

引用格式:王丽亚,刘昌辉,蔡敦波,等. 基于字符级联合网络特征融合的中文文本情感分析[J]. 微电子学与计算机,2020,37(1):80-86. [WANG L Y,LIU C H,CAI D B,et al. Chinese text sentiment analysis based on joint network and attention model [J]. Microelectronics & Computer,2020,37(1):80-86.]

基于字符级联合网络特征融合的中文文本情感分析

王丽亚,刘昌辉,蔡敦波,赵彤洲,王 梦
(武汉工程大学 计算机科学与工程学院,湖北 武汉 430205)

摘 要: 针对传统卷积神经网络(CNN)同层神经元之间信息不能互传,无法充分利用同一层次上的特征信息,以及无法提取长距离上下文相关特征的问题.该文针对中文文本,提出字符级联合网络特征融合的模型进行情感分析,在字符级的基础上采用 BiGRU 和 CNN-BiGRU 并行的联合网络提取特征,利用 CNN 的强学习能力提取深层次特征,再利用双向门限循环神经网络(BiGRU)进行深度学习,加强模型对特征的学习能力.另一方面,利用 BiGRU 提取上下文相关的特征,丰富特征信息.最后在单方面上引入注意力机制进行特征权重分配,降低噪声干扰.在数据集上进行多组对比实验,该方法取得 92.36% 的 F1 值,结果表明本文提出的模型能有效的提高文本分类的准确率.

关键词: 卷积神经网络;BiGRU;注意力机制;中文文本情感分析

中图分类号: TP391

文献标识码: A

文章编号: 1000—7180(2020)01—0080—07

Chinese text sentiment analysis based on joint network and attention model

WANG Li-ya,LIU Chang-hui,CAI Dun-bo,ZHAO Tong-zhou,WANG Meng

(College of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan 430205, China)

Abstract: For the traditional convolutional neural network (CNN), the information between the same layer of neurons cannot be transmitted to each other, and the feature information at the same level cannot be fully utilized, and the problem of long-distance context-related features cannot be extracted. In this paper, based on the Chinese text, this paper proposes a model of character-level joint network feature fusion for sentiment analysis. Based on the character level, BiGRU and CNN-BiGRU parallel joint network are used to extract features, and CNN's strong learning ability is used to extract deep features and reuse. The two-way threshold cyclic neural network (BiGRU) performs deep learning and enhances the model's ability to learn features. On the other hand, BiGRU is used to extract context-related features to enrich feature information. Finally, the attention mechanism is introduced unilaterally to perform feature weight distribution to reduce noise interference. Multi-group comparison experiments were carried out on the dataset. The method obtained 92.36% F1 value. The results show that the proposed model can effectively improve the accuracy of text classification.

Key words: convolutional neural network;BiGRU;attention;Chinese text sentiment analysis

1 引言

随着 Web2.0 的飞速发展,越来越多的人习惯

在移动终端购物,然后对商品进行评价,其中包含大量的文本情感信息.这类带有情感倾向的非结构化文本,数量成指数形式增加,如何高效、快速的了解

文本所包含的情感是用户选择消费和商业界决策的辅助^[1]. 因此, 文本情感分析^[2]是自然语言处理(NLP)领域的一个重要研究方向.

文本情感分析主要目的是从原始文本中提取出评论人的主观情感. 传统的机器学习方法对人工特征构建依赖性强, 导致特征工程人力浩大, 从而人为因素干扰大^[3]. 为解决这一问题, 深度学习的方法被提出. 深度学习^[4]是机器学习的一个重要分支, 近年来, 深度学习在文本分类上广泛应用^[5-10]. 传统 CNN 模型通过卷积和池化操作可以提取一系列更高级别的短语特征, 但缺乏句子体态的特征表示, 以及对原始文本缺乏对上下文语义特征的理解能力, 因此本文利用 CNN-BiGRU 和 BiGRU 并行的联合网络来进行特征提取.

