

引入双循环机制深度学习模型的文本情感分析

胡任远,刘建华,王璇,罗逸轩,林鸿辉

(福建工程学院 计算机科学与数学学院,福建 福州 350118)

摘要: 深度神经网络模型通常使用注意力机制或融合卷积神经网络来提取特征,但由于注意力机制抓取的特征过于单一,存在提取特征不完善的问题。将循环机制引入卷积神经网络中,构建了具有双循环结构的网络模型(DRCNN),从而改善模型的特征提取能力,将其与双向长短期记忆网络结合,提出一种带有注意力机制、特征提取能力更强的混合模型(BiLSTM-DRCNN)并应用于情感分类任务中。通过情感分类的实验分析表明,BiLSTM-DRCNN神经网络模型具有比较好的性能,与融合卷积神经网络(CNN)和双循环长短期记忆神经网络(BiLSTM)模型相比,综合评价指标提高2%以上;与BiLSTM-CNN、Fusion Model模型相比,综合评价指标提高了近1%,且收敛速度更快。

关键词: 双向长短期记忆神经网络;双循环卷积神经网络;注意力机制;文本情感分析

中图分类号: TP391.1

文献标志码: A

文章编号: 1672-4348(2022)04-0383-08

Text sentiment analysis that introduces a double recurrent mechanism deep learning model

HU Renyuan, LIU Jianhua, WANG Xuan, LUO Yixuan, LIN Honghui

(School of Information Science and Engineering, Fujian University of Technology, Fuzhou 350118, China)

Abstract: Deep neural network models usually use attention mechanism or fusion convolutional neural network for feature extraction. However, due to the fact that the features grasped by the attention mechanism are too simplified, there is a problem that the extraction features are not perfect. Therefore, the recurrent mechanism was introduced into the convolutional neural network, and a double recurrent convolutional neural network model was constructed to improve the feature extraction ability of the model. Then the model was combined with the bidirectional long and short-term memory network, so as to put forward a hybrid model with the attention mechanism and stronger feature extraction capability (BiLSTM-CNN), which was applied to sentiment classification tasks. Experimental analysis of sentiment classification shows that the BiLSTM-DRCNN neural network model has good performance, and the comprehensive evaluation index is increased by more than 2% compared with the fusion convolutional neural network (CNN) and the double-loop long short-term memory neural network (BiLSTM) model. Compared with the BiLSTM-CNN and Fusion Model models, the comprehensive evaluation index is improved by nearly 1%, and the convergence is faster.

Keywords: bidirectional long and short-term memory neural network; double recurrent convolution neural network; attention mechanism; text sentiment analysis

早期的情感分析大多使用传统的机器学习方法来实现,例如支持向量机、随机森林、决策树、K

邻近算法等机器学习方法^[1]。但是,以传统机器学习方法为核心的情感分析方法需要依赖复杂人

收稿日期: 2022-01-17

第一作者简介: 胡任远(1997—),男,江苏丹阳人,硕士研究生,研究方向:自然语言处理。

通信作者: 刘建华(1967—),男,江西吉安人,教授,博士,研究方向:智能计算,大数据分析,自然语言处理。

工规则的特征选择策略,模型参数调优和模型性能表现会受到较大的影响。

近年来,深度学习模型被广泛应用于自然语言处理领域,国内外许多学者将其应用到情感分析中^[2]。Kim 等^[3]首次将卷积神经网络应用于英文文本的语义特征提取任务中,并提出了一种可以捕捉局部语义相关性的模型 TextCNN^[4],但未能获取整体文本序列的上下文联系。梁军等^[5]使用长短期记忆网络(long short-term memory, LSTM)提取上下文信息,通过情感极性转移模型进行信息提取和分类,虽然提高了分类准确率,但未提取文本序列的特征。Wang 等^[6]为解决该问题提出了融合卷积神经网络(convolutional neural network, CNN)的 CNN-LSTM 分类模型,使用 LSTM 网络作为解码器处理 CNN 的输出序列,使模型在一定程度上学习到输入序列的情感极性和上下文信息,但特征选择能力仍不足。李洋等^[7]在 Word2Vec 训练模型的词向量基础上,使用混合的网络模型 CNN-BiLSTM 进行特征提取和语义信息获取,使模型具有更好的分类性能,但对特征的处理仅使用不同大小卷积核的 CNN 网络。Lai 等^[8]提出一种由双向循环神经网络和最大池化层组成的 RCNN 模型(regions with CNN features),使用最大池化层做特征提取工作,提高了模型的分类能力,但增加了时间成本。刘全等^[9]使用深度分层网络模型,使用区域卷积神经网络处理用户评论,将其分割为定长的不同区域来进行多特征的提取。

