

基于并行混合网络融入注意力机制的情感分析

孙 敏, 李 旻*, 庄正飞, 余大为

(安徽农业大学 信息与计算机学院, 合肥 230036)

(* 通信作者 ly-and@qq.com)

摘 要: 针对传统卷积神经网络(CNN)不仅会忽略词的上下文语义信息而且最大池化处理时会丢失大量特征信息的问题,传统循环神经网络(RNN)存在的信息记忆丢失和梯度弥散问题,和CNN和RNN都忽略了词对句子含义的重要程度的问题,提出一种并行混合网络融入注意力机制的模型。首先,将文本用Glove向量化;之后,通过嵌入层分别用CNN和双向门限循环神经网络提取不同特点的文本特征;然后,再把二者提取得到的特征进行融合,特征融合后接入注意力机制判断不同的词对句子含义的重要程度。在IMDB英文语料上进行多组对比实验,实验结果表明,所提模型在文本分类中的准确率达到91.46%而其F1-Measure达到91.36%。

关键词: 卷积神经网络;双向门限循环单元;特征融合;注意力机制;文本情感分析

中图分类号: TP391 **文献标志码:** A

Sentiment analysis based on parallel hybrid network and attention mechanism

SUN Min, LI Yang*, ZHUANG Zhengfei, YU Dawei

(School of Information and Computer, Anhui Agriculture University, Hefei Anhui 230036, China)

Abstract: Concerning the problems that the traditional Convolutional Neural Network (CNN) ignores the context and semantic information of words and loses a lot of feature information in maximal pooling processing, the traditional Recurrent Neural Network (RNN) has information memory loss and vanishing gradient, and both CNN and RNN ignore the importance of words to sentence meaning, a model based on parallel hybrid network and attention mechanism was proposed. First, the text was vectorized with Glove. After that, the CNN and the bidirectional threshold recurrent neural network were respectively used to extract text features with different characteristics through the embedding layer. Then, the features extracted by two networks were fused. And the attention mechanism was introduced to judge the importance of different words to the meaning of sentence. Multiple sets of comparative experiments were performed on the English corpus of IMDB. The experimental results show that the accuracy of the proposed model in text classification reaches 91.46% and F1-Measure reaches 91.36%.

Key words: Convolutional Neural Network (CNN); Bidirectional Gated Recurrent Unit (BGRU); feature fusion; attention mechanism; text sentiment analysis

0 引言

随着Web2.0时代的到来,越来越多的人喜欢在网络上分享自己的一些经验、知识和感受等各种信息。例如,美团、饿了么、携程等各种应用软件都包含各种用户的评论信息,充分利用这些信息不仅能指导消费者进行合理消费,而且还能够让商家认识自身的问题,对下一次决策有关键的作用^[1]。那么,如何从大量数据中高效、准确和快速获取用户的情感倾向就变得尤其重要。文本的情感分析是重要的研究方向其中之一,能够有效地分析出文本所包含的各种情感信息^[2]。

文本情感分析^[3]作为自然语言处理领域的一个重要研究方向。传统的文本情感分析的方法主要包含基于机器学习的方法^[4]和基于词典^[5]的情感分析两种。基于机器学习的方法需要大量人工标注数据,在处理复杂分类问题时泛化能力不强;基于词典的情感分析方法虽然不需要人工标注数据,但需要构建情感词典,而且分类的效果主要取决于情感词典的大

小和质量。当下,卷积神经网络(Convolutional Neural Network, CNN)^[6]和循环神经网络(Recurrent Neural Network, RNN)^[7]作为两大主流深度学习模型被应用在自然语言处理领域并且取得很好的效果,尤其是在短文本的情感分析任务中。Kim^[8]对比了词向量不同构造方法,把预训练处理好的词向量作为神经网络的输入,利用卷积神经网络实现句子级的文本分类,把神经网络应用于情感分类的任务中。Kalchbrenner等^[9]提出动态卷积神经网络模型用于提取句子的特征、处理不同长度的情感文本。由于传统的神经网络提取得到的只是局部特征,同时忽略了特征的前后联系,针对此问题本文提出并行混合网络融入注意力机制模型。

Mikolov等^[10]提出将循环神经网络应用到文本分类任务中,因为循环神经网络的输出取决于当前的输入和上一个节点输出的结果,可以更好学习到句子上下文的语义信息。但是循环神经网络在学习句子特征时容易出现梯度弥散问题,为了解决此问题,Hochreiter等^[11]提出长短期记忆(Long Short-

收稿日期:2019-11-28;修回日期:2020-01-09;录用日期:2020-01-19。 基金项目:国家自然科学基金资助项目(61402013)。

作者简介:孙敏(1994—),女,安徽蚌埠人,硕士研究生,主要研究方向:深度学习、情感分析; 李旻(1963—),男,安徽蚌埠人,教授,博士,主要研究方向:深度学习、计算机网络、智能交通、智能建筑与安全防范、农业网络信息工程; 庄正飞(1996—),男,安徽滁州人,硕士研究生,主要研究方向:机器学习、文本挖掘; 余大为(1994—),男,安徽安庆人,硕士研究生,主要研究方向:文本分类、注意力机制。

