

DOI:10.19322/j.cnki.issn.1006-4710.2021.01.016

# 融合 TCN 与 BiLSTM+Attention 模型的 疫情期间文本情感分析

贵向泉, 高 祯, 李 立

(兰州理工大学 计算机与通信学院, 甘肃 兰州 730050)

**摘要:** 鉴于目前主流的文本情感分析方法存在难以解决长期依赖和对上下文信息使用不足的缺陷, 本文首次提出将时序卷积网络(TCN)和 BiLSTM+Attention 模型融合的文本情感分析模型。该模型利用 TCN 的因果卷积和扩张卷积结构获取更高层次的文本序列特征, 并通过双向长短期记忆网络(BiLSTM)进一步学习上下文相关信息的情感特征; 最后, 引入自注意力机制(Self-Attention)帮助模型优化特征向量, 提高情感分类的准确度。在新型冠状病毒疫情期间的微博文本数据集上进行对比实验, 结果表明该模型的性能相较于其它模型有明显的提升。

**关键词:** 文本情感分析; 时序卷积网络; 双向长短期记忆网络; 自注意力机制; 疫情期间微博文本  
**中图分类号:** TP391      **文献标志码:** A      **文章编号:** 1006-4710(2021)01-0113-09

## Text sentiment analysis during the epidemic based on TCN and BiLSTM+Attention fusion model

GUI Xiangquan, GAO Zhen, LI Li

(School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China)

**Abstract:** At present, there are some difficulties in solving the problems of long-term dependence and the insufficient use of contextual information in the mainstream text sentiment analysis methods. In this paper, a text sentiment analysis model combining Temporal Convolutional Network (TCN) with BiLSTM + Attention model is proposed for the first time. This model uses TCN's causal convolution and extended convolution structures to obtain higher-level text sequence features, and further learns the emotional features of contextual information through the Bidirectional Long Short-Term Memory network (BiLSTM). Finally, the Self-Attention mechanism (Self-Attention) is added to optimize the feature vectors of this model and improve the accuracy of sentiment classification. A comparative verification experiment is conducted on the Weibo text data set during the novel coronavirus epidemic. The results show that the performance of this model is more significantly improved than that of other models.

**Key words:** text sentiment analysis; TCN; BiLSTM; Self-Attention mechanism; Weibo text during the epidemic

随着移动互联网终端技术的不断发展, 用户可以通过手机等移动端随时随地发表关于日常生活的、富有情感色彩的信息, 这种及时性和互动性也导致互联网中的信息量急剧增长, 仅凭人工很难从如此庞大的文本数据中提取到人们的情感倾向。文本

情感分析技术<sup>[1]</sup>的出现可以帮助人们更高效地从文本中挖掘出用户的情感信息, 对意见挖掘、网络舆论预测等应用领域具有重要意义<sup>[2-3]</sup>。2019年年底爆发的新型冠状病毒肺炎由于其传染能力之强和扩散范围之广在互联网上引发了强烈的公众情绪波动,

收稿日期: 2020-08-20; 网络出版日期: 2020-11-25

网络出版地址: <https://kns.cnki.net/kcms/detail/61.1294.N.20201124.1607.002.html>

基金项目: 国家自然科学基金资助项目(61862040)

第一作者: 贵向泉, 男, 博士, 副教授, 研究方向为云计算及大数据分析、复杂网络理论与应用。E-mail: xqgui@lut.edu.cn

通信作者: 高祯, 男, 硕士生, 研究方向为文本情感分析、自然语言处理。E-mail: 516944911@qq.com

结合文本情感分析技术提取出这些情感,可以反映出民众对待此次疫情的态度,为管理部门准确把握舆情信息、制定相关措施提供有效的数据支持。

近年来,深度神经网络模型凭借在分类问题上的出色性能被应用于文本情感分析任务<sup>[4-6]</sup>。Bengio 等<sup>[7]</sup>建立了神经网络语言模型并提出词向量的概念,实现了词的分布式表征。Mikolov 等<sup>[8]</sup>提出 Word2Vec,通过引入负采样等可行性措施高质量地学习出词向量。Kim 等<sup>[9]</sup>提出将词向量和卷积神经网络(Convolutional Neural Network, CNN)结合,通过卷积神经网络提取文本词向量的局部特征进行文本分类,但是由于 CNN 不具备联想上下文信息的能力,所以循环神经网络(Recurrent Neural Network, RNN)被更多应用于文本序列的处理,以便有效地利用近距离的语义信息<sup>[10-11]</sup>。Wang 等<sup>[12]</sup>使用长短期记忆网络(Long Short-Term Memory, LSTM)对文本中词汇的时序关系进行学习,并解决了 RNN 存在的梯度消失和梯度爆炸问题。双向长短期记忆网络(Bidirectional Long Short-Term Memory, BiLSTM)<sup>[13]</sup>弥补了单向 LSTM 无法有效联系上下文信息的缺陷。Bai 等<sup>[14]</sup>提出了时序卷积网络(Temporal Convolutional Network, TCN),其独特的因果卷积和扩张卷积解决了 LSTM 信息无用就会逐渐被遗忘的问题,并且可以根据一个已知序列出现的先后顺序进行预测。Bahdanau 等<sup>[15]</sup>引入注意力机制,通过对重要特征增加权重来衡量其重要程度。此后,基于 BiLSTM 引入注意力机制(BiLSTM + Attention)<sup>[16]</sup>的研究方法凭借能抓取上下文信息和提取特定的情感特征,以及可以提高情感极性判别的准确率,被更多应用于文本情感分析。