综上所述,现有文本情感分类模型的特征提取方式更多局限于使用注意力机制或在模型中添加 CNN 层。其中,注意力机制存在抓取的关系过于单一、无法提取出文本序列中字词之间联系的问题。在模型中添加 CNN 虽然可以提取出各语块的主要信息,但仅使用单次卷积池化的模型也无法有效识别短文本内容、含有转折关系等句子的情感极性。

为减少模型抓取关系过于单一对分类结果产生的影响,本研究将循环机制引入进卷积神经网络中,提出一种双循环卷积神经网络模型(double recurrent convolutional neural network, DRCNN),通过改变深度学习网络的权重共享模式,各循环结构之间使用权重私有化的方法来使模型学习到更丰富的情感特征信息。

1 相关工作

1.1 卷积神经网络

卷积神经网络(CNN)采用局部连接、权值共享等策略有效地解决了网络参数大、训练时间长等问题。CNN 主要包含了输入层、卷积层、池化层、全连接层和输出层,将词嵌入处理的文本向量作为卷积层的输入。卷积层使用卷积核对输入数据进行卷积计算来提取文本特征。卷积层输出如式(1)所示。

$$A = f(W \cdot X + b) \quad (1)$$

式中, X 为输入矩阵, W 为权重矩阵, b 为偏置量, f 为非线性激活函数。通过最大池化层来进一步提取特征,同时降低了向量维度,减少了模型参数,计算结果也可以被进一步处理。

1.2 双向长短期记忆网络

1.2.1 传统长短期记忆网络

Hochreiter 等^[10]于 1997 年提出长短期记忆网络模型 LSTM,改善了循环神经网络(recurrent neural network, RNN)中存在的长期依赖问题,可作为复杂的非线性单元用于构造更大型深度神经网络。LSTM 采用了细胞状态和门机制,结构如图 1 所示。

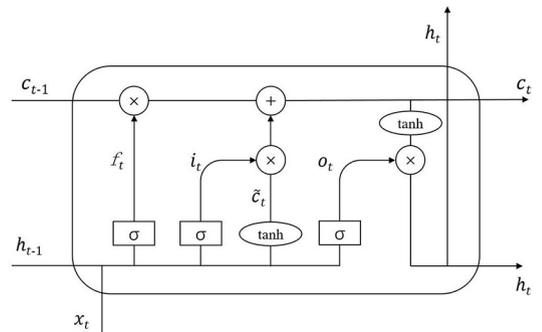


图 1 长短期记忆网络

Fig.1 Long and short-term memory network

图 1 中, h 为遗留信息的输出; i_t 为输入门,决定了有多少 t 时刻网络的输入将保存到单元状态; o_t 为输出门,控制单元状态有多少输出到 LSTM 的当前输出值; f_t 代表着遗忘门,决定了 $t-1$ 时刻的单元状态有多少能保留到 t 时刻; σ 为非线性的 sigmoid 函数。 i_t, o_t, f_t 的计算分别如式(2)~(4)所示:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (2)$$

$$o_i = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (3)$$

$$f_i = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (4)$$

式中, x_t 表示 t 时刻的当前输入, h_{t-1} 为 $t-1$ 时刻遗留的信息向量; W_i 和 U_i 为输入门的记忆细胞所对应的权重矩阵; U_o 、 W_o 为输出门的记忆细胞所对应的权重矩阵; W_f 和 U_f 为遗忘门的记忆细胞所对应的权重矩阵; b_i 、 b_o 、 b_f 分别为输入门、输出门和遗忘门的偏置量。