Term Memory, LSTM)网络模型, LSTM是RNN的众多变体之一,虽然LSTM在提取长距离上下文语义信息上是一种很有效的网络模型,但结构复杂,因此导致训练需要大量的时间和内存空间。Cho等^[12]提出的门限循环单元(Gated Recurrent Unit, GRU)是LSTM的一个变体,GRU比LSTM的架构简单,模型的训练效率得到了有效的提升。双向门限循环单元(Bidirectional Gated Recurrent Unit, BGRU)神经网络是由一个向前传播的GRU和一个向后传播GRU,前后两个方向GRU的输出状态连接组成的神经网络,本文使用BGRU来提取全局的特征挖掘文本中深层语义信息。

注意力机制的核心作用是在众多信息中注意到对当前任务更关键的信息。Mnih等^[13]进行图像处理时在RNN基础上引入注意力机制,使得图像处理效果较好。Bahdanau等^[14]把注意力机制应用在自然语言处理任务中,通过注意力机制可以把每个词学习到的表达和预测需要的词联系起来。从此之后,注意力机制被文本分析学者所广泛关注,例如,王伟等^[15]提出基于BiGRU-Attention模型的情感分类;田生伟等^[16]将BiLSTM和attention机制结合对维吾尔语事件实训关系进行识别;白静等^[17]使用BiLSTM-CNN-Attention进行两种特征融合的分类;陈洁等^[18]采用混合神经网络模型处理短文本情感分析都取得比较好的分类结果。

为了避免单一卷积神经网络忽略词的上下文语义信息和最大池化处理时会丢失大量特征信息,本文在卷积神经网络和BGRU提取得到的特征融合之后引入注意力机制提高文本情感分析的准确率和F1-Measure综合评价指标。

1 混合网络融入注意力机制模型

本文提出并行混合网络融入注意力机制模型的文本情感分析构造方法主要从以下两个方面考虑:1)把卷积神经网络和双向门限循环神经网络学习得到的特征进行融合;2)把融合之后的特征送入到注意力机制中以便判断不同的词对句子含义的重要程度。

1.1 CNN 并联 BGRU 提取特征并融合

本文利用卷积神经网络提取句子的局部特征,卷积神经网络的结构主要有输入层、卷积层、池化层和输出层。卷积神经网络结构如图1所示。

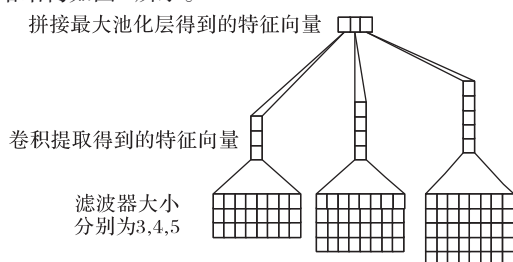


图1 卷积神经网络结构

Fig. 1 Convolutional neural network structure

输入层 文本中每个句子中每一个词都要利用Glove映射为对应的词向量 x_i ,句子中每个词向量 x_i 组成句子矩阵 S ,如式(1)所示:

$$S = \{x_1, x_2, \dots, x_n\} \quad (1)$$

其中, $x_i \in \mathbb{R}^k$, x_i 表示句子矩阵中第 i 个词的词向量, k 表示词向量的维度。 $S \in \mathbb{R}^{n \times k}$, n 表示句子矩阵中词向量的个数。本实验中词向量的维度 $k = 200$ 。

卷积层 卷积层采用 $r \times k$ 的滤波器对句子矩阵 S 执行卷积操作提取 S 的局部特征 c 。如式(2)所示:

$$c_i = f(w \cdot x_{i:i+r-1} + b) \quad (2)$$

其中: w 代表卷积核, b 代表偏置项; f 代表通过ReLU进行非线性

操作的函数; $x_{i:i+r-1}$ 代表句子矩阵中从 i 到 $i+r-1$ 共 r 行向量; c_i 代表通过卷积操作得到的局部特征。滤波器在整个句子矩阵中从上往下已设置的步长进行滑动,得到局部特征向量集合 C ,如式(3)所示:

$$C = \{c_1, c_2, \dots, c_{n-r+1}\} \quad (3)$$

池化层 池化方法分为两种,分别为最大池化和平均池化,本实验采用最大池化的方法对卷积操作得到的局部特征向量集合提取最大的特征代替整个局部特征,而且通过池化操作还可以降低特征向量的大小。如式(4)所示:

$$M_{ji} = \max(c_1, c_2, \dots, c_{n-r+1}) = \max\{C\} \quad (4)$$

其中, $1 \leq j \leq r, 1 \leq i \leq n-m+1, m$ 为卷积核个数。

输出层 将池化层中得到的所有 M_{ji} 拼接组成句子级特征向量。

文本具有很强的序列性,因此充分提取文本上下文语义信息就显得非常重要。为了体现文本的序列特征,使用LSTM建立序列模型。但是LSTM模型结构复杂,在训练时会花费更多的时间代价。GRU是LSTM的变体,主要的变化是使用更新门代替了LSTM中的输入门和遗忘门,模型结构的复杂度相对变得更加简单,训练参数减少并且时间代价比较低。其模型结构如图2所示,具体的计算过程如式(5)~(8)所示:

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (5)$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad (6)$$

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

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

$$y_t = \sigma(W_o \cdot h_t) \quad (9)$$

其中, r_t 表示重置门; z_t 表示更新门; h_t 表示隐含层状态的输出; y_t 表示GRU的输出; W_r, W_z, W_h, W_o 分别表示各个状态的权重矩阵, σ 和 $\tanh$ 表示激活函数。

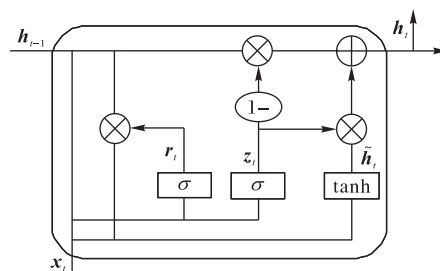


图2 GRU模型结构

Fig. 2 GRU model structure

GRU只是考虑当前词的上文信息并没有考虑到下文的信息,为了解决这个问题,本文使用双向门限循环神经网络不仅能够充分提取上文信息而且还能考虑到下文信息对当前词的影响,主要是由1个单向向前传播的GRU和1个单向向后传播的GRU组成,输出的结果由这2个GRU共同决定。其网络结构如图3所示,具体的计算过程如式(10)~(12)所示:

$$\vec{h}_t = \sigma(W_{sh} x_t + W_{hh} \vec{h}_{t-1} + b_h) \quad (10)$$

$$\overleftarrow{h}_t = \sigma(W_{sh} x_t + W_{hh} \overleftarrow{h}_{t+1} + b_h) \quad (11)$$

$$y = W_{hy} [\vec{h}_n; \overleftarrow{h}_1] + b_y \quad (12)$$

其中, $\vec{h}_t, \overleftarrow{h}_t$ 分别表示的是GRU向前传播的输出和向后传播的输出; y 表示的是BGRU的输出; W 表示的是权重矩阵; b 表示的是偏置向量; σ 为sigmoid激活函数。

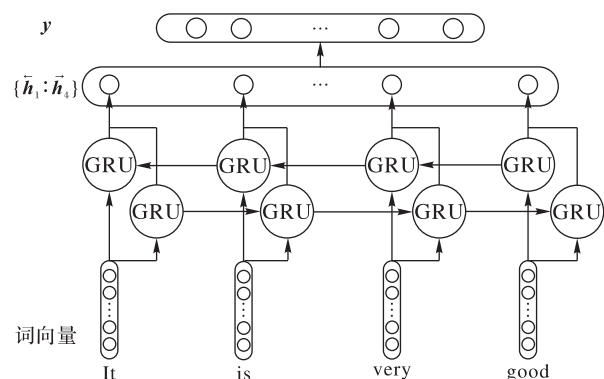


图3 BGRU情感分析网络模型

Fig. 3 BGRU sentiment analysis network model

1.2 注意力机制获取词的重要程度

注意力机制来源于对人类视觉的研究,人们会在所有的信息中选择性地关注自己在意的部分,同时还可以忽略其他的信息。把注意力机制应用在文本的情感分析中可以通过权重的分配计算不同词的词向量概率权重,使得部分词可以得到更多的关注从而突出重要的词语,提高分类的准确率。其模型结构如图4所示。

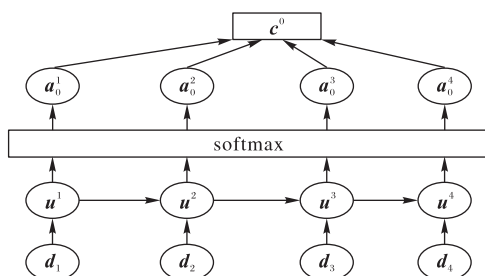


图4 注意力机制模型结构

Fig. 4 Attention mechanism model structure

假设 d_i 为卷积神经网络和双向门限循环神经网络提取得到的特征融合之后的特征向量,通过神经网络得到对 d_i 隐层的表示,如式(13)所示:

$$u_i^t = v_i \tanh(W_l d_i + b_w) \quad (13)$$

其中, W_l 权重矩阵; b_w 为偏置向量。

归一化处理,对得到的隐层表示用 softmax 函数进行归一化处理得到全局矩阵,如式(14)所示:

$$a_i^t = \text{softmax}(u_i^t) \quad (14)$$

最后用融合之后的特征向量 d_i 和全局矩阵 a_i^t 进行加权求和获取文本中每一个词的重要性信息的文本向量 c^t ,如式(15)所示:

$$c^t = \sum a_i^t h^i \quad (15)$$

1.3 并行混合网络融入注意力机制模型

1.3.1 并行混合网络融入注意力机制模型的构造

并行混合网络融入注意力机制模型结构如图5所示。本文的模型主要由卷积神经网络(CNN)、双向门限循环神经网络(BGRU)和注意力机制组成。其中,模型左边的卷积神经网络部分是由3个大小分别为3、4、5的卷积核组成,为了能够更加全面地提取文本中的局部特征;模型右边的是利用双向门限循环神经网络提取词的上下文的语义信息和文本中词的全局特征;模型的下面是注意力机制,主要是利用卷积神经网络和双向门限循环神经网络提取得到的特征计算不同的词向量概率权重,使得部分词得到更多的关注,提高分类的准确率。

本文并行混合网络融入注意力机制模型详细介绍如下:

1)将文本中的词利用 Glove 映射为对应的词向量 x_i ,然后组成句子矩阵 S 。

2)增加卷积神经网络,提取得到文本中的局部特征 M_j 。

3)增加双向门限循环神经网络,得到词的上下文语义信息和词的全局特征 y 。

4)将2)、3)提取的特征融合得到 d_i 作为注意力机制的输入,为了使部分词得到更多的关注提高分类的准确率,得到文本向量 c^t 。

5)增加 Dense 层,使用 sigmoid 激活函数对文本进行分类。

6)模型损失函数使用交叉熵损失函数 categorical_crossentropy,模型在训练的过程中不断地更新参数达到最优的效果。

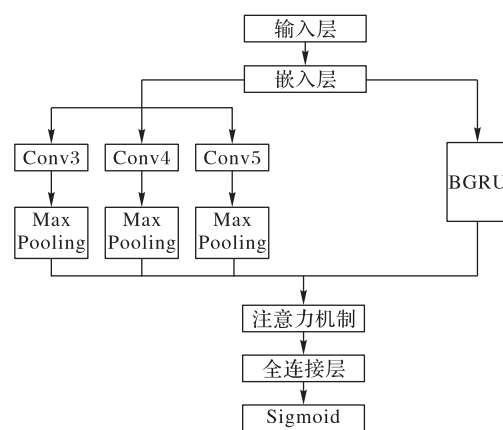


图5 并行混合网络融入注意力机制模型

Fig. 5 Model based on parallel hybrid network and attention mechanism

1.3.2 并行混合网络融入注意力机制模型的训练

本文在训练并行混合网络融入注意力机制模型中的参数包括卷积神经网络、双向门限循环神经网络和注意力机制中的全部参数。本模型使用的优化器是 Adam,因为 Adam 优化器结合了 RMSProp 和 AdaGrad 二者的优点,还可以计算参数的自适应学习率。为了防止在训练过程中过拟合,分别在三个卷积层之后加入 Dropout 函数,并把 Dropout 设置为 0.5,通过在每一次的迭代中随机丢弃部分训练参数来提高模型的泛化能力。本模型使用 Sigmoid 函数进行分类如式(16)所示:

$$p = (y = j | x, \theta) = \frac{1}{1 + \exp(-\theta^T x)} \quad (16)$$

其中: j 表示的是类别数,本文中主要分为积极和消极,取值是0或者1。 θ 为模型所有的参数。训练模型的参数 θ 使用的损失函数是交叉熵损失函数。本文样本点为 (x, y) , y 为真实的标签,取值是0或者1。现在假设 y_t 为样本点的真实标签, y_p 为该样本点属于积极类别的概率。如式(17)所示:

$$L(\theta) = -\frac{1}{N} \sum_{i=1}^N (y_i \ln y_p(x_i) + (1 - y_i) \ln (1 - y_p(x_i))) \quad (17)$$

2 实验

在公开带有情感标签的电影评论 IMDB 数据集上,对本文提出的情感模型进行验证和分析。实验环境具体配置如表1所示。

2.1 实验数据

实验的数据主要来自于电影评论 IMDB 数据集,电影情感评论的标签主要分为二种:0或者1。如果情感评论是积极的标记为1,情感评论是消极的标记为0。数据集的划分情况

如表 2 所示。

表 1 实验环境具体配置

Tab. 1 Detailed configuration of experimental environment

软件名称	软件类型	软件版本
Window 10	操作系统	Window 10 纯净版
Tensorflow	Python 深度学习库	1. 14
Keras	Python 深度学习库	2. 2. 4
Python	程序语言	3. 7
Numpy	Python 科学计算库	1. 14. 3
Jupyter notebook	工具	—
CPU	处理器	—