以上研究说明,文本情感分析方法已取得很大进步,但该领域内的大部分神经网络模型仍然存在对上下文信息利用不足的缺陷,以及远距离词汇对当前词汇影响的长期依赖问题。单一的 TCN 网络仅关注单向特征的学习,忽略了下文对上文的语义影响,而如果仅用 BiLSTM 实现双向的语义学习,就会导致对词序信息无法充分的利用。针对这些问题,本文基于深度学习的方法,首次提出融合 TCN 与引入自注意力机制的 BiLSTM 网络的情感模型(简称 TCN-BA 模型),并将其应用于新型冠状病毒疫情期间的网民评论文本情感分析,分析判别与此次疫情相关的文本信息中蕴含的情感倾向。

本文首先通过对因果卷积层进行多次堆叠和扩大卷积的感受野,将 TCN 处理时序问题的优势性

能应用到提取文本序列特征的任务中,充分考虑先出现的词对后出现的词的影响;然后,利用 BiLSTM 的双向循环结构对 TCN 层获取到的文本特征进一步学习,获取文本的上下文联系,同时避免 RNN 存在的长期依赖问题;最后,引入自注意力机制,根据疫情期间民众的关注热点以及引发情绪波动的关键词,对提取到的特征向量进行筛选优化,降低噪声对情感极性判别的影响。最终,有效地提高了情感分析模型的准确性,TCN-BA 模型的链式网络联合的有效性也通过实验各项性能的提高得以证明。

## 1 相关理论

### 1.1 词向量

词向量表示是将微博文本转化为计算机可以理解的向量表示,本文对词向量的训练和生成使用了词嵌入技术,词嵌入技术是将文本中的词转换成数字向量,具体过程就是把一个维数为所有词数量的高维空间嵌入到一个维数低得多的连续向量空间中,每个单词或词组被映射为实数域上的向量。

### 1.2 时序卷积网络

时序卷积网络本质上是对一维卷积进行变形使得其可以处理时序问题的卷积网络,近年来逐渐被应用到文本序列预测任务中。如图 1 所示(图中,  $x_t$  为原始的时序输入,  $y_t$  为网络预测值,  $d$  为扩张卷积中空洞大小),TCN 与普通一维卷积网络主要存在两点不同:

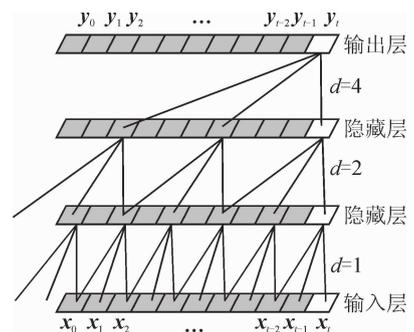


图 1 时序卷积网络结构图

Fig. 1 Network structure of TCN model

1) 扩张卷积(Dilated Convolution)。越到上层,窗口越大,卷积窗口中的空洞越多,在保证每一层隐藏层都和输入序列大小相同的同时,降低计算量增大感受野,使模型可以学习到更长时段内的信息;

2) 因果卷积(Causal Convolution)。因果卷积具备严格的时间约束,只能使用某个时刻之前的信息来预测该时刻的值,限制滑动窗口,可以确保该时

刻之后的信息不会用来预测,同时也保证了之前的信息不会被遗漏。

### 1.3 长短期记忆网络

长短期记忆网络是 RNN 的衍生网络,是针对梯度消失和梯度爆炸问题在 RNN 的基础上加入门限结构改进得来的模型。LSTM 记忆单元结构如图 2 所示,由  $t$  时刻的输入词  $x_t$ 、细胞状态  $C_t$ 、临时细胞状态  $\tilde{C}_t$ 、隐层状态  $h_t$ 、遗忘门  $f_t$ 、记忆门  $i_t$ 、输出门  $o_t$  组成。

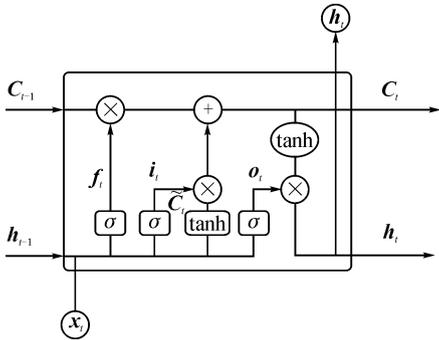


图 2 LSTM 记忆单元模型结构图

Fig. 2 Network structure of LSTM model

LSTM 网络通过上个时刻的隐层状态  $h_{t-1}$  和当前输入  $x_t$  计算出来的遗忘门  $f_t$ 、记忆门  $i_t$ 、输出门  $o_t$  控制输入信息是否被丢弃或传递。计算过程如下:

1) 计算遗忘门,控制选择要遗忘的信息:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

2) 计算记忆门,控制选择要记忆的信息:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

3) 计算当前时刻细胞状态:

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (4)$$

4) 计算输出门和当前时刻隐层状态:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (6)$$

其中,  $W$  为不同门限的权重矩阵,  $b$  为不同门限的偏置矩阵,  $\sigma$ 、 $\tanh$  为激活函数。

通过门限结构控制记忆单元中的输入信息是否被记忆或者遗忘,从而实现了对有用的信息进行传递,而无用的信息则被丢弃,弥补了 RNN 很难获取到文本语义表示的长距离依赖信息的缺陷。