2 相关研究

传统 RNN^[11-12]可以有效地学习近距离的语义特征, 但存在梯度消失或梯度弥散问题. 因此, RNN 出现多个变种循环神经网络模型, 长短期记忆网络(LSTM, Long Short-Term Memory)^[13-14]可以提取长距离的上下文语义特征, 但结构复杂, 时间代价大. GRU 网络^[15]结构相对 LSTM 而言简单些, 能有效的提高模型训练时间, 双向门限循环单元(BiGRU)^[16]则是由两个 GRU 组合而成, 一个前向传播、一个后向传播. 后为提高分类效果, 出现引入注意力机制(Attention)^[17-20]. 皆证明了通过引入注意力机制有效地提高了情感分类的准确度. 但都是以单 CNN、LSTM、BiGRU 模型的形式进行特征提取. 但模型在单方面上, 都是以单模型的形式来构建神经网络^[21-23], 对 CNN 中池化层的特征信息缺少利用.

3 字符级联合网络特征融合模型

本文提出字符级联合网络特征融合模型(CB-GA-BGA)进行情感分析. 模型以 CNN-BiGRU 和 BiGRU 并行的联合网络来进行特征提取, 利用注意力机制筛选特征, 最后经过特征融合丰富特征信息.

3.1 字符级词向量

word2vec 技术包括两种模型: 连续词袋模型(CBOW), Skip-gram 模型. CBOW 模型是根据上下文预测一个词, 而 Skip-gram 模型是利用当前词来预测周围的词. 本文使用 Skip-gram 模型. Skip-gram 模型分为输入层、投影层、输出层, 模型结构如图 1 所示. 如果上下文窗口大小设置为 4, 且设当前字 $W(t)$ 所对应的向量形式为 $V(W(t))$, 其周围 4

个字所对应的向量形式为 $V(W(t+2))$ 、 $V(W(t+1))$ 、 $V(W(t-1))$ 、 $V(W(t-2))$, 则 Skip-gram 模型通过中间字预测周围字, 是利用中间字向量 $V(W(t))$ 的条件概率值来求解, 如公式(1)所示.

$$P(V(W(i)) | V(W(t))) \quad (1)$$

式中, $V(W(i)) \in \{V(W(t+2)), V(W(t+1)), V(W(t-1)), V(W(t-2))\}$.

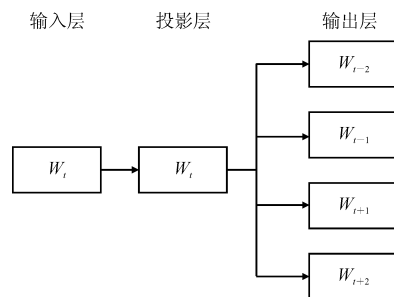


图 1 Skip-gram 模型

3.2 CNN-BiGRU 深度学习局部特征

利用 CNN 提取局部特征, CNN 的网络结构如图 2 所示. 再利用 BiGRU 学习 CNN 池化后的特征, 实现模型对局部特征的深度学习, 最后添加 Attention 进行特征筛选, 使模型学习到有效的特征.

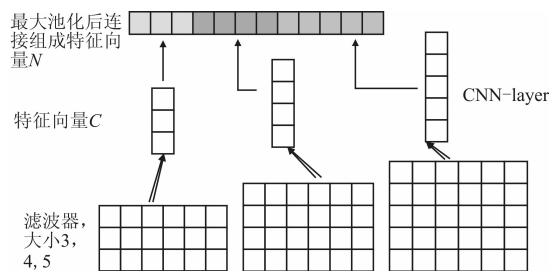


图 2 CNN 模型

(1) 将嵌入层的输出作为输入, 句子中每个 x 字向量为 x_i , $x_i \in \mathbf{R}^{n \times d}$, 其中 n 是字数, d 是向量维度, 本文定为 100 维.

(2) 卷积操作, 主要完成特征提取的工作, 通过设置好的滤波器来完成对输入文本句子特征的提取. 如公式(2) 所示.

$$c_i = f(\omega \times x_{i:i+h-1} + b) \quad (2)$$

式中, ω 为卷积核, h 为卷积核的尺寸, $x_{i:i+h-1}$ 为 i 到 $i+h-1$ 个词组成的句子向量, b 为偏置项, 通过卷积层后, 得到特征矩阵 $\mathbf{C}$, $\mathbf{C} = [c_1, c_2, \dots, c_{n-h+1}]$.