表 2 数据集

Tab. 2 Datasets

数据集	数据量
训练集	32 000
验证集	8 000
测试集	10 000
总数	50 000

2.2 实验参数设置

实验参数的设置直接影响模型的分类效果,但实验中又有很多的超参数需要设置与调整,在每一次迭代完成之后,会根据实验的准确率和损失率对所设置的超参数调整,经过多次实验以及多次迭代具体的参数设置如表 3 所示。

表 3 参数设置

Tab. 3 Parameter settings

参数名称	参数值
词向量维度	50,100,200,300
卷积核大小	3,4,5
卷积核数量	32
激活函数	ReLU
池化层大小	1
BGRU 隐藏层节点数	32,64,96,128,160,192
学习率	0. 001
Dropout	0. 5
Batch size	32
epochs	10

2.3 实验评价标准

评价是实验中一个重要的环节,可以直接反映出模型的好坏,本实验在电影评论情感分析的样本上采用准确率(Accuracy)、精确率(Precision)、召回率(Recall)和 F1-Measure 作为评价情感分析结果好坏的标准。准确率评估的是对模型正确分类的能力;精确率评估的是查准率;召回率评估的是查全率;F1-Measure 是综合评价指标。本实验针对该问题给出的分类结果混淆矩阵如表 4 所示,用此来介绍 4 种实验评价标准的计算方式。其中:TP(True Positive)指的是积极分类预测为积极分类数,FN(False Negative)指的是积极分类预测为消极分类数,FP(False Positive)指的是消极分类预测为积极分类数,TN(True Negative)指的是消极分类预测为消极分类数。

准确率(Accuracy)计算如式(18)所示:

$$Accuracy = (TP + TN)/(TP + FP + FN + TN) \quad (18)$$

精确率(Precision)计算如式(19)所示:

$$Precision = TP/(TP + FP) \quad (19)$$

召回率(Recall)计算如式(20)所示:

$$Recall = TP/(TP + FN) \quad (20)$$

F1-Measure 计算如式(21)所示:

$$F1-Measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (21)$$

表 4 分类结果混淆矩阵

Tab. 4 Confusion matrix of classification results

实际值	测试值	
	正/积极	负/消极
正/积极	TP	FN
负/消极	FP	TN

2.4 词向量维度的选择

词向量能够有效地反映词与词之间的关联度以及语义信息,理论上来说词向量维度越大越好,但是在实际情况的应用中需要考虑实验的整体性能和整体代价,从而选出最合适的词向量维度。分析表 5 可得,当词向量维度为 50 时,模型的整体性能最差,主要是因为词向量维度过小,使得词中包含的上下文相互影响的语义信息较少;当词向量维度为 200 时,模型的整体性能最好,训练的时间相对较小,主要因为词中上下文的关联度比较大,包含更丰富的文本语义信息。因此本模型把词向量的维度设置成 200。

表 5 不同词向量维度的实验结果

Tab. 5 Experimental results of different vectors dimensions of words

词向量的维度	Accuracy	Precision	Recall	F1-Measure	Time/s
50	0. 883 7	0. 884 6	0. 882 7	0. 883 5	402. 6
100	0. 898 2	0. 906 7	0. 887 7	0. 896 9	452. 8
200	0. 914 6	0. 923 2	0. 904 5	0. 913 6	616. 4
300	0. 911 0	0. 933 6	0. 884 2	0. 907 8	698. 7

2.5 BGRU 隐藏层节点数的选择

根据表 6 分析可以得出,当 BGRU 隐含层节点数为 32 时,模型的 accuracy 和 F1-Measure 最低,因为节点数过少不能够完全地提取语义信息;当节点数由 64 增加到 96 时,模型的 Accuracy 和 F1-Measure 达到最高的值;随后继续增加节点数,Accuracy 和 F1-Measure 都呈现下降的趋势,说明当节点数不断增加时,模型的整体性能越来越差。所以,本实验中 BGRU 隐含层节点数选择为 96 最好。

表 6 不同节点数的实验结果

Tab. 6 Experimental results of different numbers of nodes

节点数	Accuracy	Precision	Recall	F1-Measure	Time/s
32	0. 902 0	0. 928 3	0. 870 9	0. 898 1	429. 7
64	0. 908 0	0. 916 8	0. 897 3	0. 906 7	494. 2
96	0. 914 6	0. 923 2	0. 904 5	0. 913 6	616. 4
128	0. 913 9	0. 904 9	0. 925 7	0. 913 4	1 323. 9
160	0. 913 7	0. 918 1	0. 908 6	0. 913 2	2 163. 5
192	0. 909 5	0. 912 9	0. 905 5	0. 909 0	2 698. 0

