练模型对位置信息有一定弱化,其仅通过位置向量来表示词的位置信息,FMRo-BLA模型通过加入BiLSTM来增强词序信息,联合注意力机制增强了模型捕捉情感语义的能力。此外,由于通用数据与目标领域的数据存在内在相似性,并且本文使用的数据集规模较小,在此情况下,参数迁移和FGM对抗学习策略能够进一步增强模型的泛化性。本文模型相比目前性能最好的RoBERTa预训练语言模型,在特定领域的数据集上F1值有5.93%的提升,加入FGM扰动和迁移学习后F1值有12.38%的提升。

**结束语** 如何通过迁移学习对特定领域的文本进行情绪分类是目前情感分类方法面临的关键问题。本文针对这一问题提出了一个可迁移的情绪分类模型——FMRo-BLA。该模型针对不同领域的文本性能良好,体现了良好的可迁移性。其主要思想是首先采用预训练模型 RoBERTa 提取文本特征,并将浅层、中层和最后层的特征分别抽取后进行特征融合。然后通过基于注意力机制的网络对深层语义进行进一步提取,并对文本的情绪进行分类。最后分别在源领域数据集和目标领域相关数据集上进行实验,并通过迁移学习的方法进一步对比。实验结果表明,所构建的 FMRo-BLA 模型在性能上都有一定的提升,FGM 扰动和通用数据模型预训练也有助于跨领域的文本情感分类工作。另外,在数据集较小、数据分布不均衡的情况下,该模型的效果波动较大。如何在这种场景下,维持该模型的鲁棒性是未来研究的方向之一。此外,随着网络平台的发展,除了文字之外,图片和视频也是传达情绪的重要载体,今后的工作可以考虑融合图片、视频等多模态特征,进一步提取更准确的情感信息。

### 参 考 文 献

- [1] REBUFFI S A, BILEN H, VEDALDI A. Learning multiple visual domains with residual adapters[C]// Proceedings of the 31st Neural Information Processing Systems. 2017;506-516.
- [2] YOSINSKI J, CLUNE J, BENGIO Y, et al. How transferable are features in deep neural networks? [J]. Eprint Arxiv, 2014, 27:3320-3328.
- [3] HOWARD J, RUDER S. Universal language model fine-tuning for text classification[J]. arXiv:1801.06146, 2018.
- [4] MIYATO T, DAI A M, GOODFELLOW I. Adversarial training methods for semi-supervised text classification[J]. arXiv:1605.07725, 2016.
- [5] YANG M, ZHU D, CHOW K P. A topic model for building fine-grained domain-specific emotion lexicon[C]// Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). 2014:421-426.
- [6] XU J, XU R, ZHENG Y, et al. Chinese emotion lexicon developing via multi-lingual lexical resources integration[C]// International Conference on Intelligent Text Processing and Computational Linguistics. Berlin: Springer, 2013:174-182.
- [7] AMAN S, SZPAKOWICZ S. Identifying expressions of emotion in text[C]// International Conference on Text, Speech and Dialogue. Berlin: Springer, 2007:196-205.
- [8] LI D, CAO F Y, CAO Y D, et al. Text Sentiment Classification Based on Phrase Patterns[J]. Computer Science, 2008, 35(4):132-134.
- [9] LEE S Y M, CHEN Y, HUANG C R. A text-driven rule-based system for emotion causedetection [C] // Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text. 2010:45-53.
- [10] LI W, XU H. Text-based emotion classification using emotion cause extraction[J]. Expert Systems with Applications, 2014, 41(4):1742-1749.
- [11] GAO K, XU H, WANG J. A rule-based approach to emotion cause detection for Chinese micro-blogs[J]. Expert Systems with Applications, 2015, 42(9):4517-4528.
- [12] ZHAO J, DONG L, WU J, et al. Moodlens: an emoticon-based sentiment analysis system for Chinese tweets[C]// Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2012:1528-1531.
- [13] KANG X, REN F, WU Y. Bottom up: Exploring word emotions for Chinese sentence chief sentiment classification[C]// Proceedings of the 6th International Conference on Natural Language Processing and Knowledge Engineering (NLPKE-2010). IEEE, 2010:1-5.
- [14] ZHANG X, LI W, LU S. Emotion detection in online social network based on multi-label learning[C]// International Conference on Database Systems for Advanced Applications. Cham: Springer, 2017:659-674.
- [15] SUN X, LI C, YE J. Chinese microblogging emotion classification based on support vector machine[C]// Fifth International Conference on Computing, Communications and Networking Technologies (ICCCNT). IEEE, 2014:1-5.
- [16] SINTSOVA V, MUSAT C, PU P. Semi-supervised method for multi-category emotion recognition in tweets[C]// Proceedings of the 14th International Conference on Data Mining Workshop. IEEE, 2014:393-402.
- [17] SUTTLES J, IDE N. Distant supervision for emotion classification with discrete binary values[C]// International Conference on Intelligent Text Processing and Computational Linguistics. Berlin: Springer, 2013:121-136.
- [18] JIANG F, LIU Y Q, LUAN H B, et al. Microblog sentiment analysis with emoticon space model[J]. Journal of Computer Science and Technology, 2015, 30(5):1120-1129.
- [19] OUYANG X, ZHOU P, LI C H, et al. Sentiment analysis using convolutional neural network[C]// Proceedings of the 5th International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomous and Secure Computing; Pervasive Intelligence and Computing. IEEE, 2015:2359-2364.
- [20] SANTOS C D, GATTI M. Deep convolutional neural networks for sentiment analysis of short texts[C]// Proceedings of the 25th International Conference on Computational Linguistics: Technical Papers. 2014:69-78.
- [21] KIM Y. Convolutional Neural Networks for Sentence Classification[J]. arXiv:1408.5882, 2014.
- [22] CHIU J P C, NICHOLS E. Named entity recognition with bidirectional LSTM-CNNs[J]. Transactions of the Association for Computational Linguistics, 2016, 4:357-370.
- [23] CHEN K, LIANG B, KE W D, et al. Chinese Micro-Blog Sentiment Analysis Based on Multi-Channels Convolutional Neural Networks[J]. Journal of Computer Research and Development, 2018, 55(5):945-957.
- [24] IRSOY O, CARDIE C. Deep recursive neural networks for compositionality in language[J]. Advances in Neural Information Processing Systems, 2014, 27:2096-2104.
- [25] ZHU X, SOBIHANI P, GUO H. Long short-term memory over recursive structures[C]// Proceedings of the 32th International

