

# 结合 BERT 和卷积双向简单循环网络的文本情感分析

黄泽民 吴迎岗

(广东工业大学计算机学院 广东 广州 510006)

**摘要** 针对传统预训练模型无法解决多义词表征问题和抽取的情感特征能力不足等问题,提出一种结合 BERT 和卷积双向简单循环网络的文本情感分类模型。用 BERT 预训练得到融合句子语境的动态词向量;用多粒度卷积神经网络对词向量特征进行二次抽取,池化后注入局部响应归一化层 LRN 来归一化特征图,以生成句子整体特征表示;利用双向简单循环单元进一步学习上下文语义信息;使用 Softmax 回归得出文本的情感倾向。实验结果表明,结合 BERT 和卷积双向简单循环网络的文本情感分类模型能获得更高的准确率,注入的 LRN 层和并行化循环网络有效提高模型性能,具有较好的实用价值。

**关键词** 文本情感分析 双向解码器 上下文信息 双向简单循环单元 卷积神经网络

**中图分类号** TP391 **文献标志码** A **DOI**:10.3969/j.issn.1000-386x.2022.12.032

## TEXT EMOTION ANALYSIS BASED ON BERT AND CNN-BISRU

Huang Zemin Wu Yinggang

(School of Computer, Guangdong University of Technology, Guangzhou 510006, Guangdong, China)

**Abstract** Aimed at the problem that the traditional language model can not solve the problem of word ambiguity in word vector representation and not be able to capture more semantic features in existing text sentiment analysis methods, the text emotion analysis based on BERT and CNN-BiSRU model is proposed. The BERT was used to obtain the word dynamic vector representation that integrated text semantics. The word vector features was secondly extracted by multi-kernel convolution neural network (CNN), and the local response normalization layer (LRN) was injected to normalize the feature map after pooling to generate the overall feature representation of the sentence. Bidirectional simple recurrent unit (BiSRU) was used to further extract semantic context information. Softmax function was utilized to determine the emotional tendency of the text. The experimental results show that the model based on BERT and CNN-BiSRU can achieve higher accuracy, the injected LRN layer and the parallel circular network can improve the performance of this model. It is of great value in practical application.

**Keywords** Text sentiment analysis BERT Context information Bidirectional simple recurrent unit Convolutional neural network

## 0 引言

随着通信网络和移动应用的快速发展,各种社交媒体也渐渐兴旺,网络用户越来越多,用户通过各种通信媒体来评论热点事件、表达自己的感受,从而产生了大量带有感情色彩的文本信息,这些情感倾向对政府相关部门或商家都有很大用处,但是这些海量文本的

评论形式复杂多样,仅靠人工处理存在工作量重和实时性较差等问题,如 Hu 等<sup>[1]</sup> 依托 WordNet 词典计算待处理文本的情感得分,虽取得了较好的预测效果,但存在大量人工标注工作。自然语言处理领域的文本情感分析<sup>[2]</sup> 可以从这些复杂多样的信息中提取情感倾向,从而应用于各种需要情感倾向的场景。

目前,情感分析模型主要有基于机器学习和基于神经网络等方法。基于机器学习的方法,如 Pang 等<sup>[3]</sup>

应用支持向量机、朴素贝叶斯和最大熵等情感分析方法,在分析影评文本情感上有较好的效果,但是在各媒体平台中,存在评论文本长度不一致、表达不规整的问题,导致传统的机器学习方法需要面临样本人工结构化、表达稀疏和计算繁杂等问题,不能获得理想的结果。基于神经网络的方法,杨云龙等<sup>[4]</sup>提出通过 GRU 获得语义特征和整体标量信息;然后在初始胶囊层用动态路由算法进行迭代后,获取到融合文本整体属性的向量化表示,最后在主胶囊层获得更准确的文本特征,并根据每个特征的特征表示强弱来分析文本的情感极性,从而在测试集有着不错的预测效果。湛志群等<sup>[5]</sup>提出采用 BERT 模型生成文本的动态向量表征,然后词表征输入到双向 LSTM 网络中进行情感分类,取得了不错的预测结果。Salur 等<sup>[6]</sup>将不同的词向量化方法,如 FastText、Word2Vec 等,与不同的深度学习方法(GRU、LSTM、BiLSTM)适当整合在一起,提取了不同的特征,结合这些特征来训练和预测分类。Ma 等<sup>[7]</sup>利用特征复合内存网络(FCMN)提取三种特征去更好地表示上下文单词的词向量,而且策略性地结合特征表示和词嵌入以提高注意力机制的作用,从而在测试集的情感预测方面有较好效果。赵亚欧等<sup>[8]</sup>使用预训练好的基于字符的向量初始化 ELMO 层,然后利用了多尺度卷积神经网络二次抽取词向量的特征,将特征融合后以生成句子的语义表征,此方法能明显提高预测准确率。上述学者虽都取得较好的分类效果,但没有综合考虑中文词汇的多义词性质、上下文信息和词语之间的联系特征对情感分析的影响。