2.6 对比实验设置

本实验模型与以下几种常见模型进行对比分析。为了使得对比的结果更具有可比性和准确性,所有模型词向量的输入都是 Glove200 维,每个对比模型中的参数值选取都一样。

1)CNN^[19]:只有 CNN,由 3 个卷积核大小分别为 3、4、5 的卷积神经网络组成。

2)BLSTM^[20]:只有 BLSTM 网络,其中 BLSTM 隐含层节点数为 96。

3)BGRU^[21]:只有 BGRU 网络,其中 BGRU 隐含层节点数为 96。

4)CNN-BGRU:把 3 个卷积核大小分别为 3、4、5 的 CNN

与隐含层节点数为 96 的 BGRU 网络链式连接。

5) CNN+BGRU:此模型主要是把 3 个卷积核大小分别为 3、4、5 的 CNN 与隐含层节点数为 96 的 BGRU 网络并行连接。

6) CNN+BGRU+Attention:本文提出的模型在把 3 个卷积核大小分别为 3、4、5 的 CNN 与隐含层节点数为 96 的 BGRU 网络并行连接的基础上引入注意力机制。

2.7 实验结果分析

实验在测试集上计算出 Accuracy 值、Precision 值、Recall 值、F1-Measure 值,具体各个模型实验结果如表 7 所示。

表 7 不同模型的实验结果

Tab. 7 Experimental results of different models

模型	Accuracy	Precision	Recall	F1-Measure
CNN	0.875 1	0.874 1	0.876 8	0.875 4
BLSTM	0.900 6	0.899 9	0.901 8	0.900 7
BGRU	0.904 6	0.894 5	0.918 6	0.906 2
CNN-BGRU	0.904 8	0.904 6	0.905 0	0.904 8
CNN+BGRU	0.907 6	0.908 3	0.906 7	0.907 5
CNN+BGRU+Attention	0.914 6	0.923 2	0.904 5	0.913 6

表 7 给出了 6 种实验结果,从 Accuracy 和 F1-Measure 这两个评价指标可以得出,本文提出的 CNN+BGRU+Attention 模型分别达到了 91.46% 和 91.36%,均优于其他的 5 种模型。从 1、2 组对比的实验结果可知,只利用卷积神经网络提取文本语义信息的效果远差于利用双向长短时记忆神经网络获取文本的上下文语义信息;从 2、3 组可以看出在 BLSTM 结构上做了改进的 BGRU 不仅在训练的时间上大幅缩减而且模型的 Accuracy 和 F1-Measure 都得到了部分的提升;从 4、5 组和 1、2、3 组的实验结果可以看出,两种模型的任意组合在情感分析的任务中都要比单个 CNN 模型或者单个的 BLSTM、BGRU 模型效果好,主要是因为可以利用 CNN 提取文本的局部语义信息,以及对文本特征的强学习能力,双向循环神经网络可以提取序列化特征上下文的学习能力,因此在模型组合时对 CNN 提取得到的特征进行再加工有着积极作用;第 4 组的链式融合主要来源于 Lai 等^[22]将 CNN 模型和 BiGRU 模型以链式的方式融合,从 4、5 组对比实验的结果可知本文提出并行融合的方式比链式融合方式效果更好;从 5、6 组实验得出,在融合之后的模型中引入注意力机制能有效地提高 Accuracy 和 F1-Measure,因为注意力机制通过权重的分配计算不同词的词向量概率权重,为模型学习到的特征给予不同的关注度,有利于模型快速地抓住最重要的特征。

本文将 6 种模型在 IMDB 数据集上,利用表 3 参数设置中的参数训练,得到在验证集上的准确率、损失率与迭代次数之间的变化曲线,分别如图 6 和图 7 所示。

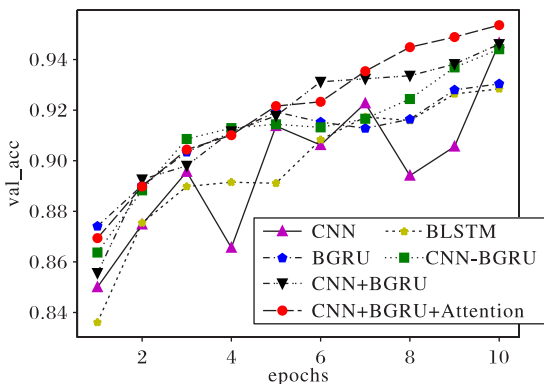


图 6 不同模型验证集准确率变化情况

Fig. 6 Variation of accuracy of different models on verification set

从图 6 中可以看出,随着迭代次数的增加,各个模型的准确率都在逐渐上升,并且在第二次迭代的时候所有模型的准确率都达到 87% 以上。本文模型的曲线相对于其他 5 条曲线变化比较平稳,波动较小,验证集准确率的值处于较高的位置,尤其在第 7 次迭代后较为明显达到 93% 以上,得出本文模型在提取文本特征情感分析任务中更稳定。