图 1 中, \tilde{c} 为记忆细胞候选项, c 是记忆细胞, 用来记录细胞状态的更新, \tilde{c} 和 c 的计算分别如式(5)、式(6)所示:

$$\tilde{c}_i = \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (5)$$

$$c_i = f_i c_{i-1} + i_i \tilde{c}_i \quad (6)$$

式中, \tanh 代表着非线性激活函数 $\tanh()$, 网络接收到 t 时刻的输入 x_t 和 $t-1$ 时刻遗留的信息向量 h_{t-1} 作为 3 个门机制的输入, 分别为 i_t 输入门、 o_t 输出门和 f_t 遗忘门在 t 时刻所得到的激活向量。最后得到最终输出 h_t 如式(7)所示:

$$h_t = o_t \tanh(c_t) \quad (7)$$

1.2.2 双向长短期记忆网络

双向长短期记忆网络 BiLSTM 是由前向 LSTM 与后向 LSTM 组合而成, 可以同时考虑文本的上下文语境、提供更加全面的语义信息, 减少因上下文语序问题出现的判断错误, 显著提升模型效果。因为神经网络易存在过拟合的问题, Hinton 等^[11] 在 2012 年提出了 Dropout 策略, 通过随机丢弃部分比例的神经元、减小模型参数从而达到一定程度上的正则化效果。BiLSTM 模型的结构如图 2 所示。

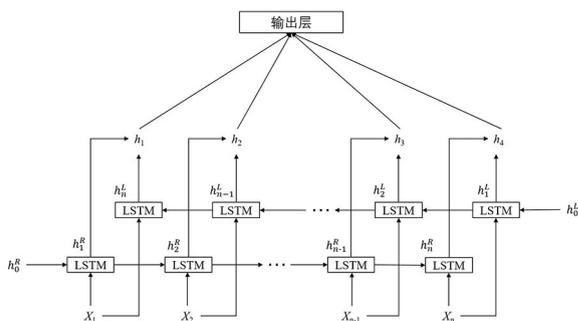


图 2 双向长短期记忆网络

Fig.2 Bidirectional long and short-term memory network

将 x_1, x_2, \dots, x_n 作为模型输入, 通过两个方向的 LSTM 网络便可以获得 $h^L =$

$[h_1^L, h_2^L, \dots, h_n^L]$ 和 $h^R = [h_1^R, h_2^R, \dots, h_n^R]$ 。连接 h^L 和 h^R 获得输出数据 $h_i (i = 1, 2, \dots, n)$ 。

2 深度分层网络模型

深度分层网络模型 (BiLSTM-DRCNN 模型) 是结合了 BiLSTM 和 DRCNN 的情感分析模型, 具体结构如图 3 所示。BiLSTM-DRCNN 模型先利用经典 BiLSTM 网络处理序列使得序列中每个元素都可以包含双向的时序信息, 并引入注意力机制使文本序列中带有情感倾向的信息被高度关注。BiLSTM-DRCNN 模型利用多次不同层面的循环机制来处理序列特征, 可以高度关注文本序列中的情感倾向, 提高情感特征表达的准确性。BiLSTM-DRCNN 模型不仅能够同时接收上下文信息, 还能有效提取语句的情感极性, 有助于提升模型对情感分析的性能。

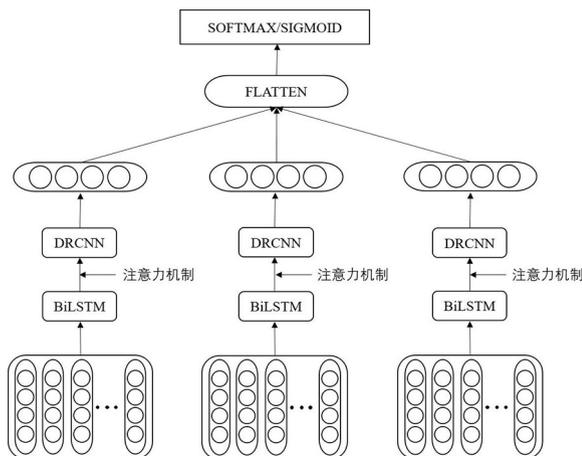


图 3 BiLSTM-DRCNN 网络模型

Fig.3 Network model of BiLSTM-DRCNN

2.1 模型输入

BiLSTM-DRCNN 模型所接收的数据是对原文本序列经过处理后的向量表示。对于中文文本数据先用正则表达式处理和停用词去除; 对于英文文本数据也需要进行停用词去除以及分词和统一大小写处理。最后使用最广泛的方法将词语用分布式词向量来表示, 作为模型的输入数据。

2.2 DRCNN 网络

本节提出的 DRCNN 为改进传统 CNN 中卷积运算并添加池化层的方式, 利用多次卷积操作反复提取所需要的局部特征之后, 最终构造了一个循环卷积块 (recurrent convolutional block,