基于统计机器学习的方法虽然提高了准确率,但人工结构化数据需要先验知识才能使得该方法应用于规模较大的样本集,导致效率下降。基于神经网络的方法可以自动提取特征,但使用单一的神经网络不能够充分提取句子特征和词语特征,且基于神经网络的模型输入是句子的词向量,若仅用 Word2vec、glove 等学习来的词向量作为向量表示,学习到的只是静态向量,即不能够学习到多义词本身具有的灵活性运用,会减低文本情感倾向的准确预测。综上,本文提出的结合 BERT 和卷积双向简单循环网络模型能解决以上问题,且在卷积层注入 LRN 和在循环网络中实现训练的并行化,有效地提高预测准确率和减少训练时间。

## 1 网络模型

### 1.1 情感分析模型

结合 BERT 和卷积双向简单循环网络模型如图 1

所示。数据预处理是对评论文本分词和去噪;BERT 层是对句子词静态向量的重构,以生成融合语境的词向量表征;卷积层二次提取词语特征,利用 LRN 优化特征图;BiSRU 提取文本的上下文信息;最后情感分类使用 Softmax 回归。

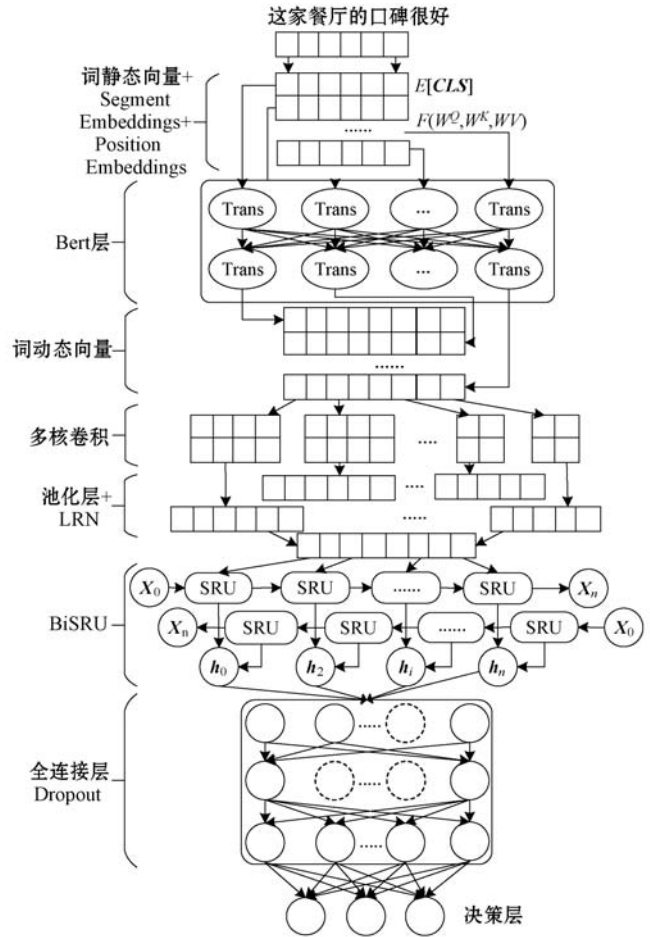


图 1 文本情感分析模型

### 1.2 数据清洗及词静态向量表征

评论文本存在各种无用信息和每个句子长短不一,需对评论文本进行预处理:对评论文本利用正则表达式进行去噪,如标点符号、表情等;然后对语料的每一条独立文本进行分词,并且会去除干扰词和停用词等,分词后只包含具有语义信息的词语集合,且保证分词数目不超过设置的最大长度减 2,因为剩下的两个位置用来存放句首标志和分句标志;预处理后对文本的每个分词进行静态向量映射,用来初始化 BERT 模型,此时的向量表示不能解决多义词问题。

