

## 基于BERT和双通道注意力的文本情感分类模型

谢润忠, 李 烨

(上海理工大学光电信息与计算机工程学院, 上海, 200093)

**摘要:** 对于句子级文本情感分析问题, 目前的深度学习方法未能充分运用情感词、否定词、程度副词等情感语言资源。提出一种基于变换器的双向编码器表征技术(Bidirectional encoder representations from transformers, BERT)和双通道注意力的新模型。基于双向门控循环单元(BiGRU)神经网络的通道负责提取语义特征, 而基于全连接神经网络的通道负责提取情感特征; 同时, 在两个通道中均引入注意力机制以更好地提取关键信息, 并且均采用预训练模型BERT提供词向量, 通过BERT依据上下文语境对词向量的动态调整, 将真实情感语义嵌入到模型; 最后, 通过对双通道的语义特征与情感特征进行融合, 获取最终语义表达。实验结果表明, 相比其他词向量工具, BERT的特征提取能力更强, 而情感信息通道和注意力机制增强了模型捕捉情感语义的能力, 明显提升了情感分类性能, 且在收敛速度和稳定性上更优。

**关键词:** 文本情感分析; 深度学习; 基于变换器的双向编码器表征技术; 双通道; 注意力; 双向门控循环单元

中图分类号: TP391      文献标志码: A

### Text Sentiment Classification Model Based on BERT and Dual Channel Attention

XIE Runzhong, LI Ye

(School of Optical-Electrical & Computer Engineering, University of Shanghai for Science & Technology, Shanghai, 200093, China)

**Abstract:** As for sentence-level emotion analysis, current deep learning methods fail to make full use of emotional language resources such as emotion words, negative words and degree adverbs. A new model is proposed based on bidirectional encoder representations from transformers (BERT) and dual channel attention. One channel based on bi-directional GRU (BiGRU) neural network is responsible for extracting semantic features, while the other based on full connection neural network is responsible for extracting emotional features. At the same time, attention mechanism is introduced into both the channels to better extract key information, and the pre-trained model Bert is used to provide word vectors and thereafter adjust them dynamically according to the context so as to embed real emotional semantic into the model. The final semantic expression is obtained through the fusion of semantic features and emotional features from the two channels. The experimental results show that, compared with other word vector tools, BERT has a better feature extraction ability, while the emotional information channel and the attention mechanism enhance the model's ability to capture emotional semantics, which significantly improves the performance of emotion classification and its convergence speed and stability as well.

**Key words:** text sentiment analysis; deep learning; BERT; dual channel; attention; BiGRU

## 引言

句子级别的文本情感分析,即针对语句的情感倾向性分析,是对带有情感色彩的主观性文本进行分析、处理、归纳和推理的过程。随着论坛、博客和推特等社交媒体的发展,我们拥有了海量的情绪化数据,情感分析技术扮演着越来越重要的角色。

现有的情感分析技术有基于情感词典的方法、基于机器学习的方法、基于深度学习的方法。基于情感词典的方法通过对文本进行词语和句法分析,并计算情感值作为判断文本情感倾向的依据。该方法根据情感词典可以很好地锁定文本情感信息,且实现较为简单。个体在进行语言表达时,会增添必要的情感词汇,情感褒贬词、程度副词、否定词等对情感语义增强或减弱有着重要的促进作用<sup>[1]</sup>。赵妍妍等<sup>[2]</sup>通过构建否定词、副词、情感表情等相关词典来扩充情感词典,大大增强了文本情感极性的判断能力。Keshavarz等<sup>[3]</sup>将语料库与词典相结合,构建自适应情感词典,以改善微博中情感的极性分类。蒋翠清等<sup>[4]</sup>利用初始情感种子词构建了一个面向中文社交媒体文本的领域情感词典,有效地改进了句子级文本的情感分类效果。

基于机器学习的方法主要是将文本的情感倾向分析转换为一个分类问题,然后利用经典的支持向量机、朴素贝叶斯等机器学习算法<sup>[5]</sup>,通过有监督的训练得到一个模型,进而根据该模型进行文本情感倾向分析。然而由于传统的机器学习算法多采用词袋模型来表示文本,其面临数据特征稀疏,且不能很好地抽取文本中蕴涵的情感信息等问题。

近年来兴起的深度学习方法能够较好地弥补传统机器学习方法的缺陷,由卷积神经网络和循环神经网络(Recurrent neural network, RNN)为代表的深度学习模型被广泛地应用于文本情感分析领域。王煜涵等<sup>[6]</sup>采用CNN模型学习文本中的深层语义信息,挖掘Twitter文本的情感倾向,比传统机器学习方法取得了显著提高。Zhu等<sup>[7]</sup>提出了一种统一的CNN-RNN模型,用于视觉情感识别。该体系结构利用CNN的多个层次,在多任务学习框架内提取不同层次的特征,并提出了一种双向RNN来集成CNN模型中不同层次的学习特征,极大地提高了情感分类性能。Zhang等<sup>[8]</sup>提出了一种基于BiGRU的分层多输入输出模型,该模型同时考虑了情感表达的语义信息和词汇信息,实现了对客户评论情感分类的突破性改进。Peng等<sup>[9]</sup>将BiGRU与注意力机制融合应用于细粒度文本情感分析,在不同的数据集上都取得了良好的性能。此后,基于CNN、BiGRU、注意力机制等的混合神经网络模型在文本情感分析任务中得到了广泛应用。