### 1.4 注意力机制

注意力机制的实质是仿效人类的视觉系统,根据一些特定的信息得到视觉焦点,抓取到有关信息的重要特征。一般的注意力机制是通过多个查询(Query)和键值对(Key-Value)组成的映射函数来实现,通过计算 Query 和每个 Key 的相似度得到相

应的权重,再利用 Softmax 函数对权重归一化,并将处理后的权重与对应的键值加权求和得到最终的注意力值,计算过程如下:

1) 计算每一个 Query 和各个 Key 的相关性得到每个 Key 对应 Value 的权重系数:

$$Sim(Q, K_i) = Q K_i^T \quad (7)$$

2) 使用 Softmax 函数对权重系数进行归一化处理:

$$\alpha_i = \text{Softmax}(Sim_i) = \frac{e^{Sim_i}}{\sum_{j=1}^{L_x} e^{Sim_j}} \quad (8)$$

3) 将权重与键值加权求和得到最终的注意力值:

$$A(Q, K, V) = \sum_{i=1}^{L_x} \alpha_i \cdot V_i \quad (9)$$

其中,  $A$  是最终所求的注意力值;  $L_x$  为数据长度;  $Q$  代表查询;  $K$  代表键;  $V$  代表值。

在文本情感分析任务中,注意力机制主要用于表示句子中的词汇和输出之间的相关程度,对文本语义表示中的情感特征增加不同的注意力权重,区分不同的特征对文本信息的重要性,进而提高文本情感分类的准确性。

## 2 TCN-BA 文本情感分析模型

文本情感分析需要考虑文本序列、上下文信息以及情感特征重要程度对最终情感分类的影响,本文提出的 TCN-BA 模型共包含了 6 层结构,如图 3 所示:第一层为输入层,输入微博文本;第二层为文本向量表示层,将输入的文本转化为词向量;第三层为 TCN 网络层,利用 TCN 网络层抓取文本序列特征;第四层为 BiLSTM 网络层,将 TCN 网络层的输出作为 BiLSTM 网络层的输入,对文本的上下文信息进一步学习;第五层为自注意力机制层,引入自注意力机制,生成权重向量,将情感分类重要特征突出表示,并将词汇级的特征拼接为句子级特征用于情感分类;第六层为输出层,使用 Softmax 分类器完成情感分类。

### 2.1 文本向量表示层

本文使用 Word2Vec 中的 Skip-gram 模型(Continuous Skip-gram Model)将词汇转化为词向量,词向量的维度需在训练前定义。Skip-gram 模型根据给定词,对词汇表中的词汇出现在该词上下文概率作出判断,表达式为:

$$P(w_{t+j} | w_t) = \text{Softmax}(v_{w_t}^T v'_{w_{t+j}}) = \frac{\exp(v_{w_t}^T v'_{w_{t+j}})}{\sum_{w' \in V} \exp(v_{w_t}^T v'_{w'})} \quad (10)$$

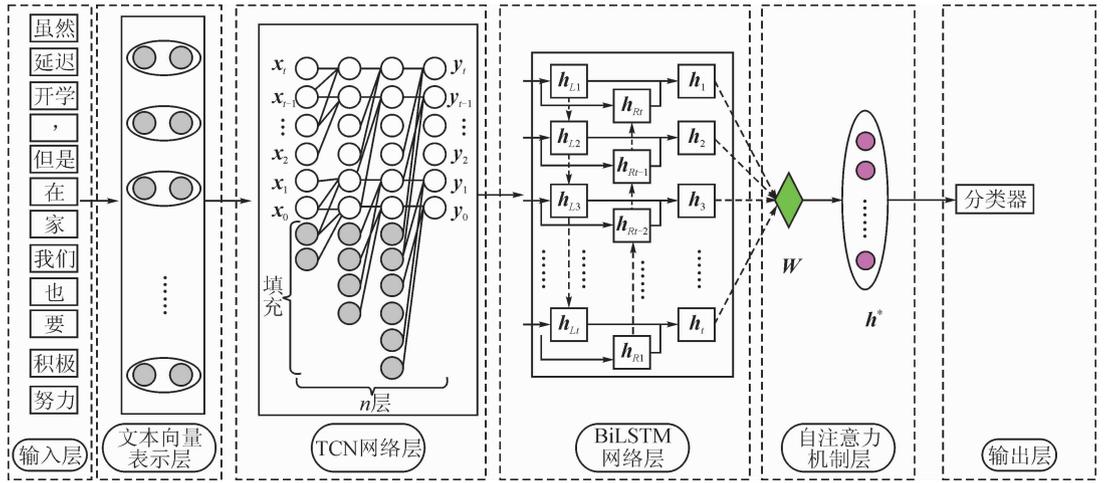


图3 TCN-BA网络模型结构图

Fig. 3 Network structure of TCN-BA model

其中,  $w_i$  代表给定词汇;  $j$  代表与当前输入词汇的距离;  $v_{w_i}$  表示当前时刻输入矩阵中的词向量;  $v'_{w_{i+j}}$  表示输出矩阵中的词向量;  $w' \in v$  表示训练样本词汇  $w'$  包含于词汇表  $v$  中。如果输入句子  $S$  长度为  $n$ , 则可以通过向量矩阵  $S' \in \mathbf{R}^{n \times k}$  ( $k$  为词汇表中的词汇数量) 表示, 而  $S'_i \in \mathbf{R}^k$  则代表句子中第  $i$  个词汇的词向量表示。