### 1.3 BERT 预训练模型

自然语言处理的预训练模型<sup>[9]</sup>有动态模型 BERT<sup>[10]</sup>,其采用了融合特征能力强的双向 Transformer 模型,预训练效果相比静态模型 Word2vec<sup>[11]</sup>能解决多义词表征问题;相比动态模型 ELMO 模型<sup>[12]</sup>和 GPT 模型<sup>[13]</sup>有更强的语境融合能力。BERT 预训练模型如图 2 所示。

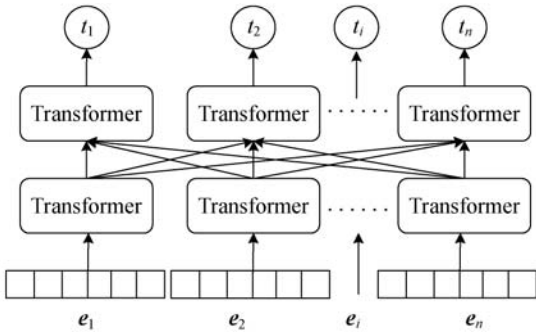


图 2 BERT 的网络模型

BERT 模型的输入向量  $e_i$  由三种不同的向量对应元素相加而成。第一种为词静态向量,可由 Word2vec 工具得到;第二种为位置向量,用来保存分词在语料序列中的相对位置或绝对位置;第三种为句向量,由于本文的情感预测是句子级别的情感分析,即输入是一个句子,故仅使用一个句向量,且对每个句子添加一个 [CLS] 标志向量(可用于下游的分类任务)和句尾 [SEP] 标志向量(用作两个句子的分隔符)。

本文利用了谷歌的预训练模型来微调 BERT 模型,由于 BERT 是随机掩码模型,掩码词可能在微调阶段未出现过,故对被掩码词用 [mask] 标志替代、语料其他词替代或不变处理,来保证模型的泛化性。

Transformer 结构如图 3 所示。其中多头注意力子层是词向量化的关键,由 8 个自注意力机制组成。

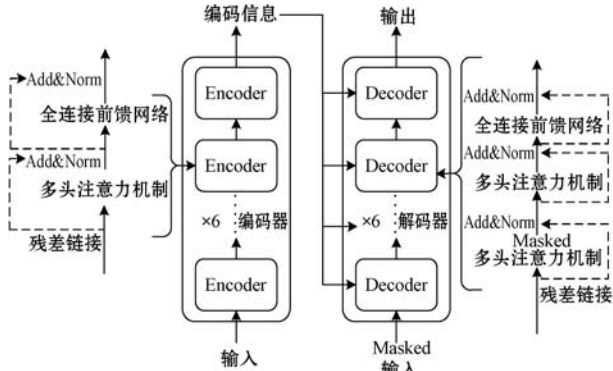


图 3 Transformer 模型结构

BERT 在操作中是基于矩阵计算的,全部输入向量拼接成向量矩阵  $E, E = \{e_1, e_2, \dots, e_n\}$ , 交付给自注意力机制,其输出见式(1)和式(2)。

$$Q = EW^Q, K = EW^K, V = EW^V \quad (1)$$

$$Attention(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

式中:  $W^Q, W^K$  和  $W^V$  为线性变换矩阵;  $Q$  为查询矩阵,  $K$  为键矩阵,  $V$  为值矩阵;  $d_k$  的值为向量维度;惩罚因子  $\sqrt{d_k}$  是为防止  $QK^T$  内积过大; Softmax 函数是归一化  $\frac{QK^T}{\sqrt{d_k}}$  运算后的每个行向量,目的是计算出每个单词对

于其他单词的重要程度。

计算出自注意力的输出后,通过式(3)和式(4)得出多头注意力机制的输出,记为  $X$ 。

$$Attention_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (3)$$

$$X = \text{MultiHead}(Q, K, V) =$$

$$\text{concat}(Attention_1, Attention_2, \dots, Attention_8) \cdot W \quad (4)$$