根据图 7 可以分析出,CNN 模型和 BLSTM 模型相对于其他 4 种模型的损失率变化波动最大极其不稳定;而本文所使用的模型 loss 值下降速度较快并且 loss 值很快到达了一个相对低的稳定值,总结出本文设计的模型取得了较好的收敛效果。

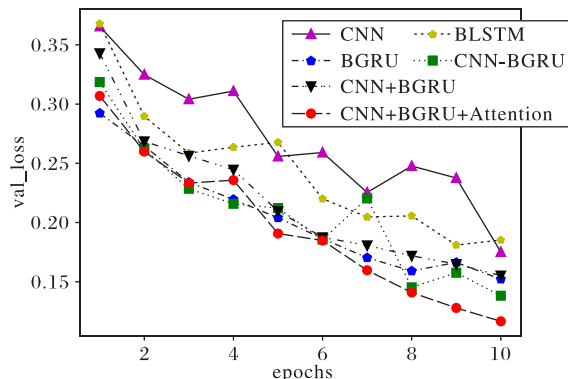


图 7 不同模型验证集损失率变化情况

Fig. 7 Variation of loss of different models on verification set

迭代时间指的是完成一次实验所需要的时间。图 8 给出了迭代时间与迭代次数之间的变化曲线。总体来看各个模型的迭代时间没有很大的波动,整体时间呈平稳趋势。CNN 模型的迭代时间最短,主要的原因是卷积神经网络的最大池化层的训练迭代收敛;BLSTM 的迭代时间最长,这与 BLSTM 结构比较复杂、参数太多分不开,BGRU 是 BLSTM 的一种改进,减少“门”的数量,并把细胞状态和隐藏状态结合在了一起,结构更加简单,训练参数减少,因此训练的时间也降低很多;本文提出模型的迭代时间高于 CNN-BGRU 和 CNN+BGRU,主要的原因是引入注意力机制在能够获取重点信息的同时也增加了加权计算时间。

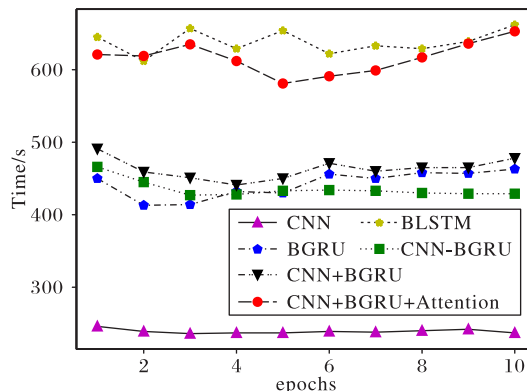


图 8 不同模型迭代时间变化情况

Fig. 8 Iteration time variation of different models

3 结语

本文提出一种基于并行混合网络融入注意力机制的文本情感分析方法,该方法将文本转换为词向量后分别通过 3 个大小不同的卷积核更全面、更细致地提取文本的特征和双向门限循环网络获取上下文的语义信息,在提取得到的特征进行融合之后引入注意力机制使部分词可以得到更多的关注突

出重要的词语,从而提高模型对文本类别的分类能力。从实验结果的分析中可以看出,在相同的实验条件下本模型的整体性能比单个CNN或者单个BGRU要好得多,虽然本模型在IMDB数据集上的准确率、综合评价指标较高,损失率较低,但是如果数据集的数量比较大,准确率、综合评价指标可能会有所下降,同时训练的时间大大增加。需要不断地改进模型以及算法,使得模型在数据量比较大时,得到较高的准确率、较低的损失率,同时减少训练的时间。

参考文献 (References)