Word2vec<sup>[10]</sup>是目前NLP领域中最常用的词向量工具。Glove模型于2014年由Pennington等<sup>[11]</sup>提出,因其提高了词向量在大语料上的训练速度且稳定性高,在最近几年得以流行。预训练模型(Pre-trained model)是一种基于大型基准数据集训练得到的深度学习架构,在此基础上可以进行后续任务。预训练模型对改进许多NLP任务都有非常大的帮助<sup>[12]</sup>。随着预训练模型研究的深入,ELMo<sup>[13]</sup>、ULMFiT<sup>[14]</sup>、基于变换器的双向编码器表征技术(Bidirectional encoder representations from transformers, BERT)<sup>[15]</sup>等众多NLP预训练模型相继被提出,通过将大型文本语料库作为语言模型进行预训练,为给定句子中的每个单词创建上下文关联的嵌入(Embedding),这些嵌入将被输入到后续任务中。选择高效的词向量表示工具对深度学习的应用研究有着极其重要的影响。

对比传统情感分类方法,情感词、否定词、强度词等情感信息词起着至关重要的作用<sup>[16]</sup>。尽管情感语言词很有用,但在近几年的卷积神经网络(Convolutional neural network, CNN)和长短时记忆网络(Long short-term memory, LSTM)等深度神经网络模型中,情感语言知识的应用还很有限。受文献[15, 17]启发,结合深度学习和情感词典方法的优点,本文提出一种基于BERT和双通道注意力(Dual-channel attention, DCA)的新模型(BERT-DCA)应用于句子级文本情感分析。

## 1 相关技术与BERT-DCA模型

### 1.1 情感信息集合的提取

修饰词词典一般包括否定词、程度副词、连词等部分。当情感词被这些修饰词围绕时,有很大概率伴随着整句的情感极性变化,如极性反转、加强或减弱等。因此,综合考虑情感词和修饰词对判断文本情感极性至关重要。参考Lei等<sup>[18]</sup>构造情感语言库的方法,本文构建的情感语言库主要考虑情感词、否定词和程度副词。通过构建的情感语言库,提取文本句子中包含的情感信息,从而得到每一条文本句子所对应的情感信息集合。设定策略方案如下:

**策略1:**若当前词语为情感词,直接将当前词语加入到情感信息集合中。

**策略2:**若当前词语为程度副词,且下一个词语为情感词,则将二者作为一个整体加入到情感信息集合中;若该情感词已存在于情感信息集合中,则将其删除。

**策略3:**若当前词语为否定词,且下一个词语为情感词,则将二者作为一个整体加入到情感信息集合中;若该情感词已存在于情感信息集合中,则将其删除。或者若否定词依次紧跟副词和情感词,则将三者作为一个整体加入到情感信息集合中;类似的,若副词与情感词的整体存在于情感信息集合中,则将其删除。

### 1.2 BERT预训练模型

Word2vec等传统词向量工具只是简单地提供词嵌入作为特性,相比之下,BERT还可以集成到下游任务中,并作为特定于任务的体系结构进行动态调整。与ELMo使用独立训练的从左到右和从右到左LSTM的级联来生成下游任务的特征不同,BERT使用的是更为强大的双向Transformer编码器<sup>[19]</sup>,如图1所示,并以遮蔽语言建模(Masked language model, MLM)和下一句话预测(Next sentence prediction, NSP)为无监督目标,使模型输出的每个字与词的向量表示都能尽可能全面、准确地刻画输入文本的整体信息,为后续的微调任务提供更好的模型参数初始值;且其输入表征(Input embedding)是通过将相应词的词块嵌入(Token embedding)、段嵌入(Segment embedding)和位置嵌入(Position embedding)求和来构造的,包含了更多的参数,因此具有更强的词向量表示能力。

BERT在大量句子级和Token级任务上获得了最佳性能,优于许多具有任务特定体系结构的系统,包括情感分析领域中的任务<sup>[20]</sup>。

### 1.3 双向GRU神经网络

门控循环单元(Gated recurrent unit, GRU)由Cho等<sup>[21]</sup>提出,是一种对LSTM改进的深度网络模型,其模型结构如图2所示。GRU最大的优点在于