由此可见, Skip-gram 模型在考虑到上下文信息的同时, 词向量维度也相对较少, 能够表现出词汇之间的关系, 帮助提取文本中的语法语义信息。

## 2.2 时序卷积网络层

TCN 网络层利用因果卷积限制滑动窗口来实现严格的时间束缚, 利用扩大扩张卷积结构感受野学习到更大范围的信息, 并通过增加残差链接 (Residual Connections) 防止网络层过高导致梯度消失来获取更高层文本序列的语义信息特征。

### 2.2.1 因果卷积

因果卷积使得 TCN 具备很强的因果性质, 当前时刻的信息只能依靠当前时刻之前的信息进行预测, 即只能通过当前的时刻输入  $x_i$  和之前的输入  $x_1, x_2, \dots, x_{i-1}$  进行计算, 在这之后信息不能用于预测, 从左至右进行计算提取特征。

$$p(x) = \prod_{i=1}^T p(x_i | x_1, x_2, \dots, x_{i-1}) \quad (11)$$

### 2.2.2 扩张卷积

由于因果卷积对特征的提取被限定于卷积核的大小范围之内, 添加扩张卷积可以弥补这一缺陷。每隔一定的空洞数对文本序列进行采样计算, 相当于在卷积核内部插入了  $d$  个 0 ( $d$  为空洞数), 添加扩张卷积以后的卷积核大小为:

$$f_{k-d} = (d-1) \times (f_k - 1) + f_k \quad (12)$$

其中,  $f_k$  为当前层的卷积核大小。

### 2.2.3 残差链接

扩张卷积的添加需要增加网络层数、扩大网络层的感受野, 而网络层数过深时, 很容易出现梯度消失的现象。添加残差链接结构可以形成短路链接, 通过对输入  $x$  和经过非线性变换以后的  $F(x)$  求和, 使得时序卷积网络避免了由网络层数增加引起的梯度消失。

$$R = x + F(x) \quad (13)$$

本文所采取的扩张因果卷积模块就是在每一次进行扩张因果卷积计算  $Conv()$  之后将参数层级归一化  $Hi-norm()$ , 然后采用 ReLU 激活函数进行非线性计算, 将得到的结果与输入进行求和, 实现残差链接。计算过程如下:

$$T_i = Conv(W_i \times F_j + b_i) \quad (14)$$

$$\{T_0, T_1, \dots, T_n\} = Hi-norm(\{T_0, T_1, \dots, T_n\}) \quad (15)$$

$$\{T_0, T_1, \dots, T_n\} = Relu(\{T_0, T_1, \dots, T_n\}) \quad (16)$$

其中,  $T_i$  是  $i$  时刻卷积计算得到的状态值;  $W_i$  为  $i$  时刻卷积计算的词的矩阵;  $F_j$  为第  $j$  层的卷积核;  $b_i$  为偏置矩阵;  $\{T_0, T_1, \dots, T_n\}$  是序列经过一次完整的卷积计算后的编码。

TCN 网络层通过堆叠多个扩张因果卷积层, 扩大卷积的感受野, 获取到更完整的文本序列特征, 从而能提取到更深层次的语义信息。

## 2.3 双向长短期记忆网络层

单一的 TCN 对句子进行建模时, 存在无法编码从后到前的信息的问题, 本文加入 BiLSTM 层 (网络结构如图 4 所示), 其隐藏层由正、反两个传播方向的 LSTM 结合而成, 正向传播层单元学习当前文本上文信息, 后向传播层单元学习下文信息, 然后

将两个方向学习到的特征进行拼接,提取到文本序列的上下文信息,获取正反两个传播方向的情感特征。

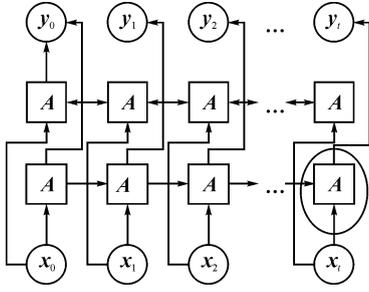


图 4 BiLSTM 网络结构图

Fig. 4 Network structure of BiLSTM model

正向的长短期记忆网络对正向输入的词语计算后可以得到与句子长度相同的隐层状态序列:  $\vec{H} = \{h_{L_1}, h_{L_2}, \dots, h_{L_i}\}$ , 反向的长短期记忆网络则是对逆向输入的词语计算后得到与句子长度相同的隐层状态序列:  $\overleftarrow{H} = \{h_{R_i}, h_{R_{i-1}}, \dots, h_{R_1}\}$ , 最后将正向和反向的隐层状态输出的向量进行拼接, 得到 BiLSTM 网络层的最终输出:

$$\mathbf{H} = \{\vec{H}, \overleftarrow{H}\} \quad (17)$$

## 2.4 自注意力机制层

自注意力机制是注意力机制的一个变形。注意力机制通常用于任务目标与数据源不同的情况(如, 机器翻译), 而自注意力机制则更擅长处理数据源内部、任务目标内部元素的计算。

对文本情感分析而言, 自注意力机制可以更好地捕获到同一个句子之间的语法、语义特征, 对语句中的情感特征重要程度进行区分。自注意力机制并不依赖于外部信息, 它是对自身的数据信息进行训练并不断更新参数, 从而得到注意力值, 即更关注句子之间词汇的关系。