取 BERT 模型最后一层的输出向量矩阵记为  $T = \{t_1, t_2, \dots, t_n\}$ , 矩阵  $T$  的维度与 BERT 输入矩阵  $E$  相同,每个向量表示分词的无歧义深度向量表示,作为下游任务的输入。

### 1.4 局部特征提取

卷积神经网络是能在给定的结构中总结出具有特征性的局部预测器,将它们结合以生成特征矩阵来表示该结构,捕获最具有任务特征的局部特征。BERT 层的输出向量矩阵  $T$  作为卷积神经网络的输入,卷积核在句子词向量矩阵上滑动,卷积核与对应的句子词向量矩阵窗口上进行对应元素相乘后作和,其值作为最后特征矩阵的一个特征值,所有特征值组成一幅特征图,其计算式为式(5)、式(6)和式(7)。

$$x_i = \bigoplus T_{i:i+t-1} \quad (5)$$

$$c_i = f(w^c \cdot x_i + b^c) \quad (6)$$

$$y = \{c_1, c_2, \dots, c_i, \dots, c_{n-t+1}\} \quad (7)$$

式中:卷积核大小为  $t$ ;  $x_i$  为第  $i$  个窗口的拼接向量,即  $x_i = \bigoplus T_{i:i+t-1} = [T_i, T_{i+1}, \dots, T_{i+t-1}]$ ,  $x_i \in \mathbf{R}^{t \times M}$  ( $M$  为词向量维度);权重矩阵  $w^c \in \mathbf{R}^{t \times M}$  和偏置  $b^c$  都通过参数估计过程进行学习;  $f$  为激励函数,去线性化处理;  $c_i$  表示特征图的第  $i$  个特征值,为标量,  $c_i \in \mathbf{R}$ ;  $y$  为卷积完成得到的特征图。

本文模型使用多个卷积核,(1)充分提取特征;(2)为 LRN 优化特征图的归一化做前提。将所有的特征图经过 ReLU 去线性化和最大池化处理,送入 LRN 局部优化层,来优化所有的特征图。

### 1.5 LRN 优化层

局部响应归一化 LRN 层<sup>[14]</sup>模仿了生物神经系统的“侧抑制”机制,使局部神经元之间竞争响应机会,让响应较大的值变得相对更大,并抑制其他响应较小的神经元。卷积神经网络在池化后利用 LRN 优化使同一个特征图的特征值都有相似的方差,能加快模型的收敛速度和提高特征提取能力。本文的 LRN 使用通道内归一化模式,即在同一个特征图上操作。LRN 的计算公式见式(8)。

$$b = \frac{a}{\left(k + \frac{\alpha}{n} \sum_i (x_i)^2\right)^\beta} \quad (8)$$

式中: $a, b$  分别表示在同一个特征图的原特征值和 LRN 优化后的特征值; $n$  表示局部归一化区域的大小, 设为 5; $\alpha$  为缩放因子, 设为 1; $\beta$  为指数项, 设为 5; $k$  为平滑操作, 设为 1; $\Sigma$  是以当前特征值为中心的区域计算平方和, 即  $1 \times n \times n$  个特征值的平方和, 不满  $1 \times n \times n$  个特征值的会补零后再求平方和。

每个原始特征图的每一个特征值都是除以一个系数来达到归一化的目的, 该系数为式(8)等式右边的分母部分。经过 LRN 优化后的各个特征图与原始特征图的维度相同, 对所有经 LRN 优化后的所有特征图连接成表示句子特征的特征向量, 作为 BiSRU 层的输入。

## 1.6 上下文语义提取

长短期记忆网络 LSTM 的出现, 通过对循环层的重新设计来避免了 RNN 网络的长期依赖问题和梯度消失的问题, 并有效解决了历史链接问题, 但 LSTM 神经网络的训练很难并行化。本文提出的 SRU 模型不仅具有 LSTM 的能力, 也具有并行计算的能力, 它平衡了输入的顺序依赖和独立性, 简化了计算状态的过程。SRU 模型<sup>[15]</sup>的网络结构如图 4 所示。