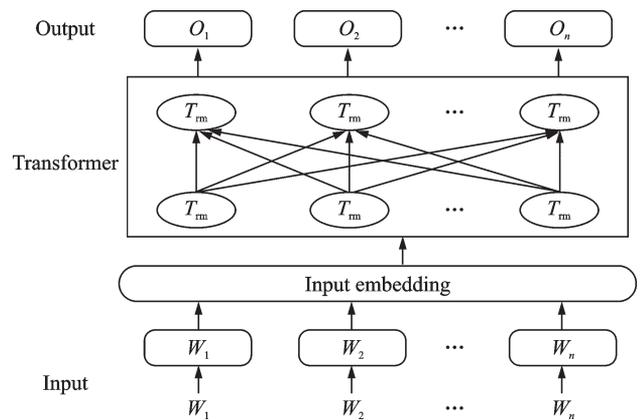


图1 BERT模型结构图

Fig.1 BERT model structure

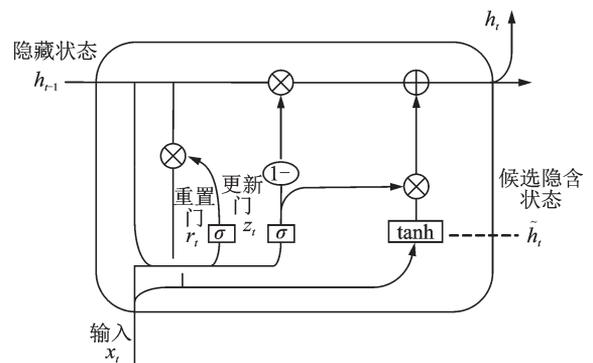


图2 GRU神经元结构图

Fig.2 Structure of a GRU neuron

很好地解决了循环神经网络中的长期依赖问题,且被认为更易于计算和实施。它保留了LSTM对解决梯度消失问题的优点,但内部结构更简单,只有2个控制门:更新门和重置门。GRU神经网络的参数比LSTM减少了1/3,不易产生过拟合,同时由于采取对Cell融合和其他一些改进,在收敛时间和需要的迭代次数上更胜一筹。

GRU神经网络的更新方式如下

$$r_t = \sigma(w_r \cdot [h_{t-1}, x_t]) \tag{1}$$

$$z_t = \sigma(w_z \cdot [h_{t-1}, x_t]) \tag{2}$$

$$\tilde{h}_t = \tanh(w \cdot [r_t * h_{t-1}, x_t]) \tag{3}$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \tag{4}$$

式中: $r_t$ 为 $t$ 时刻的重置门, $z_t$ 为 $t$ 时刻的更新门, $\tilde{h}_t$ 为 $t$ 时刻的候选激活状态, $h_t$ 为 $t$ 时刻的激活状态, $h_{t-1}$ 为 $t-1$ 时刻的隐层状态, $w_r$ 、 $w_z$ 、 $w$ 为相应的权重矩阵, $\sigma$ 为sigmoid激活函数, $\tanh$ 为双曲正切激活函数。更新门由当前状态需要被遗忘的历史信息和接受的新信息决定;重置门由候选状态从历史信息中得到的信息决定。

单向GRU在使用时是从上文向下文推进的,容易导致后面的词比前面的词更重要,而双向GRU(BiGRU)通过增加从后往前传递信息的隐藏层,能更充分利用上下文信息,克服了这一缺陷。BiGRU的模型结构如图3所示。

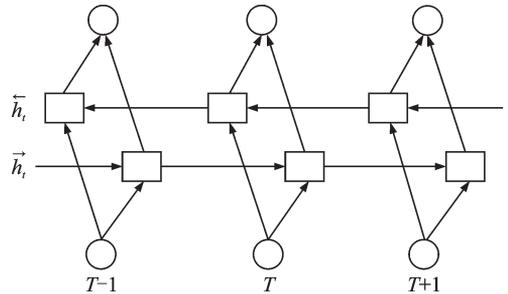


图3 BiGRU结构图

Fig.3 Structure of BiGRU

### 1.4 注意力机制

2014年,Mnih等<sup>[22]</sup>在图像分类任务中首次提出注意力机制,使得结合注意力机制的神经网络成为研究的热点。Bahdanau等<sup>[23]</sup>将注意力机制和RNN结合以解决机器翻译任务,将注意力机制引入到自然语言处理领域。

通过计算概率分布,选择出对当前任务目标更关键的信息,可对深度学习模型起到优化作用,注意力机制在文本情感分析领域中已经得到广泛应用。

### 1.5 BERT-DCA模型

如图4所示,BERT-DCA模型结构包含输入层、信息提取层、特征融合层和输出层4个信息处理层次,在结构上则采用了2个信息处理通道:左侧为语义信息注意力通道(Semantic information attention channel, SAC),右侧为情感信息注意力通道(Emotional information attention channel, EAC)。

#### 1.5.1 输入层

对于一条文本句子序列,经分词后的词语序列 $\{W_1, W_2, \dots, W_n\}$ 作为SAC的输入,通过情感信息集合的提取策略,进而得到由该文本句子提取的情感信息词集合 $\{E_1, E_2, \dots, E_m\}$ 作为EAC的输入;然后利用预训练模型BERT为整个模型提供词向量,能配合上下文语境实现词向量的动态调整,更好地将真实情感语义嵌入模型训练,从而得到语义信息词向量矩阵 $R_x$ 和情感信息词向量矩阵 $R_e$ 。