本文将 BiLSTM 网络层生成的向量集合作为自注意力机制的输入, 与自注意力机制层生成的权重向量相乘, 并将词汇级的特征汇总成句子级特征用于最终的情感分类。实现过程如下:

1) 获取 BiLSTM 网络层的输出, 即向量集合  $\mathbf{H}: \{\vec{H}, \overleftarrow{H}\}: [h_1, h_2, \dots, h_i]$ , 生成权重向量:

$$\mathbf{W} = \frac{\mathbf{H} \cdot \mathbf{H}^T}{\sqrt{d_k}} \quad (18)$$

2) 使用 Softmax 函数对生成的权重矩阵做归一化处理, 突出重要特征的权重。权重越大, 表明对微博用户在疫情期间的情感倾向越重要。

$$\alpha = \text{Softmax}(\mathbf{W}) \quad (19)$$

3) 将权重与向量集合  $\mathbf{H}$  相乘, 即得到最终的自注意力值:

$$A(\mathbf{H}) = \alpha \mathbf{H} \quad (20)$$

其中,  $\sqrt{d_k}$  为输入向量的维度, 即词向量维度, 用于防止  $\mathbf{H} \cdot \mathbf{H}^T$  太大导致 Softmax 函数的结果非 0 即 1。最后, 用于微博用户情感分类的句子表示为:

$$\mathbf{h}^* = \tanh(A(\mathbf{H})) \quad (21)$$

## 2.5 模型训练

TCN-BA 模型在训练过程中采取有监督学习。本文以训练集中每个疫情文本的用户情感极性值  $y$  与模型计算出来的预测输出值  $\hat{y}$  之间的交叉熵误差作为损失函数, 目标是 minimized 训练样本值与实际样本值之间的交叉熵, 设定交叉熵损失函数为:

$$L = -\frac{1}{N} \sum_x [y \ln \hat{y} + (1 - y) \ln(1 - \hat{y})] \quad (22)$$

其中,  $L$  为损失值;  $x$  为样本;  $N$  为样本总数。采取的是 Mini-batch 梯度下降优化算法, 相较于其他算法, 该法只需对损失函数一阶求导, 计算代价相对较小的同时, 也提高了模型的训练速度。

## 3 实验与分析

### 3.1 实验数据及环境搭建

实验数据来源于新浪微博(Weibo)的用户微博文本。新浪微博是中国最大的社交媒体平台之一, 其内容具有很高的及时性和开放性, 对于一些突发事件(如此次新冠疫情), 可以提供大量的研究数据。

本实验依据“新冠肺炎”、“冠状病毒”、“疫情”、“传染”、“武汉”等关键词, 利用爬虫技术对 2020 年 1 月 24 日至 2020 年 2 月 24 日期间的相关微博文本进行爬取, 共获得原发微博 87 563 条, 实验数据统计信息如表 1 所示。由于采集的数据存在噪声, 会对最终的实验结果产生影响, 本文对数据进行了筛选和预处理:

1) 字体转换(繁体字转换成简体字);

2) 分词和停用词去除(把中文的汉字序列切分成有意义的词, 并适当减少电脑检索中的虚字和非检索用字);

3) 特殊符号替换(使用正则匹配的方法将 @ 符号、URL 分别转换为 <NAME>、<URL> 等标准形式);

4) 无效微博的过滤(文本内容中词数少于 5 个, 全部都是重复符号、字母、汉字的情况, 微博中的广告和推广性的内容等);

5) 数据格式的标准化。

表 1 实验数据统计信息

Tab. 1 Experimental data statistics

数据名称	数据格式
微博 ID	整型
微博发布时间	xx 月 xx 日 xx : xx
发布人账号	字符串
中文内容	字符串
情感倾向	{1,0,-1}

微博数据经过初步处理后,根据 BosonNLP 情感词典(情感词汇:积极情感词、消极情感词、积极评价词以及消极评价词;程度副词:控制情感词的强弱;否定词汇:判断情感极性是否发生转变)进行人工标注,如表 2 实验数据样例所示,情感标签分为三类:-1,消极;0,中性;1,积极。根据情感词典进行情感极性的标注,在一定程度上可以避免主观性判断出现的偏差。标注后对数据作进一步欠采样处理,保留了 68 657 条微博文本,使得三种情感倾向的比例趋近 1:1:1,保证了数据分布平衡。将标注好的数据作为实验数据集,按照 8:1:1 的比例划分成训练集、验证集及测试集进行实验。

表 2 实验数据样例

Tab. 2 Sample experimental data

情感倾向	标签	微博文本
消极	-1	最近/新型/冠状病毒/事情/真/太/可怕/这/几天/门诊/班/上/我/心态/爆炸
中性	0	公布/疫情/是/对/不过/要/注意/方式/方法
积极	1	虽然/延迟/开学/但是/在家/我们/也/要/积极/努力/中国/加油/武汉/加油

在标注数据的过程中,发现微博用户的情感倾向分布与年龄、性别之间呈现出一定的关联(如图 5 所示):多数 25 岁以下的用户在面对疫情时更为积极,50 岁以上的用户其消极评论占比更大,25~50 岁的用户其评论文本则相对更为中性;在性别方面,男性和女性之间的差异也比较明显,男性用户中的中性文本和积极文本相对于女性分别高出 7.5% 和 15.2%,消极文本中女性用户则占到了 61.8%,可见,男性在面对疫情时,其评论文本更为理性和积极。