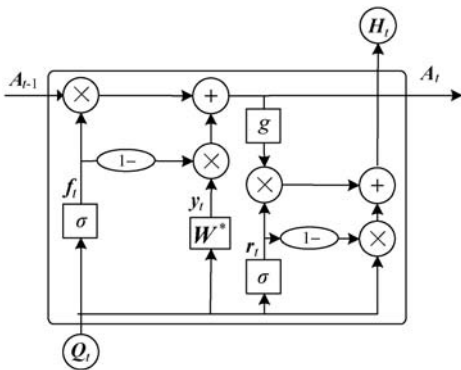


图 4 SRU 模型的网络结构

在图 4 中,  $Q_t$  表示在  $t$  时刻的输入,  $f_t$  表示忘记门,  $r_t$  表示在  $t$  时刻复位门,  $A_t$  和  $H_t$  分别表示当前时刻  $t$  的细胞状态和最终输出,  $A_{t-1}$  表示上一时刻  $t-1$  的细胞状态,  $\sigma$  和  $g$  代表 Sigmoid 函数和激活函数(双曲正切或 ReLU),  $\odot$  表示矩阵对应元素的乘法运算。

SRU 模型使用丢弃连接, 即当前时刻的输入消除了对上一时刻  $H_{t-1}$  的依赖性, 以下是 SRU 模型的计算公式:

$$f_t = \sigma(W_f Q_t + b_f) \quad (9)$$

$$r_t = \sigma(W_r Q_t + b_r) \quad (10)$$

$$y_t = \sigma(W^* Q_t) \quad (11)$$

$$A_t = f_t \odot A_{t-1} + (1 - f_t) \odot y_t \quad (12)$$

$$H_t = r_t \odot g(A_t) + (1 - r_t) \odot Q_t \quad (13)$$

式中:  $W_f, W_r$  和  $W^*$  是参数矩阵;  $b_f, b_r$  是偏置, 都是训

练时要学习的参数向量。

从式(9)到式(11)可以看到  $H_{t-1}$  的依赖性被消除了, 通过式(12)和式(13)可以快速地实现, 因为输出不再依赖于输入  $H_{t-1}$ , 从而可以批次处理所有时间步的矩阵相乘, 明显提高计算速度。

利用双向 SRU 允许模型在序列内部向前和向后看任意远的距离。BiSRU 模型是由两个 SRU 网络通过上下叠加形成的, 每一个时刻会同时存在相反方向的 SRU 中, 将卷积层的输出向量分别输入到正反序 SRU, 会得到 SRU 每个时间步的输出, 每一个时间步的计算公式为:

$$h_t = [\vec{h}_t, \overleftarrow{h}_t] = \text{BiSRU}(Q_t) \quad (14)$$

式中:  $\vec{h}_t$  表示  $t$  时刻 SRU 正向输出;  $\overleftarrow{h}_t$  表示  $t$  时刻 SRU 反向输出;  $h_t$  为  $t$  时刻的 BiSRU 输出;  $Q_t$  为卷积层的输出向量。

将每一个时间步的正反序叠加值进行拼接, 拼接的方式能在线性时间内高效地运算, 最后得到全部时间步的输出。BiSRU 模型的输出输入到全连接 dropout 层, 该层在训练时会概率性地随机作用于隐藏层的节点, 使其暂时被选择隐藏掉而不工作, 但是其权重会被保留, 从而提高了模型的泛化能力, 在本文的 dropout 比率为 0.5。最后, 利用 Softmax 回归得到预测概率分布, 取概率值最大的为预测结果。

## 2 实验与结果分析

### 2.1 数据集及实验评价标准

实验的数据集为中文版本的 Twitter 数据集, 包含 2 万多条中文标注语料, 其中积极情感句子 8 700 多条, 消极句子 8 100 多条, 中性句子 5 500 条。由于数据不平衡会造成分类器在训练的过程中严重偏向样本较多的类别, 使模型的综合性能大大下降, 因此, 本文为保证训练集各类别的数目相同, 都设为 5 000 条, 各类别剩余的数据集作为测试集或者验证集。为验证模型在文本情感分析的有效性, 使用评价标准为准确率 (Precision)、召回率 (Recall) 和 F1 (F-measure) 值。计算式如式(15) - 式(17)所示。公式的参数含义见表 1。