$$R_x = x_1 \oplus x_2 \oplus \dots \oplus x_n \tag{5}$$

$$R_e = e_1 \oplus e_2 \oplus \dots \oplus e_m \tag{6}$$

式中: $\oplus$ 为行向量连接运算符, $R_x$ 和 $R_e$ 的维数即为评论文本中词语的数目和情感信息词的数目。

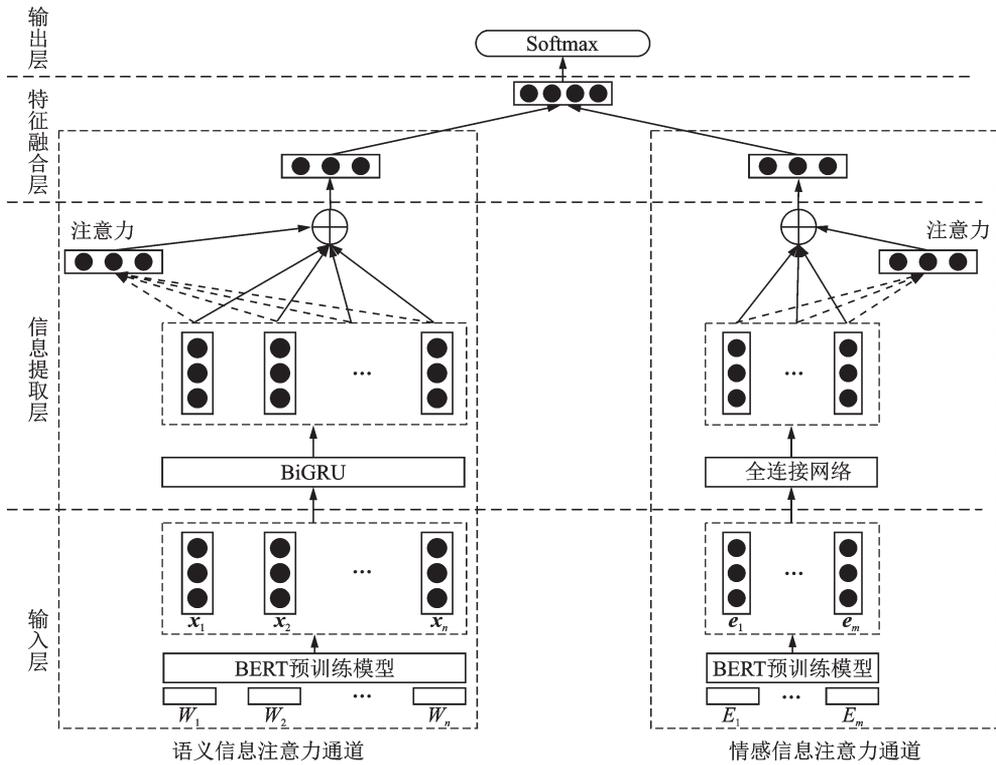


图4 BERT-DCA 模型结构示意图

Fig.4 Structure of BERT-DCA model

1.5.2 信息提取层

对于语义信息文本,首先利用BiGRU神经网络同时处理正向和反向文本序列,对文本深层次的信息进行特征提取,然后利用注意力机制对提取的特征信息分配相应的权重。对于情感信息集合,采用全连接网络与注意力机制相结合的方式对情感信息词进行编码,以获取最重要的情感信号。

某一时刻 \$t\$ 的BiGRU信息提取模块输出状态由正向GRU和反向GRU的输出相连接组成,其计算方法为

$$x_t = W_e w_t, \quad t \in [1, T] \tag{7}$$

$$\bar{h}_t = \overline{\text{GRU}}(x_t), \quad t \in [1, T] \tag{8}$$

$$\bar{h}_t = \overline{\text{GRU}}(x_t), \quad t \in [T, 1] \tag{9}$$

$$s_t = [\bar{h}_t, \bar{h}_t] \tag{10}$$

通过结合 \$\bar{h}\_t\$ 与 \$\bar{h}\_t\$ 得到隐藏层的语义表示 \$s\_t\$,正向的语义信息和反向的语义信息以相同的地位被考虑。

为了捕获更直接的语义依赖关系(使模型在训练时聚焦到数据中的重要信息),将评论文本 BiGRU 模型的输出 \$H^s = \{h\_1^s, h\_2^s, \dots, h\_n^s\}\$ 和情感信息集合全连接网络的输出 \$H^e = \{h\_1^e, h\_2^e, \dots, h\_m^e\}\$ 分别输入到注意力机制中,采用双通道注意力分别对评论文本句子和情感信息进行编码。注意力计算方法如下

$$u_t = \tanh(w_w h_t + b_w) \tag{11}$$

$$\alpha_t = \frac{\exp(\mathbf{u}_t^\top \mathbf{u}_w)}{\sum_t \exp(\mathbf{u}_t^\top \mathbf{u}_w)} \quad (12)$$

$$\mathbf{V} = \sum_t \alpha_t \mathbf{h}_t \quad (13)$$

式中: $\mathbf{w}_w$ 与 $\mathbf{b}_w$ 为注意力机制的可调节权重和偏置项, $\mathbf{h}_t$ 为BiGRU或全连接网络的输出; $\mathbf{u}_t$ 为 $\mathbf{h}_t$ 的隐含状态, $\mathbf{u}_w$ 为Softmax分类器的权重参数, $\alpha_t$ 表示句子中每个词的重要度信息; $\mathbf{V}$ 即为经过注意力模型计算后的特征向量。