(3) 池化层, 通过对卷积层之后得到的句子局部特征矩阵 $\mathbf{C}$ 进行下采样, 求得局部值的最优解. 这里采用 MaxPooling 技术, 如公式(3) 所示.

$$M = \max(c_1, c_2, \dots, c_{n-h+1}) = \max\{\mathbf{C}\} \quad (3)$$

(4) 由于 BiGRU 输入必须是序列化结构, 池化将中断序列结构 C , 所以需要添加全连接层, 将池化层后的 M_i 向量连接成向量 N , 如公式(4) 所示.

$$N = \{M_1, M_2, \dots, M_n\} \quad (4)$$

(5) 将新的连续高阶窗口 N 作为 BiGRU 的输入.

(6) 利用 BiGRU 网络学习特征 N . BiGRU 由正向 GRU、反向 GRU、正反向 GRU 的输出状态连接层组成的神经网络. GRU 的原理图如图 3 所示. 计算过程如公式(5~8) 所示. BiGRU 的网络结构如图 4 所示. 若记 t 时刻正向 GRU 输出的隐藏状态为 $\vec{h}_t$, 反向 GRU 输出的隐藏状态为 $\overleftarrow{h}_t$, 则 BiGRU 输出的隐藏状态 h_t , 其具体的计算过程如公式(9~11) 所示.

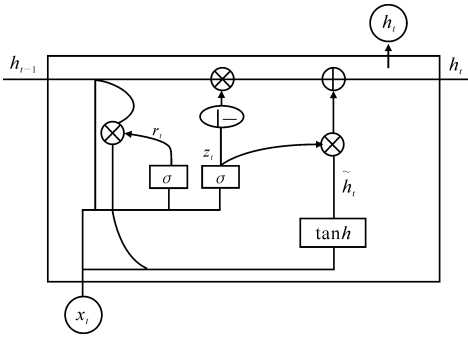


图 3 GRU 模型

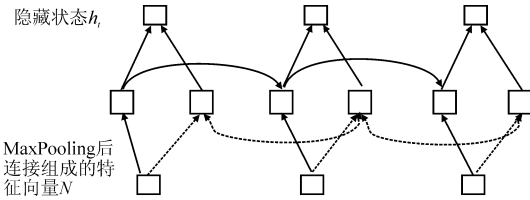


图 4 BiGRU 模型

$$z_t = \sigma(W_z \cdot [h_{t-1}, N_t]) \quad (5)$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, N_t]) \quad (6)$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, N_t]) \quad (7)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (8)$$

式中, W_z, W_r, W 为权值矩阵; z_t 为更新门; r_t 为重置门; $\tilde{h}_t$ 为备选激活函数; h_t 为激活函数; N_t 为 t 时刻 GRU 的输入; $\tanh$: $\tanh$ 激活函数; σ : σ sigmoid 激活函数.

$$\vec{h}_t = \text{GRU}(\vec{h}_{t-1}, N_t) \quad (9)$$

$$\overleftarrow{h}_t = \text{GRU}(\overleftarrow{h}_{t-1}, N_t) \quad (10)$$

$$h_t = W_t \vec{h}_t + V_t \overleftarrow{h}_t + b_t \quad (11)$$

式中, W_t, V_t 为权值矩阵; GRU 为单 GRU 函数; b_t 为偏置向量.

3.3 BiGRU-Attention 筛选特征

注意力机制通过对特征值分配不同的注意力权重, 实现模型对 BiGRU 网络处理后的特征进行筛选. 由于联合网络最后一层的结构皆是通过 BiGRU 处理后的特征, 类型相似, 所以可使用相同的注意力机制. 本文使用前馈注意力模型, 注意力模型结构图如图 5 所示.