- [1] 张蓊. 基于餐饮评论的情感倾向性分析[D]. 昆明:昆明理工大学, 2016: 1. (ZHANG L. Analysis of sentiment orientation based on restaurant reviews [D]. Kunming: Kunming University of Science and Technology, 2016: 1.)
- [2] 李然,林政,林海伦,等. 文本情绪分析综述[J]. 计算机研究与发展, 2018, 55(1): 30-52. (LI R, LIN Z, LIN H L, et al. Text emotion analysis: a survey [J]. Journal of Computer Research and Development, 2018, 55(1): 30-52.)
- [3] 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. (WEI W, XIANG Y, CHEN Q. Survey on Chinese text sentiment analysis [J]. Journal of Computer Applications, 2011, 31(12): 3321-3323.)
- [4] 程正双,王亮. 基于支持向量机的网络评论情感分析方法[J]. 电子技术与软件工程, 2019(16): 3-4. (CHENG Z S, WANG L. Sentiment analysis method of network reviews based on support vector machine [J]. Electronic Technology and Software Engineering, 2019(16): 3-4.)
- [5] 吴杰胜,陆奎,王诗兵. 基于多部情感词典与SVM的电影评论情感分析[J]. 阜阳师范学院学报(自然科学版), 2019, 36(2): 68-72. (WU J S, LU K, WANG S B. Sentiment analysis of movie reviews based on multiple sentiment dictionaries and SVM [J]. Journal of Fuyang Normal University (Natural Science), 2019, 36(2): 68-72.)
- [6] 刘政,卫志华,张韧弦. 基于卷积神经网络的谣言检测[J]. 计算机应用, 2017, 37(11): 3053-3056, 3100. (LIU Z, WEI Z H, ZHANG R X. Rumor detection based on convolutional neural network [J]. Journal of Computer Applications, 2017, 37(11): 3053-3056, 3100.)
- [7] 梁丕军. 基于循环神经网络和卷积神经网络的中文情感分类研究[D]. 长沙:湖南大学, 2017: 1. (LIANG P J. Research on Chinese sentiment classification based on recurrent neural network and convolutional neural network [D]. Changsha: Hunan University, 2017: 1.)
- [8] KIM Y. Convolutional neural networks for sentence classification [C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: Association for Computational Linguistics, 2014: 1746-1751.
- [9] KALCHBRENNER N, GREFENSTETTE E, BLUNSON P. A convolutional neural network for modelling sentences [C]// Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2014: 655-665.
- [10] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality [C]// Proceedings of the 26th International Conference on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc., 2013, 26: 3111-3119.
- [11] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1780.
- [12] CHO K, VAN MERRIËNBOER 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 Processing. Stroudsburg, PA: Association for Computational Linguistics, 2014: 1724-1734.
- [13] MNH V, HEES N, GRAVES A, et al. Recurrent models of visual attention [C]// Proceedings of the 27th International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2014: 2204-2212.
- [14] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate [EB/OL]. [2019-03-20]. <https://arxiv.org/pdf/1409.0473v7.pdf>.
- [15] 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]. 计算机应用研究, 2019, 36(12): 3558-3564. (WANG W, SUN Y X, QI Q J, et al. Text sentiment classification model based on BiGRU-attention neural network [J]. Application Research of Computers, 2019, 36(12): 3558-3564.)
- [16] 田生伟,胡伟,禹龙,等. 结合注意力机制的Bi-LSTM维吾尔语事件时序关系识别[J]. 东南大学学报(自然科学版), 2018, 48(3): 393-399. (TIAN S W, HU W, YU L, et al. Temporal relation identification of Uyghur event based on Bi-LSTM with attention mechanism [J]. Journal of Southeast University (Natural Science Edition), 2018, 48(3): 393-399.)
- [17] 白静,李霏,姬东鸿. 基于注意力的BiLSTM-CNN中文微博立场检测模型[J]. 计算机应用与软件, 2018, 35(3): 266-274. (BAI J, LI F, JI D H. Attention based BiLSTM-CNN Chinese microblog position detection model [J]. Computer Applications and Software, 2018, 35(3): 266-274.)
- [18] 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]. 计算机应用, 2019, 39(8): 2192-2197. (CHEN J, SHAO Z Q, ZHANG H H, et al. Short text sentiment analysis based on parallel hybrid neural network model [J]. Journal of Computer Applications, 2019, 39(8): 2192-2197.)
- [19] 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 Y H, ZHANG C Y, ZHAO B L, et al. Sentiment analysis of Twitter data based on CNN [J]. Journal of Data Acquisition and Processing, 2018, 33(5): 921-927.)
- [20] XIAO Z, LIANG P. Chinese sentiment analysis using bidirectional LSTM with word embedding [C]// Proceedings of the 2016 International Conference on Cloud Computing and Security, LNCS 10040. Cham: Springer, 2016: 601-610.
- [21] 曹宇,李天瑞,贾真,等. BGRU: 中文文本情感分析的新方法[J]. 计算机科学与探索, 2019, 13(6): 973-981. (CAO Y, LI T R, JIA Z, et al. BGRU: new method of Chinese text sentiment analysis [J]. Journal of Frontiers of Computer Science and Technology, 2019, 13(6): 973-981.)
- [22] LAI S, XU L, LIU K, et al. Recurrent convolutional neural networks for text classification [C]// Proceedings of the 29th AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI, 2015: 2267-2273.

This work is partially supported by the National Natural Science Foundation of China (61402013).

SUN Min, born in 1994, M. S. candidate. Her research interests include deep learning, sentiment analysis.

LI Yang, born in 1963, Ph. D., professor. His research interests include deep learning, computer network, smart transportation, intelligent building and security, agricultural network information engineering.

ZHUANG Zhengfei, born in 1996, M. S. candidate. His research interests include machine learning, text mining.

YU Dawei, born in 1994, M. S. candidate. His research interests include text classification, attention mechanism.