本文在 PyTorch 深度学习框架下进行实验,具体实验环境如表 3 所示。

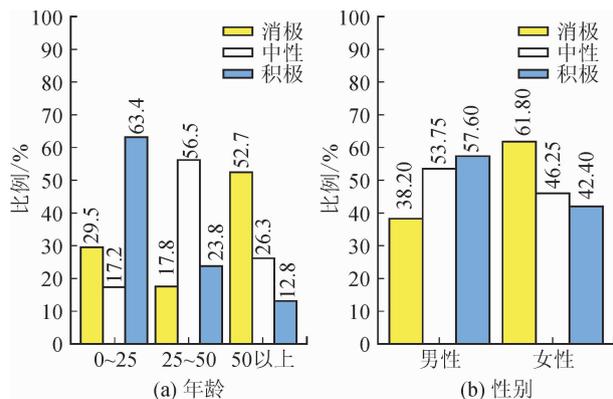


图 5 各类情感倾向在不同年龄和性别之间的分布  
Fig. 5 Distribution of various emotional tendencies among different ages and genders

表 3 实验环境搭建

Tab. 3 Experimental environment construction

实验环境	具体信息
操作系统	WIN10
CPU	Intel CoreTM i7-7700 CPU@3.60 GHz
内存	16 GB
显卡	GTX 1080 Ti
开发框架	PyTorch 1.2.0
开发语言	Python 3.7.7

### 3.2 评价指标

本文采用准确率、召回率以及  $Macro-F1$  值作为评估模型有效性的指标。计算过程为:

$$\text{准确率: } P = \frac{TP}{TP + FP} \quad (23)$$

$$\text{召回率: } R = \frac{TP}{TP + FN} \quad (24)$$

$$\text{F1 值: } F1 = \frac{2 \times P \times R}{P + R} \quad (25)$$

其中,  $TP$  表示真阳例;  $FP$  表示假阳例;  $FN$  表示假阴例。准确率  $P$  越大,代表模型预测准确度越高;对每一个类别分别计算  $F1$  值,求平均即得到  $Macro-F1$  值,其值越大,表示模型分类效果越好。

### 3.3 对照实验设置

在相同的实验环境设置了 6 组对照实验,用于验证本文提出的 TCN-BA 模型的有效性。除 BERT 模型外,其余 5 组实验的词向量均由 Word2Vec 中的 Skip-gram 模型训练得来。

1) LSTM:在一般 RNN 的基础上添加了门限结构,选择性提取信息<sup>[17]</sup>。

2) BiLSTM:正向 LSTM 和反向 LSTM 分别从正、反两个方向对文本语义特征进行学习,并将得到的特征进行融合,用于最终的情感极性判断<sup>[18]</sup>。

3) BiLSTM+Attention:在 BiLSTM 网络中引

入注意力机制,由 BiLSTM 提取文本序列的上下文信息,注意力机制负责区分所提取特征对最终情感分类的重要程度<sup>[19-20]</sup>。

4) TCN:对于单一的时序卷积网络,设置多层因果卷积及扩张卷积对文本序列进行计算,将最后时刻状态视为最终语义表示用于情感分类。

5) BERT:利用 Transformer 的双向编码表示对基于架构微调的方法进行改进,通过联合调节所有层中的上下文来预先训练深度双向表示<sup>[21]</sup>。

6) TCN-BA:即本文模型。

### 3.4 可调参数设置

TCN-BA 模型在实验中使用 Word2Vec 中的 Skip-gram 模型训练词向量,维度设置为 300,学习率为 0.03;TCN 网络层中扩张因果卷积层数为 4,卷积核大小设置为 7,空洞因子为  $2n$ ,使用 ReLU 函数作为激活函数;BiLSTM 网络层中隐藏层数设置为 128,隐藏层有 64 个神经元;Dropout 参数分别设置为 0.3、0.3、0.25;Adam 优化器学习率为 0.002;训练批次长度 *batch\_size* 设置为 64。

### 3.5 实验结果分析

本文在疫情期间微博用户文本数据集上进行了 6 组对照实验,以模型在数据集上的准确率、召回率和 *Macro-F1* 值为评价标准,具体实验结果如表 4、图 6 所示。

表 4 对照实验结果

Tab. 4 Comparative experimental results

评价标准	准确率/%	召回率/%	<i>Macro-F1</i>
LSTM	83.17	84.24	0.765 1
BiLSTM	85.96	86.13	0.784 0
BiLSTM+Attention	89.43	88.73	0.812 3
TCN	90.29	89.98	0.825 6
BERT	91.16	90.84	0.837 9
TCN-BA	92.68	91.53	0.839 5

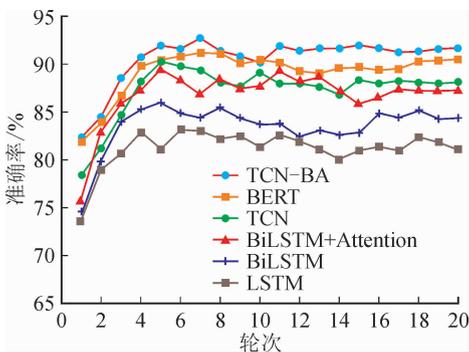


图 6 6 种模型准确率对比图

Fig. 6 Comparison of accuracy of 6 models

1) 相对于单一的 LSTM, BiLSTM 增加了后向传播单元,可以抓取到下文对当前词汇的影响,融合了上下文信息特征,效果更优;