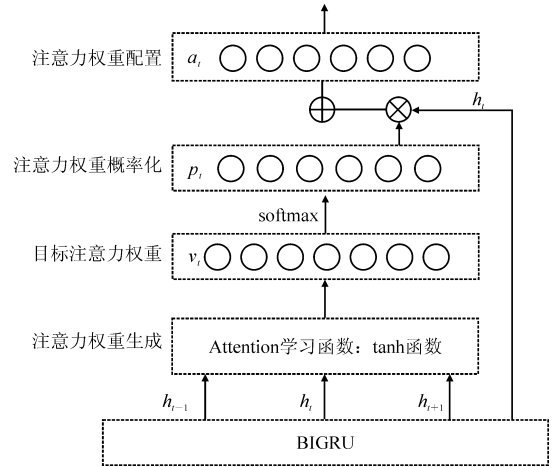


图 5 注意力机制

(1) 生成目标注意力权重 v_t , σ 是一种注意力函数 $\tanh$, 如公式(12) 所示.

$$v_t = \sigma(h_t) \quad (12)$$

(2) 注意力权重概率化, 通过 softmax 函数生成概率向量 p_t , 如公式(13) 所示.

$$p_t = \frac{\exp(v_t)}{\sum_{t=1}^m \exp(v_t)} \quad (13)$$

(3) 注意力权重配置, 将生成的注意力权重配置给对应的隐层状态语义编码 h_t , 使模型生成的注意力权重发挥作用, a_t 是 h_t 的加权平均值, 权值是 p_t , 如公式(14) 所示.

$$a_t = \sum_{t=1}^m p_t \cdot h_t \quad (14)$$

3.4 CBGA-BGA 模型

CBGA-BGA 模型网络结构如图 6 所示. 其主要包括两个部分: 一是 BiGRU 和 CNN-BiGRU 并行的联合网络构建. 二是引入注意力机制.

(1) 将文本输入到 word2vec 模型中训练, 句子中每个字向量为 x_i , $x_i \in R^{n \times d}$, 其中 n 是字数, d 是向量维度, 则句子矩阵 S 表示为:

$$S = \{x_1, x_2, \dots, x_n\}$$

(2) CNN-BiGRU-Attention 网络的构建

(3) 添加 CNN 层, 得到句子局部特征矩阵 C .

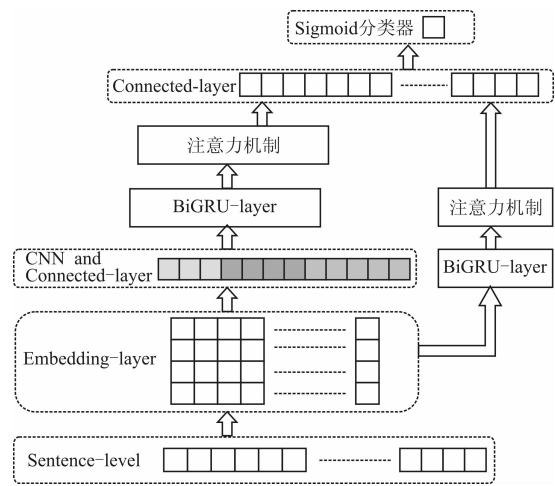


图 6 CBGA-BGA 模型

- (4)利用 MaxPooling 技术,得到矩阵 C 的局部最优解 M .
- (5)经过全连接层,将局部特征矩阵的最优解 M 拼接成新的矩阵 N .
- (6)添加 BiGRU 层,深度学习特征 N ,得到新的特征 H_{ca} .
- (7)引入注意力机制,进行重点特征的筛选,得到 H_{ca} 的加权平均值 A'_{ca} .
- (8)经过 dropout 层后,得到特征向量 A'_{ca} .
- (9)BiGRU-Attention 网络的构建
- (10)添加 BiGRU 层,提取文本上下文语义信息,得到特征 H_a .
- (11)引入注意力机制,进行重点特征的筛选,得到 H_a 的加权平均值 A_a .
- (12)经过 dropout 层后,得到特征向量 A'_a .
- (13)增加全连接层,将所得特征矩阵 A'_{ca} 、 A'_a 拼接成句子最终的特征矩阵 A .
- (14)添加 Dense 层,参数为 1,激活函数为 sigmoid 函数,对文本特征进行分类:

$$p(y \mid \mathbf{A}, \mathbf{W}, b) = \text{sigmoid}(\mathbf{WA} + b)$$

(15)模型损失函数定为对数损失函数,模型会通过更新参数权值矩阵 $\mathbf{W}$ 和偏置向量 b ,从而达到优化模型的效果.