### 1.5.3 特征融合层

特征融合层的主要任务是将SAC中生成的特征向量 $\mathbf{V}^s$ 和EAC中生成的特征向量 $\mathbf{V}^e$ 进行合并,从而构建文本整体的情感特征向量。为了简化模型的计算量,采用行连接的方式进行特征融合,构建一个 $(r_s + r_e) \times c$ 的矩阵 $\mathbf{V}^*$ ,生成最终情感特征向量,其中 $r_s$ 和 $r_e$ 分别为 $\mathbf{V}^s$ 和 $\mathbf{V}^e$ 的行数, $c$ 为 $\mathbf{V}^s$ 和 $\mathbf{V}^e$ 的列数。

### 1.5.4 输出层

将特征融合层生成的情感特征向量 $\mathbf{V}^*$ 输入Softmax分类器,从而得到模型最终预测的情感分类结果

$$p = \text{softmax}(\mathbf{w}_o \mathbf{V}^* + \mathbf{b}_o) \quad (14)$$

式中: $\mathbf{w}_o$ 为权重系数矩阵, $\mathbf{b}_o$ 为偏置矩阵, $p$ 为输出的预测情感标签。

### 1.5.5 模型训练

情感分析模型的训练采用端到端的反向传播方式,最小化所有文本中已知情感类别和预测情感类别的交叉熵。

$$f_{\text{loss}} = - \sum_{i=1}^D \sum_{j=1}^C y_i^j \log \hat{y}_i^j + \lambda \|\theta\|^2 \quad (15)$$

式中: $D$ 为训练数据集, $C$ 为情感标签的类别数, $y$ 为实际情感类别, $\hat{y}$ 为预测情感类别, $\lambda$ 为 $L_2$ 正则化, $\theta$ 为设置的参数。

## 2 实验与分析

### 2.1 实验设置

选取2个语料库进行实验:

**实验数据1** 采用流行的中文情感挖掘酒店评论语料ChnSentiCorp,该语料规模为10 000篇。为了方便起见,将语料整理为4个子集。选择ChnSentiCorp-Htl-ba-6000的数据进行实验,该语料为平衡语料,包含正负类各3 000篇。

**实验数据2** 利用爬虫工具Pyspider在豆瓣电影平台上抓取了近30部热播电影用户评论及评分,并进行了去短文本(少于30个字符)处理操作,再以影评的评分作为判定依据,2星及以下判定为情感消极的评论,反之4星及以上判定为情感积极的评论,分别获取消极情感评论文本10 000条,积极情感评论文本20 000条。

情感词词典来自大连理工大学的情感词本体数据库,程度副词和否定词来自于知网中文词库HowNet。为了构建语义信息和情感信息的词向量,采用结巴分词系统对酒店评论语料文本句子进行分词,在分词过程中,将构建的情感语言库作为自定义分词词典,如表1,使其在分词后作为一个基本的语言单元存在。此外,还进行了去重、过滤、去停用词等文本预处理。

实验设置的超参数如表2所示。

表1 情感语言库

Table 1 Emotional language library

情感信息	数目/个	实例
情感词	27 466	杰出、勤奋、没礼貌、弱不禁风
否定词	59	不、从不、不如、没有、白搭
程度副词	219	稍、过于、极度、愈发、完完全全

表2 模型参数设置

Table 2 Model parameter setting

参数	含义	值
batchSize	每批次使用的样本数	64
embeddingSize	词嵌入维度	200
gru_size	GRU网络单元数	128
dropoutKeepProb	丢失率	0.5
learningRate	学习率	0.001
l2RegLambda	$L_2$ 正则化系数	0.001

## 2.2 实验评测指标

采用文本情感分析中常用的准确率  $P$  (Precision)、召回率  $R$  (Recall)、 $F_1$  测度 ( $F_1$  measure) 和损失率  $L$  (Loss rate) 作为实验数据的评测指标。

$$P = \frac{n_{TP}}{n_{TP} + n_{FP}} \quad (16)$$

$$R = \frac{n_{TP}}{n_{TP} + n_{FN}} \quad (17)$$

$$F_1 = \frac{2n_{PR}}{P + R} \quad (18)$$

$$L = 1 - \frac{\sum_{k=1}^G y_k = \hat{y}_k}{G} \quad (19)$$

式中,对于某一类别,  $n_{TP}$  表示正确判断属于该类别的数量,  $n_{FP}$  表示误判为该类别的数量,  $n_{FN}$  指误判为错误类别的数量,  $G$  为测试数据集。

## 2.3 对比实验

进行3组对比实验,所有实验均基于情感二分类(积极和消极)的任务进行十折交叉验证。第1组比较词向量工具对模型的影响,只采用SAC通道;第2组实验对比SAC和EAC对模型分类性能的影响;第3组实验探究模型情感分类迭代过程中准确率和损失率的变化情况。