表 1 公式参数含义

实际结果	预测结果	
	属于类别 A	不属于类别 A
属于类别 A	$s$	$t$
不属于类别 A	$m$	$n$

*Precision* 表示预测为某个类别的样本数目与真正类别的样本数目的比值。

$$Precision = \frac{s}{s+m} \quad (15)$$

*Recall* 表示预测为某个类别中的真实类别数目与所有真实类别数目的比值。

$$Recall = \frac{s}{s+t} \quad (16)$$

由于上述两种经常矛盾,故引入综合评价指标 *F1*, 计算如下:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (17)$$

本文的情感分类是三分类,因此在计算出各个类别对应的准确率和召回率后,采用平均值作为分类器性能的评价指标。

## 2.2 实验参数的选择

本文模型的各层参数设置如下:词向量维度为 768 维;由于汉语中词语之间的距离过大会导致语义联系较弱,故卷积核粒度分别为 3、4 和 5,激活函数选择 ReLU,因 ReLU 是无上界函数,适应 LRN 的“侧抑制”原理;SRU 隐藏层节数为 128;损失函数为交叉熵函数(学习率为 0.000 01);模型优化器为 Adam;训练批次长度为 32;epoch 为 10;最后的 Dense 层使用 3 分类。

## 2.3 实验模型对比分析

为验证本文方法的有效性,在相同的实验环境下,将本文情感分析模型与其他情感分析混合模型进行对比,最后结果如表 2 所示。

表 2 与其他模型的性能比较

模型	<i>Precision</i>	<i>Recall</i>	<i>F1</i>
Baseline	87.93	86.87	87.39
W2V-CNN(LRN)-BiSRU <sup>[16]</sup>	85.71	85.07	85.38
ELMO-CNN(LRN)-BiSRU	87.97	87.69	87.82
BERT-CNN(LRN)-BiLSTM	91.88	90.69	91.28
BERT-CNN(LRN)-BiRNN	89.17	88.83	89.01
BERT-CNN-BiSRU	89.45	89.75	89.52
BERT-CNN(LRN)	88.56	89.17	88.84
BERT-BiSRU	88.95	88.78	88.86
BERT-CNN(LRN)-BiSRU	91.38	91.26	91.27

基线模型 1 是直接将 BERT 的输出连接到全连接层后输入到 Softmax 分类器得出的结果,相对基线模型,本文模型的准确率能提高近 4 个百分点;对比模型 2 和模型 3 是为了验证 BERT 预训练模型的有效性,因为 BERT 预训练解决了多义词表征问题和利用了融合

文本能力强的 Transformer 作为预训练模型的子结构,从而使模型分析能力大大提升,且 BERT 模型使用的是预训练的方式,是从大语料集中训练语义向量,而非在任务语料中训练,使得训练出来的词向量表征拥有更高的泛化程度;与模型 4 和模型 5 作对比是为了验证 BiSRU 模型的语义学习的有效性,由于 LSTM 和 SRU 模型能对历史信息进行策略性的过滤和保留,使得语义学习的能力更强;与模型 6 对比是为了验证卷积层注入 LRN 层的预测效果,由表 2 可知注入 LRN 层后预测准确率明显提高;对比模型 7 和模型 8 是为了验证本文模型与基于 BERT 的单一模型的有效性,由表 2 可知,相比基于 BERT 的单一特征提取模型,本文利用的混合模型能更好地提取文本特征,在各个评价指标都有不同程度的提升。最终的实验结果表明,本文提出的结合 BERT 和卷积双向简单循环网络的文本情感分析在测试集上有很好的表现,证明了本文模型的有效性。

本文的模型分别与 BERT-CNN-BiLSTM、BERT-CNN-LSTM 等模型的每一个 epoch 的训练时间进行比较,结果如图 5 所示,在相同的实验环境下,本文模型的训练时间明显缩小。BERT 预训练模型总参数大小为 110 MB,如果利用计算复杂度高的 LSTM 网络作为下游任务的模型,无疑会导致训练时间大大增加,故本文利用了 SRU 模型的并行化的计算能力,在保证准确率的条件下,能有效地减少训练时间。