3.5 模型训练

本文将情感分析问题看成一个二分类的问题,分类函数选为 sigmoid 函数,如公式(15)所示.

$$p(y = 1 \mid x, \omega) = h_{\omega}(x) = g(\omega^T x)$$
$$= 1/(1 + \exp(-\omega^T x)) \quad (15)$$

式中,样本是 $\{x, y\}$; y 是 0 或者 1;表示正类或者负类; x 是样本特征向量; ω 是可训练参数.

模型训练的目标实质就是最小化损失函数. 本文在编译模型时,损失函数为对数损失函数,一般是与 sigmoid 相对应的损失函数. 训练模型参数 ω : y_i 为输入 x_i 的真实类别, $h_{\omega}(x_i)$ 为预测输入 x_i 属于类别 1 的概率. 如公式 16 所示.

$$L(\omega) = -\frac{1}{N} \sum_{i=1}^N (y_i \log h_{\omega}(x_i) + (1 - y_i) \log(1 - h_{\omega}(x_i))) \quad (16)$$

模型的优化器选为 Adam, Adam 优化算法是一种计算每个参数的自适应学习率的方法,结合了 AdaGrad 和 RMSProp 两种优化算法的优点. 参数的更新不受梯度的伸缩变换影响,变化平稳.

4 实验

实验环境配置数据如表 1 所示.

表 1 实验环境配置

实验环境	配置数据
语言	Python3. 6
工具	Jupyter notebook
框架	Keras2. 1. 5
处理器	CPU

4.1 实验数据

数据来自中文网络购物的评论,共有 21 105 条,数据内容对象类型有酒店、牛奶、书籍、手机等. 评论情感标签分为两类 $[0, 1]$,其中消极的情感为 0,积极的情感为 1. 数据集设置如表 2 所示.

表 2 实验数据设置

数据集	训练集	验证集	测试集	总数
Data	16 884	2 000	2 221	21 105

4.2 参数设置

参数设置会直接影响后续模型的分类效果. 具体的参数设置如表 3 所示.

表 3 参数设置

参数名	参数值
Vec_window	7
Vec_dim	100
Cnn_win	3, 4, 5
Cnn_win_num	128
Cnn_activation	relu
Pool_size	4
BiGRU_layers	2
BiGRU_dim	100
Dropout	0. 5
lr	0. 001
Bach_size	100
epochs	10

4.3 评价指标

设总的测试集个数为: TP+TN+FP+FN. 其具体含义如表 4 所示.

表 4 评价指标相关参数含义

实际值 \ 预测值	Positive	Negative
	TP	FN
Positive	TP	FN
Negative	FP	TN

Accuracy 是对模型正确分类能力的评估, 准确度越高, 模型的分类能力越好, Precision 是精确率, Recall 是对查全率的评估, F 值中 θ 常用值为 1, 是综合评价指标. 如公式(17)~(20)所示.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (17)$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (18)$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \quad (19)$$

$$F_1 = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (20)$$

4.4 对比实验设置

(1)BiLSTM^[24]: 单一的 BiLSTM, 网络输入是利用 word2vec 训练得词向量.

(2)BiGRU^[16]: 单一的 BiGRU, 网络输入是利用 word2vec 训练得词向量. 证明在本文所用数据集上单模型 BiGRU 优于 BiLSTM.

(3)CNN^[25]: 单一的 CNN, 网络输入是利用 word2vec 训练得词向量.

(4)CNN+BiGRU: 先添加一个 CNN, 再添加一个 BiGRU, 网络输入是利用 word2vec 训练得词向量. 证明本文 CNN+BiGRU 联合网络特征融合模型分类效果优于单模型.

(5)CA-BGA^[23]: 先添加一个 CNN, 再引入注意力机制, 添加一个 BiGRU, 再引入注意力机制, 最后将 CNN-Attetion 和 BiGRU-Attention 所得的结果拼接在一起. 网络输入是利用 word2vec 训练得字符级词向量. 证明利用 BiGRU 学习 CNN 池化后的特征, 加强了模型对特征的学习能力, 在文本分类上起到正影响的作用.

(6)CBGA-BGA: 先添加一个 CNN, 再添加一个 BiGRU, 再引入注意力机制. 添加一个 BiGRU, 再引入注意力机制. 最后将 CNN-BiGRU-Attetion 和 BiGRU-Attention 所得的结果拼接在一起. 网络输入是利用 word2vec 训练得词向量.