### 2.3.1 词向量工具对比实验

基于SAC通道信息提取层中的BiGRU-Attention神经网络模型对比不同词向量工具的性能。

Word2vec-SAC:采用Word2vec作为词向量工具,然后通过BiGRU-Attention神经网络进行训练。

Glove-SAC:将词通过Glove模型转换为词向量以后,利用BiGRU-Attention神经网络进行训练。

ELMo-SAC:将词通过ELMo模型转换为词向量以后,利用BiGRU-Attention神经网络进行训练。

BERT-SAC:采用Google预训练好的中文模型BERT-Base-Chinese训练数据文本转换为词向量以后,利用BiGRU-Attention神经网络进行训练。

对表3中的实验结果进行分析:

(1)由于Glove模型是在统计词向量模型和预测词向量模型基础上,通过矩阵分解的方法利用词共现信

表3 基于不同词向量模型的对比结果

Table 3 Comparison results based on different word vector models

语料库	模型	$P/\%$	$R/\%$	$F_1/\%$
数据1	Word2vec-SAC	87.46	86.05	86.75
	Glove-SAC	87.88	86.46	87.16
	ELMo-SAC	92.11	92.01	92.06
	BERT-SAC	94.09	93.88	93.98
数据2	Word2vec-SAC	87.96	86.65	87.30
	Glove-SAC	88.95	86.93	87.75
	ELMo-SAC	92.29	92.58	91.93
	BERT-SAC	94.96	94.01	94.48

息,不仅关注 Word2vec 窗口(Context)大小的上下文,而且用到了全局信息,能较好地提高语义表征能力。因此,Glove-SAC 的模型评价结果略优于 Word2vec-SAC。

(2)ELMo与BERT作为基于语言模型的动态词向量,可以根据下游任务对语义表征进行微调,提高多义词的识别效率,因而评价指标均大幅提升。对比当前应用最广的 Word2vec 词向量,ELMo-SAC 模型在 2 个数据中的准确率、召回率、 $F_1$ 测度分别提高了 4.65%,5.96%,5.31%和 4.33%,5.93%,4.63%。

(3)相比 ELMo 采用 LSTM 进行提取词向量特征,BERT 采用更为强大的 Transformer 编码器,进一步提高了特征提取的能力,有效增强了情感表征,在 2 组数据中 BERT-SAC 均获得了最优结果。

### 2.3.2 分类模型对比实验

本组实验参照文献[24]选择了现阶段句子级情感分析领域中较为常用的几种分类模型进行对比研究,其中 5 种模型均采用 BERT 作为词向量工具。实验结果如表 4。

对表 4 中的实验结果进行分析:

(1)相比于 CNN,BiLSTM 更擅长时序特征的捕获,因此 BERT-CNN 实验效果稍逊于 BERT-BiLSTM;而 BERT-BiGRU 的表现性能更优,这与 Xu 等<sup>[25]</sup>针对 CNN 与 BiLSTM 的评论文本情感分析实验的结论一致。

(2)相比 BERT-BiGRU 模型,BERT-SAC 模型在 BiGRU 神经网络基础上增加了注意力机制,有助于提取文本重点信息,因而性能取得了明显的提升。

(3)相比于 BERT-SAC 模型,BERT-DCA 模型的性能有进一步提升,这是由于 BERT-DCA 构建了 SAC 和 EAC 双通道,其中 EAC 中采用由全连接网络和注意力机制相结合的深度网络模型,针对文本中包含否定词、程度副词、情感词等具情感色彩的词进行提取,结构化的情感语言词及全局上下文都得到了充分考虑,有助于捕捉文本潜在的语义特征,更好地挖掘出文本深层情感信息。

(4)相对于数据 1 的实验结果,数据 2 中 BERT-DCA 比 BERT-SAC 在准确率、召回率、 $F_1$ 测度上分别提升了 2.94%,3.31%,3.09%,且 5 个综合模型的综合效果均优于数据 1,这是因为本组实验选取了大于等于 30 个字符的长文本实验数据,其情感信息词频更高,文本数据中所含的情感信息更丰富,有利于情感信息注意力通道中注意力机制发挥更好的效果,准确捕捉文本情感极性。

(4)相对于数据 1 的实验结果,数据 2 中 BERT-DCA 比 BERT-SAC 在准确率、召回率、 $F_1$ 测度上分别提升了 2.94%,3.31%,3.09%,且 5 个综合模型的综合效果均优于数据 1,这是因为本组实验选取了大于等于 30 个字符的长文本实验数据,其情感信息词频更高,文本数据中所含的情感信息更丰富,有利于情感信息注意力通道中注意力机制发挥更好的效果,准确捕捉文本情感极性。

### 2.3.3 迭代过程性能对比实验

为进一步说明 BERT-DCA 模型的有效性,且更直观地体现注意力机制和语义信息与情感信息的双通道构造对模型的性能提升效果,本组实验选择均包含了双向 GRU 神经网络架构、较为相近的 3 种模型 BERT-BiGRU、BERT-SAC、BERT-DCA,对比分析其分别在 2 个数据集实验中 10 次迭代的准确率和损失率变化,实验结果如图 5,6 所示。

可以观察到,在 2 个数据集上 BERT-DCA 的准确率始终高于其他 2 种模型,而损失率则始终更低。而且,虽然随着迭代的进行,3 个模型均出现了不同程度的性能退化,但相对而言 BERT-DCA 的性能更稳定得多,准确度下降和损失率上升的幅度小,显然这得益于双通道的构建。在收敛速度上,BERT-