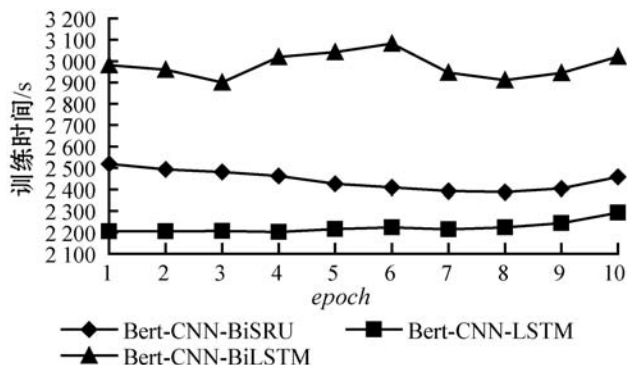


图 5 每个 epoch 的训练时间

## 3 结 语

本文提出的结合 BERT 和卷积双向简单循环网络的文本分析模型实现了文本情感倾向的三分类,在实验数据集上的预测结果表明,本文提出的模型能够提取融合文本语境的词表征,然后利用该词表征作为下游策略模型的输入,并在卷积层注入 LRN 层和循环网络中并行化参数估计过程,从而有效提升模型的情感分析的总体性能。

## 参 考 文 献

- [ 1 ] Hu M Q, Liu B. Mining and summarizing customer reviews [ C ] // 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2004 : 22 - 25.
- [ 2 ] 洪巍, 李敏. 文本情感分析方法研究综述 [ J ]. 计算机工程与科学, 2019, 41 ( 4 ) : 750 - 757
- [ 3 ] Pang B, Lee L, Vaithyanathan S. Thumbs up sentiment classification using machine learning techniques [ C ] // Conference on Empirical Methods in Natural Language Processing, 2002 : 79 - 86.
- [ 4 ] 杨云龙, 孙建强, 宋国超. 基于门控循环单元和胶囊特征的文本情感分析 [ J ]. 计算机应用, 2020, 9 : 2531 - 2535.
- [ 5 ] 湛志群, 鞠婷. 基于 BERT 和双向 LSTM 的微博评论倾向性分析研究 [ J ]. 情报理论与实践, 2020, 8 : 173 - 177.
- [ 6 ] Salur M U, Aydin I. Novel hybrid deep learning model for sentiment classification [ J ]. IEEE Access, 2020, 8 : 58080 - 58093.
- [ 7 ] Ma R X, Wang K, Qiu T, et al. Feature-based compositing memory networks for aspect-based sentiment classification in social Internet of things [ J ]. Future Generation Computer Systems, 2017, 92 : 879 - 888.
- [ 8 ] 赵亚欧, 张家重, 李贻斌, 等. 融合基于语言模型的词嵌入和多尺度卷积神经网络的情感分析 [ J ]. 计算机应用, 2020, 40 ( 3 ) : 651 - 657.
- [ 9 ] 李舟军, 范宇, 吴贤杰. 面向自然语言处理的预训练技术研究综述 [ J ]. 计算机科学, 2020, 47 ( 3 ) : 162 - 173.
- [ 10 ] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [ C ] // Advances in Neural Information Processing Systems, 2017 : 5998 - 6008.
- [ 11 ] Mikolov T, Sutskever I, Kai C, et al. Distributed representations of words and phrases and their compositionality [ J ]. Advances in Neural Information Processing Systems, 2013, 26 : 3111 - 3119.
- [ 12 ] Peters M E, Neumann M, Iyyer M, et al. Deep contextualized word representations [ EB ]. arXiv : 1802.05365, 05365
- [ 13 ] Radford A, Narasimhan K, Salimans T, et al. Improving language understanding by generative pre-training [ EB/OL ]. [ 2020 - 05 - 20 ]. [https://s3-us-west-2, amazonaws. com/ openai-assets/ research-covers/ languageunsupervised/ language-understanding-paper. pdf](https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/languageunsupervised/language-understanding-paper.pdf).
- [ 14 ] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [ C ] // International Conference on Neural Information Processing, 2012.