4.5 实验结果分析

通过计算测试集的 Accuracy 值、Precision 值、Recall 值、F1 值, 进行模型评估. 模型的对比结果如

表 5 所示.

表 5 模型对比结果

	Accuracy	Precision	Recall	F1
BiLSTM	0. 876 2	0. 872 7	0. 8751	0. 873 9
BiGRU	0. 881 1	0. 873 3	0. 886 1	0. 879 7
CNN	0. 887 0	0. 882 3	0. 888 0	0. 885 1
CNN+BiGRU	0. 907 2	0. 924 1	0. 883 4	0. 903 3
CA-BGA	0. 913 6	0. 898 7	0. 928 4	0. 913 3
CBGA-BGA	0. 923 9	0. 910 0	0. 937 6	0. 923 6

表 5 显示了 6 组模型在测试集上的对比结果. 从两个综合评测指标 Accuracy、F1 上来看. CBGA-BGA 的测试结果 Accuracy 取得 92. 39%, F1 值取得 92. 36%, 均高于其他 5 组模型, 说明本文模型能有效的提高文本分类效果.

第 4 模型与前 3 组模型的对比, 证明本文 CNN+BiGRU 联合网络特征融合模型分类效果优于单模型, 因为 CNN 对文本特征的强学习能力有助于联合模型对深层次特征的学习, 而 BiGRU 对序列化特征前后的学习能力, 对 CNN 提取的特征进行再加工, 增强了模型对特征的学习能力.

添加第 6 组模型与第 5 组模型的对比, 可知在联合模型的基础上引入注意力机制能有效的提高模型分类的准确度, 因为注意力机制分配给特征不同的权重, 让模型学习到特征之间的轻重不同之分, 有助于模型快速掌握重要的特征. 其中, 可看出 CBGA-BGA 模型优于 CA-BGA 模型, 证明利用 BiGRU 学习 CNN 池化后的特征, 在文本分类上起到正影响的作用.

可通过绘画模型的准确率和损失率的变化图, 从而更直观的反应 6 组模型的优劣. 由于验证集上的变化要比在训练集上更为真实的反映一个模型的优劣, 本文选择绘画验证集的准确率(val_acc)和损失率(val_loss)变化图. val_acc 的变化如图 7 所示, val_loss 的变化如图 8 所示. 从图 7val_acc 的变化图上来看, 总体上 6 组模型准确度都不断上升, CBGA-BGA 模型的准确度均达 90%以上, 在第 4 次迭代时达到最高值 92. 55%, 趋势平稳, 不似其余 5 组模型波动性大, 且多数居于其余模型上方. 可见 CBGA-BGA 模型在提取文本特征上更为优秀和稳定, 收敛性快, 在迭代次数较少的情况下也可以达到较高的准确度. 从图 8val_loss 的变化图上来看, CNN 的波动最大, 其余 5 组都较为集中. CBGA-BGA 下降到最小值 0. 223 6 后, 回升幅度小, 总体居于其余