- Conference on Machine Learning. PMLR, 2015:1604-1612.
- [26] BRAHMA S. Improved sentence modeling using suffix bidirectional LSTM[J]. arXiv:1805.07340, 2018.
- [27] DU Y, HE M, WANG L, et al. Wasserstein based transfer network for cross-domain sentiment classification[J]. Knowledge-Based Systems, 2020, 204:106162.
- [28] LIU J, ZHENG S, XU G, et al. Cross-domain sentiment aware word embeddings for review sentiment analysis[J]. International Journal of Machine Learning and Cybernetics, 2020, 12:343-354.
- [29] BLITZER J, DREDZE M, PEREIRA F. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification[C]// Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics. 2007:440-447.
- [30] PAN S J, NI X, SUN J T, et al. Cross-domain sentiment classification via spectral feature alignment[C]// Proceedings of the 19th International Conference on World Wide Web. 2010:751-760.
- [31] DU C, SUN H, WANG J, et al. Adversarial and domain-aware bert for cross-domain sentiment analysis[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020:4019-4028.
- [32] BORGWARDT K M, GRETTON A, RASCH M J, et al. Integrating structured biological data by kernel maximum mean discrepancy[J]. Bioinformatics, 2006, 22(14):49-57.
- [33] PENG M, ZHANG Q, JIANG Y, et al. Cross-domain sentiment classification with target domain specific information[C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 2018:2505-2513.
- [34] YANG M, YIN W, QU Q, et al. Neural attentive network for cross-domain aspect-level sentiment classification[C]// IEEE Transactions on Affective Computing. 2019:761-775.
- [35] ZHANG Y, MIAO D, WANG J. Hierarchical attention generative adversarial networks for cross-domain sentiment classification[J]. arXiv:1903.11334, 2019.
- [36] JI J, LUO C, CHEN X, et al. Cross-domain sentiment classification via a bifurcated-LSTM[C]// Pacific-Asia Conference on Knowledge Discovery and Data Mining. Cham; Springer, 2018:681-693.
- [37] JAWAHAR G, SAGOT B, SEDDAH D. What does BERT learn about the structure of language? [C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019:3651-3657.
- [38] CHEN X, QIU X, ZHU C. Bi-directional Long short-term memory neural networks for chinese word segmentation[C]// Proceedings of the 21st Conference on Empirical Methods in Natural Language Processing. 2015:1197-1206.
- [39] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv:1810.04805, 2018.
- [40] LIU Y, OTT M, GOYAL N, et al. RoBERTa: A robustly optimized bert pretraining approach[J]. arXiv:1907.11692, 2019.



**ZHANG Shu-meng**, born in 1996, post-graduate. Her main research interests include natural language processing and sentiment analysis.



**LI Tian-ruì**, born in 1969, Ph.D, professor, Ph.D supervisor, is a distinguished member of China Computer Federation. His main research interests include big data intelligence, rough sets and granular computing.

(责任编辑:李亚辉)