RCB)。RCB 层通过多次的卷积运算来处理序列特征,再将输出传给池化层进行特征提取,使模型可以更好地分析出文本的情感倾向。DRCNN 模型更关注句中含有情感极性信息的内容,它拥有多个不同循环结构组成的网络,其结构主要包含数个循环卷积层、数个池化层、展平层以及最终输出层,如图 4 所示。

从图 4 可见,DRCNN 模型在输入层之后连接着多个 RCB 层 ($RCB_1 \sim RCB_m$) 对序列进行处理,在层与层之间加入池化层,降低数据维度,并同时提取特征;最后使用 FLATTEN 将多维数据展平成一维数据输入给分类器。每一层 RCB 都是由若干个卷积层所组建而成,即需要经过数次特征提取的过程,图 5 为 RCB 层的具体展开图。

RCB 可看作深度为 m 的前馈子网,由多个卷积层组成,其中 f 为卷积核激活函数, A_0 为经过 BiLSTM 和注意力机制处理过后的序列信息。 $f(A_i)$ 只能存在一个输入和一个输出,其中 $i = 1, 2, 3, \dots, m - 1$ 。每个循环卷积层之后都会有一个最大池化层来进行数据降维,并且每个 RCB 层的权重不共享,从而使模型拥有更高的容错率。其中 $f(A_i)$ 的计算如式(8)所示:

$$f(A_i) = \text{relu}(W_1 A_i + W_2(a_i \otimes r_L)) \quad (8)$$

式中, A_i 为输入序列, W_1, W_2 分别为权重矩阵, $a_i \otimes r_L$ 为将 a_i 重复拼接 L 次, a_i 为序列对应的情感标签, L 为固定句子长度。

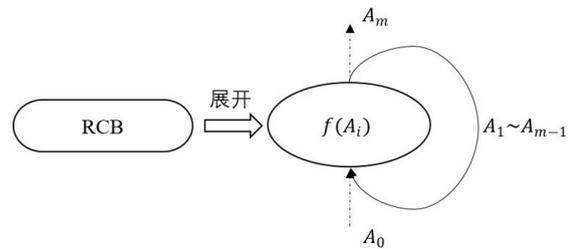


图 5 RCB 展开图

Fig.5 RCB expansion diagram

2.3 模型训练

模型对不同的文本分类数据使用不同的激活函数,接收经过 DRCNN 特征提取和 FLATTEN 数据展平后的输出,随后进行分类处理。其中二分类数据使用了 sigmoid 函数如式(9)所示,三分类数据采用了 softmax 函数计算待分类句子情感极性,如(10)所示:

$$\hat{y} = \text{sigmoid}(W \cdot x + b) \quad (9)$$

$$\hat{y} = \text{softmax}(W \cdot x + b) \quad (10)$$

式中, W 为权重矩阵, b 为偏置量, x 为 DRCNN 输出层数据, \hat{y} 为网络模型的预测输出。分类问题的损失函数为交叉熵函数,如式(11)所示,其中 y 为实际结果:

$$\text{loss} = -\frac{1}{n} \sum_x [y \ln \hat{y} + (1 - y) \ln(1 - \hat{y})] \quad (11)$$

2.4 注意力机制

注意力机制是一种能够聚焦于局部信息的机制。计算机视觉领域最早提出了注意力机制分配权重的方法,帮助神经网络具备专注于其输入特征的能力。Wang 等^[12]将注意力机制进入 LSTM 网络中,用于处理情感分析任务,最终结果证明了注意力机制在自然语言处理中的有效性。李洋等^[6]将 CNN 与 BiLSTM 结合并且引入注意力机制用于处理情感分析中,利用 CNN 提取局部特征后作为 BiLSTM 层的输入,结合注意力机制对特定情感进行高度关注,更有效地关注输入序列的句子情感信息,比传统网络的分类效果更佳。本研究在 BiLSTM 层后加入注意力机制,可以使模型更好地关注情感极性特征,随后使用 DRCNN 接收处理好的序列进一步进行特征提取,在注意力机制的处理序列的前提下将文本的情感特征更好地提取出来,从而提高模型分类结果。