模型下方,说明模型取得较好的收敛效果。

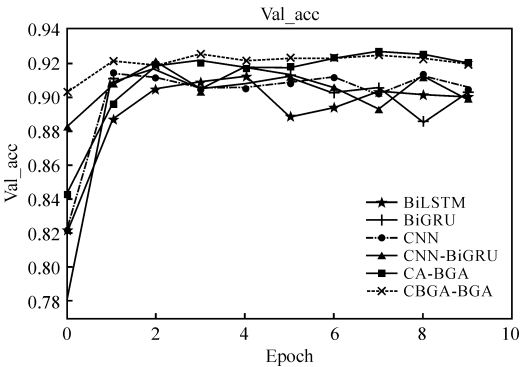


图 7 验证集准确度变化图

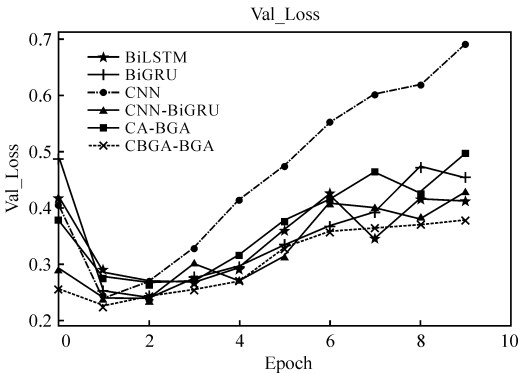


图 8 验证集损失率变化图

综合图 7、8 的分析结果,CBGA-BGA 模型具有收敛性快、准确度高、稳定的特点,相比较其他 6 组模型在文本分类上更具有优势。

为了量化模型的优劣,本文在测试集上进行预测,其中负样本 1 132 个,正样本 1 089 个。预测值统计结果如表 6 所示。Right 为模型预测正确的样本总数,Wrong 为模型预测错误的样本总数。输出结果为预测样本是 1 概率,为方便统计,将值大于 0.5 的定为 1,其余为 0。实际标签值是[0,1],如预测值与实际值相等,则为正确,相应项的值加 1。

表 6 预测值统计结果

	TP	FP	TN	FN	Right	Wrong
BiLSTM	953	139	993	136	1 946	275
BiGRU	965	140	992	124	1 957	264
CNN	967	129	1 003	122	1 970	251
CNN+BiGRU	962	79	1 053	127	2 015	206
CA-BGA	1 011	114	1 018	78	2 029	192
CBGA-BGA	1 021	101	1 031	68	2 052	169

表 6 中,CBGA-BGA 模型预测正确最多:2 052 个,错误最少:169 个,可见 CBGA-BGA 模型最优。

综上,针对本文所用数据集,利用 CNN-BiGRU 联合模型提取特征优于单模型 CNN 或 BiGRU。

5 结束语

本文提出了一种字符级联合网络特征融合的模型的中文文本情感分析方法。以字符级词向量作为输入,利用 CNN-BiGRU 和 BiGRU 分别同时进行特征提取,再引入注意力机制筛选特征,在丰富特征信息的同时减少噪声干扰。实验在中文网络购物评论文本上进行训练和测试,结果表明本文模型较并行单模型更优,但由于网络更加复杂化,一定程度上增加了模型时间代价。今后研究如何提高分类准确率且时间代价更小的模型,是下一步工作的目标。

参考文献:

[1] 付广成. 基于深度神经网络的电商评论情感分析研究[D]. 南昌:南昌大学,2018.

[2] 魏韩,向阳,陈千. 中文文本情感分析综述[J]. 计算机应用,2011,31(12):3321-3323.

[3] LIU B. Sentiment analysis and opinion mining[C]// Proceedings of the 2012 Synthesis Lectures on Human Language Technologies. Vermont, Australia: Morgan & Claypool Publishers. 2012: 152-153.

[4] 孙志军,薛磊,许阳明,等. 深度学习研究综述[J]. 计算机应用研究,2012,29(8):2806-2810.

[5] BENGIO Y, SCHWENK H, SENÉCAL J S, et al. A neural probabilistic language model[J]. Journal of Machine Learning Research, 2003, 3(6):1137-1155.

[6] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient Estimation of Word Representations in Vector Space [J]. Computer Science, 2013. arXiv: 1301.3781v3.

[7] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed Representations of Words and Phrases and their Compositionality[J]. Advances in Neural Information Processing Systems, 2013(26):3111- 3119.

[8] MNH A, HINTON G E. A scalable hierarchical distributed language model[C]//Proceedings of the 22nd Annual Conference on Neural Information Processing, Istanbul, 2008. New York: NIPS, 2009: 1081-1088.

[9] KALCHBRENNER N, GREFFENSTETTE E, BLUN-SOM P. A convolutional neural network for modelling sentences[J]. arXiv:1404.2188v1.

[10] KIM Y. Convolutional neural networks for sentence classification[J]. Eprint Arxiv, arxiv.org/abs/1408.5882, 2014.