- [ 15 ] Tao L, Yu Z, Sida I. Simple recurrent units for highly parallelizable recurrence [ C ] // Conference on Empirical Methods in Natural Language Processing, 2018 : 4470 - 4481.
- [ 16 ] 廖小琴, 徐杨. 面向 BSP-CNN 的短文本情感倾向性分类研究 [ J ]. 计算机工程与应用, 2019, 55 ( 23 ) : 120 - 124.
- [ 17 ] Zhang L, Tjondronegoro D, Chandran V. Random Gabor based templates for facial expression recognition in images with facial occlusion [ J ]. Neurocomputing, 2014, 145 : 451 - 464.
- [ 18 ] 周志铭, 余松煜, 张瑞, 等. 一种基于 SIFT 算子的人脸识别方法 [ J ]. 中国图象图形学报, 2008, 13 ( 10 ) : 1882 - 1885.
- [ 19 ] 黄奕玮. 基于深度学习的面部表情识别研究与实现 [ D ]. 北京 : 北京邮电大学, 2019.
- [ 20 ] Shan C, Gong S, Mcowan P W. Facial expression recognition based on Local Binary Patterns: A comprehensive study [ J ]. Image and Vision Computing, 2009, 27 ( 6 ) : 803 - 816.
- [ 21 ] 莫修飞. 基于 SIFT 算法的人脸表情识别 [ D ]. 西安 : 西安电子科技大学, 2012.
- [ 22 ] 曾贤灏, 李向伟. 基于 Fisher 准则改进线性判别回归分类的人脸识别 [ J ]. 计算机应用与软件, 2014, 31 ( 9 ) : 184 - 186, 271.
- [ 23 ] 杨凡, 张磊. 基于 Gabor 参数矩阵与改进 Adaboost 的人脸表情识别 [ J ]. 计算机应用, 2014, 34 ( 4 ) : 1134 - 1138.
- [ 24 ] 陆嘉慧, 张树美, 赵俊莉. 基于深度学习的面部表情识别研究 [ J ]. 计算机应用研究, 2020, 37 ( 4 ) : 966 - 972.
- [ 25 ] Chen J, Lv Y, Xu R, et al. Automatic social signal analysis: Facial expression recognition using difference convolution neural network [ J ]. Journal of Parallel and Distributed Computing, 2019, 131 : 97 - 102.
- [ 26 ] Shao J, Qian Y. Three convolutional neural network models for facial expression recognition in the wild [ J ]. Neurocomputing, 2019, 355 : 82 - 92.
- [ 27 ] 张俞晴, 何宁, 魏润辰. 基于卷积神经网络融合 SIFT 特征的人脸表情识别 [ J ]. 计算机应用与软件, 2019, 36 ( 11 ) : 161 - 167.
- [ 28 ] Ojala T, Pietikäinen M, Harwood D. A comparative study of texture measures with classification based on featured distributions [ J ]. Pattern Recognition, 1996, 29 ( 1 ) : 51 - 59.
- [ 29 ] 徐琳琳, 张树美, 赵俊莉. 构建并行卷积神经网络的表情识别算法 [ J ]. 中国图象图形学报, 2019, 24 ( 2 ) : 227 - 236.
- [ 30 ] 陈航, 邱晓晖. 基于卷积神经网络和池化算法的表情识别研究 [ J ]. 计算机技术与发展, 2019, 29 ( 1 ) : 61 - 65.
- [ 31 ] 蒋昂波, 王维维. ReLU 激活函数优化研究 [ J ]. 传感器与微系统, 2018, 37 ( 2 ) : 50 - 52.
- [ 32 ] 丁名都, 李琳. 基于 CNN 和 HOG 双路特征融合的人脸表情识别 [ J ]. 信息与控制, 2020, 49 ( 1 ) : 47 - 54.
- [ 33 ] 陆嘉慧, 张树美, 赵俊莉. 基于 CNN 集成的面部表情识别 [ J ]. 青岛大学学报 ( 工程技术版 ), 2020, 35 ( 2 ) : 24 - 29, 42.
- [ 34 ] Mollahosseini A, Chan D, Mahoor M H. Going deeper in facial expression recognition using deep neural networks [ C ] // 2016 IEEE Winter Conference on Applications of Computer Vision ( WACV ), 2016.

( 上接第 179 页 )