表 4 分类模型实验的对比结果

Table 4 Comparison results of classification model experiments

语料库	模型	P/%	R/%	$F_1$ /%
数据 1	BERT-CNN	89.71	88.98	89.34
	BERT-BiLSTM	90.32	90.00	90.16
	BERT-BiGRU	90.51	90.32	90.41
	BERT-SAC	94.09	93.88	93.98
	BERT-DCA	95.08	94.95	95.01
数据 2	BERT-CNN	88.95	88.16	88.55
	BERT-BiLSTM	90.23	90.01	90.12
	BERT-BiGRU	90.55	90.12	90.33
	BERT-SAC	94.91	93.58	94.24
	BERT-DCA	97.85	96.89	97.33

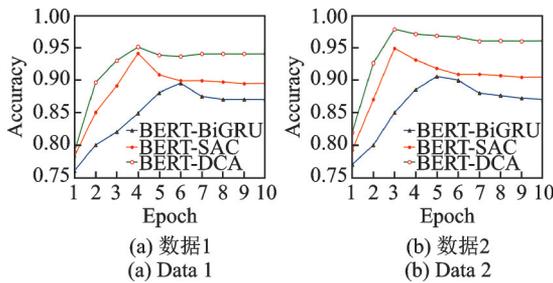


图5 3种模型准确率变化曲线

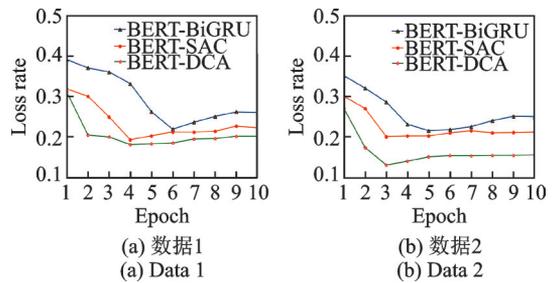


图6 3种模型损失率变化曲线

DCA和BERT-SAC模型在2个数据集上均优于BERT-BiGRU模型,这归功于注意力机制带来的优化效果。在数据2上3个模型的收敛速度均比在数据1上更快些,这是由于数据2选择的文本序列长度较为统一,情感信息更为丰富,为模型准确快速识别评论文本情感极性提供了良好条件。

表5给出了部分评论文本示例在BERT-DCA模型训练后的情感分类效果。

表5 实例展示

Table 5 Case demonstration

文本示例	情感极性	分词结果	模型结果
景色还很不错,那一片花海好漂亮。付辛博好帅,跟白冰有点不搭,剧情嘛,轻轻松松,看电影而已,别较真,哈哈,有笑点 还不错啊 大家去看看吧。	积极	景色/还/很/不错/那/一/片/花海/好/漂亮/付辛博/好/帅/跟/白冰/有点/不/搭/剧情/嘛/轻轻松松/看/个/电影/而已/别/较真/哈哈/有/笑/点/还不错/啊/大家/去/看看/吧/	积极
故事太多太杂,都没法好好展开。结尾转折普遍生硬。有种一锅大乱炖的感觉。	消极	故事/太多/太杂/都/没/法/好好/展开/结尾/转折/普遍/生硬/有/种/一/锅/大乱/炖/的/感觉/	消极
这个男主角很厉害,我从头到尾都没看到他的脸!很有演技很有爱! 不过确实比同类型的都好看。	积极	这个/男/主角/很/厉害/我/从/头/到/尾/都/没/看到/他/的/脸/很/有/演技/很/有/爱/不过/确实/比/同/类型/的/都好看/	积极

### 3 结束语

针对句子级文本情感分析任务,提出了一种基于BERT和双通道注意力的新模型。采用BERT作为词向量工具,一方面,由BiGRU-Attention神经网络作为语义信息注意力通道,另一方面,通过自定义的情感信息集合提取策略和引入注意力机制,构建并行的情感信息注意力通道;最后进行双通道特征融合,利用Softmax分类器实现情感分类。实验结果表明,相比其他词向量工具,BERT的特征提取能力更为优异,而情感信息通道和注意力机制增强了模型捕捉情感语义的能力,明显提升了情感分类性能,且在收敛速度和稳定性上表现更优。

#### 参考文献:

[1] LIU Z, YANG N, CAO S. Sentiment-analysis of review text for micro-video[C]//Proceedings of 2016 2nd IEEE International Conference on Computer and Communications (ICCC). [S.l.]: IEEE, 2016: 526-530.  
 [2] 赵妍妍, 秦兵, 石秋慧, 等. 大规模情感词典的构建及其在情感分类中的应用[J]. 中文信息学报, 2017, 31(2): 187-193.  
 ZHAO Yanyan, QIN Bing, SHI Qiuhui, et al. Large-scale sentiment lexicon collection and its application in sentiment classifi-