2) 相较于单独的 BiLSTM 网络模型, BiLSTM+Attention 网络模型在引入注意力机制以后,能更好地捕捉到句子的局部特征,区分情感特征重要程度,其准确率、召回率以及 *Macro-F1* 值分别提高了 3.47%、2.60%、0.028 3。

3) TCN 由于利用因果卷积、扩张卷积可以获得更多的文本序列特征,并且残差链接结构可以抑制网络层数过多而引起的梯度消失现象,因此其实验效果优于 BiLSTM+Attention 模型。

4) BERT 使用遮蔽语言模型来实现预训练的深度双向表示,由于在所有层中共同依赖于左右上下文,使得 BERT 模型在文本情感分析任务中可以较好地提取到文本序列中的情感特征。

5) TCN-BA 模型结合 TCN 学习文本序列特征,增加 BiLSTM 网络层补足单向 TCN 无法联系上下文的缺陷,并利用自注意力机制优化特征向量。对比 TCN-BA 和其他 5 组模型的实验结果可知,TCN-BA 模型的准确率和召回率都得到了有效提升,分别达到 92.68%、91.53%,*Macro-F1* 值达到 0.839 5。

### 3.6 模型性能分析与对比

除了对 6 种模型的准确率、召回率以及 *Macro-F1* 值进行对比,本文还在实验过程中对各个模型的性能进行了比较,如图 7 所示。

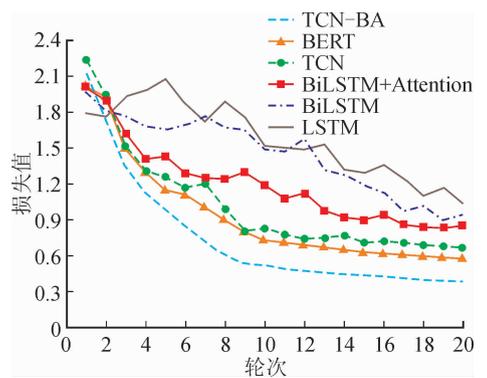


图 7 6 种模型损失对比图

Fig. 7 Comparison of loss of 6 models

通过对比发现,由于 BERT 模型预训练过程中的标记在实际预测中不会出现,过多使用标记往往会影响模型表现,并且每个批次只有 15% 的符号(包含字符和标点)被预测,这在一定程度上影响了收敛速度。相对于 BERT 模型,TCN-BA 融合模型收敛速度更快,在疫情期间微博文本数据集上的性

能表现总体上更为优秀。由图7可知,单一的时序卷积网络模型、BiLSTM + Attention 模型以及BERT模型的损失值下降到稳定值的速度尽管比TCN-BA模型慢,但在最后的几次迭代中也逐渐趋于收敛。

#### 4 结 语

本文首次提出TCN-BA模型,并将其应用于疫情期间的微博文本情感分析。首先,通过TCN网络层提取文本的序列特征;然后,将其输出作为BiLSTM网络层的输入,学习文本的上下文信息和情感特征,得到文本的语义表示;最后,引入自注意力机制,结合上下文信息与疫情期间热点词汇特征,优化特征向量,减少噪音干扰。6组对照实验证明了TCN-BA模型在疫情期间微博文本数据集上进行情感分析的有效性。在总结实验期间,发现时序卷积网络的网络结构在理论上可以对文本特征进行双向学习,提取文本的上下文信息。因此,今后的研究将会聚焦于如何对TCN的网络结构进行调整来进一步提高文本情感分析准确度的问题上。

#### 参考文献:

- [1] BAKSHI R K, KAUR N, KAUR R, et al. Opinion mining and sentiment analysis[C]//2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom). IEEE, 2016: 452-455.
- [2] 何炎祥,孙松涛,牛菲菲,等.用于微博情感分析的一种情感语义增强的深度学习模型[J].计算机学报,2017,40(4):773-790.  
HE Yanxiang, SUN Songtao, NIU Feifei, et al. A deep learning model enhanced with emotion semantics for microblog sentiment analysis[J]. Chinese Journal of Computers, 2017, 40(4): 773-790.
- [3] 姜杰.社交媒体文本情感分析[D].南京:南京理工大学,2017.  
JIANG Jie. Social media text sentiment analysis[D]. Nanjing: Nanjing University of Science and Technology, 2017.
- [4] RUDER S, GHAFFARI P, BRESLIN J G. A hierarchical model of reviews for aspect-based sentiment analysis[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016.
- [5] 李然.基于深度学习的短文本情感倾向性研究[D].北京:北京理工大学,2015.  
LI Ran. Sentiment analysis of short text with deep neural networks[D]. Beijing: Beijing Institute of Technology, 2015.
- [6] 梁军,柴玉梅,原慧斌,等.基于极性转移和LSTM递归

网络的情感分析[J].中文信息学报,2015,29(5):152-159.

LIANG Jun, CHAI Yumei, YUAN Huibin, et al. Polarity shifting and LSTM based recursive networks for sentiment analysis[J]. Journal of Chinese Information Processing, 2015, 29(5): 152-159.