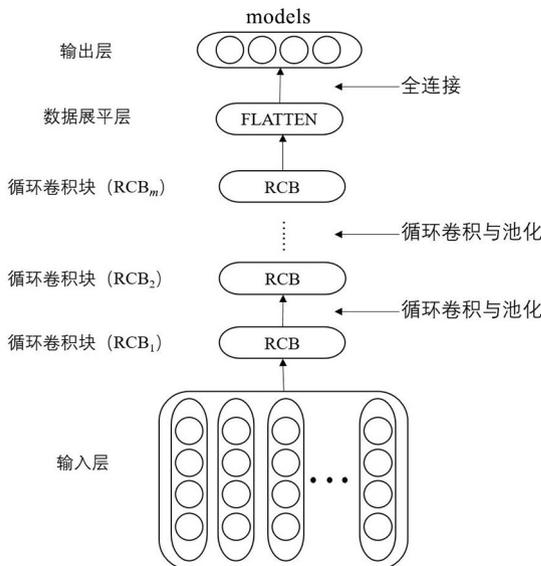


图 4 DRCNN 的结构分布图

Fig.4 Distribution map of DRCNN structure

注意力机制的原理如式(12)~(14)所示:

$$c_{i'} = \sum_{i=1}^T a_{i'} h_i \quad (12)$$

$$a_{i'} = \text{softmax}(\sigma(s_{i'-1}, h_i)) \quad (13)$$

$$\sigma(s_{i'-1}, h_i) = v^T \tanh(W_s s_{i'-1} + W_h h_i) \quad (14)$$

式中 $c_{i'}$ 代表着输出变量, h_i 为隐藏层, $a_{i'}$ 表示一个权重的概率分布, σ 为变换函数, $s_{i'-1}$ 为查询项。

3 实验结果与分析

3.1 实验环境

操作系统为 Windows 10, CPU 为 Intel Core i5-8300H, GPU 为 GeForce GTX 1060, 内存大小为 16 GB, 开发环境为 TensorFlow 2.2.0-GPU, 开发工具使用了 JetBrains Pycharm。

3.2 实验数据

采用的情感分类数据集如表 1 所示, 其中 CN 代表中文文本数据, EN 代表英文文本数据。训练集和测试集是独立的两个数据集, 通过不同语言和分类目的来确保本文提出方法的有效性。中文实验数据采用了 Data Fountain 的开源数据 o2o 商铺食品相关评论数据(缩写为 o2o)以及疫情期间网民情绪数据(缩写为 Cov19), 英文实验数据则是 IMDB 和 Twitter 评论数据。

表 1 试验数据统计

Tab.1 Test data statistics

语言	数据集	积极	消极	中性
		句子数	句子数	句子数
CN	o2o-Train	6793	2414	0
	o2o-Test	1698	604	0
	Cov19-Train	20313	13521	46095
	Cov19-Test	5079	3381	11524
EN	IMDB-Train	12500	12500	0
	IMDB-Test	2454	2456	0
	Twitter-Train	1883	2465	7335
	Twitter-test	471	617	1834

3.3 评价标准

因实验数据中各情感数据分布不均的问题, 使用的评价模型的指标为代表正确预测正样本占实际预测为正样本的比例精准率(precision)、代表正确预测正样本的比例召回率(recall)和 F1 值

(F1-score)。F1 作为一种综合指标, 可以作为模型分类结果的评价标准之一, 如式(15)~(17)所示:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (15)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (16)$$

$$\text{F1} = \frac{2\text{PR}}{\text{P} + \text{R}} \quad (17)$$

其中, TP (true positive)是指将原本为正类文本判定为正类文本的个数, FP (false positive)是指将原本为负类的文本判定为正类文本的个数, FN (false negative)是指将原本为正类文本判定为负类文本的个数。

3.4 模型参数设置

由于模型参数的选择对结果会有较大的影响, 实验采用控制变量法, 模型采用 Adam 优化器, 将卷积层填充模式设为 same。BiLSTM 隐层节点分别取 32、64 和 128, DRCNN 隐层节点数分别取 64、128 和 256 进行对比试验。通过多次实验发现取表 2 参数时, BiLSTM-DRCNN 深度分层具有最优性能。

表 2 模型参数设置

Tab.2 Model parameter settings

参数	值
BiLSTM 隐层节点数	128
DRCNN 隐层节点数	256
Optimizer	Adam
学习率	5e-4
Batch_size	32
Dropout	0.3