- cation[J]. *Journal of Chinese Information Processing*, 2017, 31(2): 187-193.
- [3] KESHAVARZ H, ABADDEH M S. ALGA: Adaptive lexicon learning using genetic algorithm for sentiment analysis of microblogs[J]. *Knowledge-Based Systems*, 2017, 122: 1-16.
- [4] 蒋翠清, 郭轶博, 刘尧. 基于中文社交媒体文本的领域情感词典构建方法研究[J]. *数据分析与知识发现*, 2019, 3(2): 98-107.
- JIANG Cuiqing, GUO Yibo, LIU Yao. Constructing a domain sentiment lexicon based on chinese social media text[J]. *Data Analysis and Knowledge Discovery*, 2019, 3(2): 98-107.
- [5] PERIKOS I, HATZILYGEROUDIS I. Recognizing emotions in text using ensemble of classifiers[J]. *Engineering Applications of Artificial Intelligence*, 2016, 51: 191-201.
- [6] 王煜涵, 张春云, 赵宝林, 等. 卷积神经网络下的 Twitter 文本情感分析[J]. *数据采集与处理*, 2018, 33(5): 921-927.
- WANG Yuhan, ZHANG Chunyun, ZHAO Baolin, et al. Emotional analysis of Twitter text in convolutional neural network [J]. *Journal of Data Acquisition and Processing*, 2018, 33(5): 921-927.
- [7] ZHU X, LI L, ZHANG W, et al. Dependency exploitation: A unified CNN-RNN approach for visual emotion recognition[C]// *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. New York, USA: ACM, 2017: 3595-3601.
- [8] ZHANG L, ZHOU Y, DUAN X, et al. A hierarchical multi-input and output bi-GRU model for sentiment analysis on customer reviews[C]// *Proceedings of IOP Conference Series: Materials Science and Engineering*. Berlin, German: Springer, 2018.
- [9] ZHAI Penghua, ZHANG Dingyi. Bidirectional-GRU based on attention mechanism for aspect-level sentiment analysis[C]// *Proceedings of the 2019 11th International Conference on Machine Learning and Computing*. New York, USA: ACM, 2019.
- [10] MIKOLOV T. Distributed representations of words and phrases and their compositionality[C]// *Proceedings of Advances in Neural Information Processing Systems*. [S.l.]: MIT Press, 2013, 26:3111-3119.
- [11] PENNINGTON J, SOCHER R, MANNING C D. Glove: Global vectors for word representation[C]// *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, 2014.
- [12] DAI A M, LE Q V. Semi-supervised sequence learning[C]// *Proceedings of Advances in Neural Information Processing Systems*. [S.l.]: MIT Press, 2015.
- [13] PETERS M E, NEUMANN M, IYYER M, et al. Deep contextualized word representations[C]// *Proceedings of NAACL-HLT*. Los Angeles, California: Association for Computational Linguistics, 2018.
- [14] HOWARD J, RUDER S. Universal language model fine-tuning for text classification[C]// *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Melbourne, Australia: Association for Computational Linguistics, 2018.
- [15] 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*. Minneapolis, Minnesota: Association for Computational Linguistics, 2019.
- [16] FU Xianghua, LIU Guo, GUO Yanyan, et al. Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon[J]. *Knowledge-Based Systems*, 2013, 37: 186-195.
- [17] MINH D L, SADEGHI-NIARAKI A, HUY H D, et al. Deep learning approach for short-term stock trends prediction based on two-stream gated recurrent unit network[J]. *IEEE Access*, 2018, 6: 55392-55404.
- [18] LEI Z, YANG Y, YANG M, et al. A multi-sentiment-resource enhanced attention network for sentiment classification[C]// *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*. Melbourne, Australia: Association for Computational Linguistics, 2018.
- [19] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// *Proceedings of Advances in Neural Information Processing Systems*. Doha, Qatar: Association for Computational Linguistics, 2017: 5998-6008.
- [20] ADHIKARI A, RAM A, TANG R, et al. DocBERT: BERT for document classification[EB/OL]. (2019-08-22)[2019-09-07]. <https://arxiv.org/abs/1904.08398v3>.
- [21] CHO K, VAN MERRIENBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for

- statistical machine translation[C]//Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP 2014). Doha, Qatar: Association for Computational Linguistics, 2014.
- [22] MNIH V, HEESS N, GRAVES A. Recurrent models of visual attention[C]//Proceedings of Advances in Neural Information Processing Systems. Doha, Qatar: Association for Computational Linguistics, 2014.
- [23] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[C]//Proceedings of International Conference on Learning Representations. Doha, Qatar: Association for Computational Linguistics, 2015.
- [24] ZHOU K, LONG F. Sentiment analysis of text based on CNN and bi-directional LSTM model[C]// Proceedings of 2018 24th International Conference on Automation and Computing (ICAC). Piscataway, NJ: IEEE, 2018.
- [25] XU G, MENG Y, QIU X, et al. Sentiment analysis of comment texts based on BiLSTM[J]. IEEE Access, 2019, 7: 51522-51532.

**作者简介:**

谢润忠(1992-),男,硕士研究生,研究方向:自然语言处理、深度学习, E-mail: 1228905230@qq.com。



李焯(1974-),通信作者,男,高级工程师,研究方向:自然语言处理、机器学习、人工智能等, E-mail: liye@usst.edu.cn。

(编辑:张彤)