- [7] BENGIO Y, DUCHARME R, VINCENT P, et al. A neural probabilistic language model[J]. Journal of Machine Learning Research, 2003, 3(6): 1137-1155.
- [8] MIKOLOV T, CHEN K, CORRADO G, et al. Distributed representations of words and phrases and their compositionality[J]. Advances in Neural Information Processing Systems, 2013, 26:3111-3119.
- [9] KIM Y. Convolutional neural networks for sentence classification [J]. arXiv preprint arXiv: 1408.5882, 2014.
- [10] EBRAHIMI J, DOU D. Chain based RNN for relation classification[C]//Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2015: 1244-1249.
- [11] MNIH A, HINTON G E. A scalable hierarchical distributed language model[C]//Advances in Neural Information Processing Systems. 2009: 1081-1088.
- [12] WANG X, LIU Y, SUN C J, et al. Predicting polarities of tweets by composing word embeddings with long short-term memory[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. 2015, 1: 1343-1353.
- [13] CHEN T, XU R, HE Y, et al. Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN[J]. Expert Systems with Applications, 2017, 72: 221-230.
- [14] BAI S, KOLTER J Z, KOLTUN V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling[J]. arXiv preprint arXiv:1803.01271, 2018.
- [15] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint arXiv:1409.0473, 2014.
- [16] 林原,李家平,许侃,等.基于多头注意力的双向LSTM情感分析模型研究[J].山西大学学报(自然科学版),2020,43(1): 1-7.  
LIN Yuan, LI Jiaping, XU Kan, et al. Sentiment analysis with multi-head attention-based Bi-LSTM model[J]. Journal of Shanxi University (Natural Science Edition), 2020, 43(1): 1-7.
- [17] HOCHREITER S, SCHMIDHUBER J. Long short-

- term memory[J]. *Neural Computation*, 1997, 9(8): 1735-1780.
- [18] 卢强,朱振方,徐富永,等. 融合语法规则的 Bi-LSTM 中文情感分类方法研究[J]. *数据分析与知识发现*, 2019,3(11):99-107.  
LU Qiang, ZHU Zhenfang, XU Fuyong, et al. Chinese sentiment classification method with Bi-LSTM and grammar rules[J]. *Data Analysis and Knowledge Discovery*, 2019, 3(11): 99-107.
- [19] 翟社平,杨媛媛,邱程,等. 基于注意力机制 Bi-LSTM 算法的双语文本情感分析[J]. *计算机应用与软件*, 2019,36(12):251-255.  
ZHAI Sheping, YANG Yuanyuan, QIU Cheng, et al. Bilingual text sentiment analysis based on attention mechanism Bi-LSTM [J]. *Computer Applications and Software*, 2019, 36(12): 251-255.
- [20] 王丽亚,刘昌辉,蔡敦波,等. 基于 CNN-BiLSTM 网络引入注意力模型的文本情感分析[J]. *武汉工程大学学报*, 2019,41(4):386-391.  
WANG Liya, LIU Changhui, CAI Dunbo, et al. Text sentiment analysis based on CNN-BiLSTM network and attention model[J]. *Journal of Wuhan Institute of Technology*, 2019, 41(4):386-391.
- [21] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[J]. *arXiv preprint arXiv: 1810.04805*, 2018.

(责任编辑 周 蓓)

(上接第 82 页)

- [3] 毛金晖,高建丰,邱云钦,等. 海底管道原油油气燃烧爆炸实验与数值仿真模拟[J]. *中国水运(下半月)*, 2017, 17(6):101-102,128.
- [4] 蔡运雄,杜扬,王世茂,等. 封闭管道油气爆炸超压及火焰传播特性[J]. *中国安全生产科学技术*, 2019, 15(6): 61-66.  
CAI Yunxiong, DU Yang, WANG Shimao, et al. Overpressure and flame propagation characteristics of gasoline-air mixture explosion in closed tube[J]. *Journal of Safety Science and Technology*, 2019, 15(6): 61-66.
- [5] 高建丰,杜扬,蒋新生,等. 油气爆炸冲击载荷对金属油罐断裂损伤的模拟实验研究[J]. *中国储运*, 2016(3): 129-134.
- [6] 陈超,周利坤. 汽车油罐车爆炸燃烧特性的数值模拟分析[J]. *安全与环境工程*, 2016, 23(6): 148-154, 159.  
CHEN Chao, ZHOU Likun. Numerical simulation analysis on explosion and combustion characteristics of oil tank trucks[J]. *Safety and Environmental Engineering*, 2016, 23(6): 148-154, 159.
- [7] 刘冲,杜扬,梁建军,等. 含双侧分支受限空间油气爆炸火焰行为与超压特性大涡模拟[J]. *爆炸与冲击*, 2020, 40(6):83-92.  
LIU CHONG, DU Yang, LIANG Jianjun, et al. Large eddy simulation of gasoline/air mixture explosion in a semi-confined space with bilateral branches[J]. *Explosion and Shock Waves*, 2020, 40(6): 83-92.
- [8] 江丙友,林柏泉,朱传杰,等. 瓦斯爆炸冲击波在采煤工作面巷网中传播特性的数值模拟[J]. *煤炭学报*, 2011, 36(6):968-972.  
JIANG Bingyou, LIN Baiquan, ZHU Chuanjie, et al. Numerical simulation on propagation characteristics of gas explosion shock waves in coalface tunnel network [J]. *Journal of China Coal Society*, 2011, 36(6): 968-972.
- [9] 高佳丽. 不同类型管道内瓦斯爆炸冲击波传播特性数值模拟研究[D]. 阜新: 辽宁工程技术大学, 2017.  
GAO Jiali. Numerical simulation of gas explosion shock wave propagation characteristics in different types of pipelines [D]. Fuxin: Liaoning Technical University, 2017.

(责任编辑 王绪迪)