[11] MERRIENBOER B V, GULCEHRE C. Learning

- phrase representations using RNN Encoder Decoder for statistical machine translation [C] //Computer Science 2014,[S. L]. 2014:1 -12.
- [12] EBRAHIMI J, DOU D. Chain based RNN for relation classification [C] //Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Denver, USA, 2015: 1244 -1249.
- [13] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997 (9): 1735-1780.
- [14] 曹宇慧. 基于深度学习的文本情感分析研究[D]. 哈尔滨: 哈尔滨工业大学, 2016.
- [15] CHO K, VAN MERRIENBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language. Doha: CoRR, 2014: 1724-1734.
- [16] 曹宇, 李天瑞, 贾真, 等. BGRU: 中文文本情感分析的新方法[J/OL]. 计算机科学与探索: 1-11 [2019-04-07]. <http://kns.cnki.net/kcms/detail/11.5602.TP.20180723.1614.008.html>.
- [17] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[J]. Computer Science, 2014, 10(4): <http://arxiv.org/abs/1409.0473>.
- [18] MNIH V, HEES N, GRAVES A. Recurrent models of visual attention[C]//Advances in Neural Information Processing Systems. [S. L]. 2014: 2204-2212.
- [19] 胡荣磊, 芮璐, 齐筱, 等. 基于循环神经网络和注意力模型的文本情感分析[J/OL]. 计算机应用研究, 2019 (11): 1-7. [2019-04-07]. <http://kns.cnki.net/kcms/detail/51.1196.TP.20180811.1330.064.html>.
- [20] 王伟, 孙玉霞, 齐庆杰, 等. 基于 BiGRU-Attention 神经网络的文本情感分类模型[J/OL]. 计算机应用研究: 1-10 [2019-04-07]. <http://kns.cnki.net/kcms/detail/51.1196.TP.20181011.1246.010.html>.
- [21] ZHOU C, SUN C, LIU Z, et al. A C-LSTM Neural Network for Text Classification [J]. Computer Science, 2015, 1(4): 39-44.
- [22] 李洋, 董红斌. 基于卷积神经网络和 BLSTM 网络特征融合的文本情感分析[J/OL]. 计算机应用: 1-7. [2019-04-07]. <http://kns.cnki.net/kcms/detail/51.1307.tp.20180719.1604.080.html>.
- [23] 陈洁, 邵志清, 张欢欢, 等. 基于并行混合神经网络模型的短文本情感分析[J/OL]. 计算机应用: 1-6. [2019-04-07]. <http://kns.cnki.net/kcms/detail/51.1307.TP.20190329.1643.008.html>.
- [24] 任勉, 甘刚. 基于双向 LSTM 模型的文本情感分类[J]. 计算机工程与设计, 2018, 39(7): 2064-2068.
- [25] 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.

作者简介:

王丽亚 女, (1994-), 硕士研究生. 研究方向为文本挖掘. E-mail: 1578938844@qq.com.

刘昌辉 男, (1965-), 博士, 教授, 硕士生导师. 研究方向为机器人控制技术、软件系统与智能计算、信息处理.

蔡敦波 男, (1981-), 博士, 教授, 硕士生导师. 研究方向为人工智能的智能规划、自动推理、约束优化、文本挖掘.

赵彬洲 女, (1970-), 硕士, 副教授, 硕士生导师. 研究方向为机器视觉、数据分析、文本挖掘.

王梦 女, (1994-), 硕士研究生. 研究方向为数据挖掘.

(上接第 79 页)

- [12] CAO Y, BO J, JIN T, et al. Image set representation with ℓ_1/ℓ_2 -Norm optimal mean robust principal component analysis[C]// International Conference on Image & Graphics. Shanghai, China, 2017.
- [13] LAI Z, XU Y, YANG J, et al. Rotational invariant dimensionality reduction algorithms[J]. IEEE Transactions on Cybernetics, 2016 (99): 1-14, DOI: 10.1109/TCYB.2016.2578642.
- [14] MA X, YE Q, YAN H. L2P-norm distance twin support vector machine [J]. IEEE Access, 2017 (5): 23473-23483.
- [15] TANVEER M. Robust and sparse linear programming

twin support vector machines[J]. Cognitive Computation, 2015, 7(1): 137-149.

- [16] HUA W, NIE F, HUANG H. Robust distance metric learning via simultaneous ℓ_1 -norm minimization and maximization[C]// International Conference on International Conference on Machine Learning. Banff, Canada, 2014.

作者简介:

沈洋 男, (1994-), 硕士研究生. 研究方向为人工智能、机器学习. E-mail: 1825472824@qq.com.

戴月明 男, (1964-), 硕士, 副教授. 研究方向为人工智能、软件工程.