本研究在设计 DRCNN 循环机制时采用了较少的循环次数, 主要由于多次卷积特征提取在信息量上的削减会导致重要特征的遗漏, 从而对结果造成负面影响, 并且会增加计算时间开销。在多次试验后得到了 RCB 层和 DRCNN 层的最佳参数, 可以得出 DRCNN 层循环次数为 2 且 RCB 层深度为 2 时, 模型效果最优。

3.5 对比实验

(1) CNN 模型: 基于文献[13]作为经典网络模型具有强大的特征提取能力, 但是存在忽略句

子长距离依赖关系及时序的问题。

(2) BiLSTM 模型: 基于文献[14]提出的经典网络模型, 可以处理时间序列但输入的文本序列, 但冗余信息过多缺乏特征提取能力, 无法有效判别句子的情感极性。

(3) ATT-CNN 模型: 基于文献[15]将注意力机制引入传统网络模型 CNN 中, 可以关注特定的情感词汇, 有效的判别出句子所表达的情感极性。但该模型无法捕捉句子上下文信息。

(4) ATT-BiLSTM 模型: 基于文献[16]提出的带有注意力机制的双向长短期记忆网络。该模型克服了 CNN 无法获取句子长距离依赖关系的问题, 在情感分析任务中效果显著。在高度关注特定情感词的同时, 有效识别用户不同的情感极性。但该模型未进一步提取序列特征, 模型效果还有提升空间。

(5) Fusion Model 模型: 基于文献[6]提出的结合 CNN 和 BiLSTM 的模型, 输入序列通过该模型用

不同大小卷积核的 CNN 层提取特征后拼接 BiLSTM 的输出, 用 CNN 代替注意力机制运作, 可以获取整句评论的长距离依赖关系。但该模型没有融入循环机制在 CNN 中, 无法进行特征多次提取。

(6) BiLSTM-CNN 模型: 基于文献[17]提出的网络模型, 型结合了 BiLSTM 和 CNN 的优点, 在全局语义的基础上进行特征提取, 模型结果超过单一模型。

(7) BiLSTM-DRCNN 模型: 本研究提出的深度网络模型, 使用 BiLSTM 可以使每个字都包含着上下文的信息, 同时结合适当深度的 RCNN 结构将 BiLSTM 处理过后的文本序列进行特征提取, 充分提取情感极性的同时不会忽略边缘词汇, 大大提升了分类结果的准确率。

3.6 实验结果与分析

本研究使用 7 个应用在情感分析领域的模型在两种语言 4 种的数据集上进行 7 组对比实验, 模型实验结果如表 3 所示。

表 3 不同模型在 4 种数据集上的结果

Tab.3 Results of different models on 4 datasets

模型	o2o			Cov19			IMDB			Twitter		
	P/%	R/%	F/%	P/%	R/%	F/%	P/%	R/%	F/%	P/%	R/%	F/%
CNN	86.67	79.32	82.21	72.99	67.87	70.40	86.55	86.06	86.26	75.9	75.58	76.75
BiLSTM	92.83	82.55	87.18	74.93	65.45	69.86	87.02	86.00	86.60	77.92	77.21	78.44
ATT-CNN	87.72	83.89	84.96	73.39	67.62	70.48	84.52	88.36	86.47	78.01	77.63	77.81
ATT-BiLSTM	94.23	82.21	87.26	73.10	68.97	71.07	86.63	89.54	87.88	79.61	77.70	78.94
Fusion Model	89.83	83.92	86.77	75.09	68.75	71.82	86.38	89.52	87.92	79.36	77.35	79.15
BiLSTM-CNN	91.17	86.58	88.04	74.19	69.43	71.83	85.08	91.15	87.99	79.11	78.05	79.39
BiLSTM-DRCNN	91.87	87.25	89.04	74.66	69.97	72.31	86.80	89.68	88.20	79.28	77.77	80.17

注: P、R、F 分别代表精准率、召回率和 F1 值。

从表 3 可见, 未引入注意力机制的基础网络模型分类效果并不理想, CNN 和 BiLSTM 的平均 F1 分数只有 78.07% 和 79.54%; 而 ATT-CNN 和 ATT-BiLSTM 模型的平均 F1 分数比基础的 CNN 和 BiLSTM 模型总体提高了 0.8% 和 0.68%。与 BiLSTM-CNN 网络模型相比, BiLSTM-DRCNN 模型在 4 个数据集上的平均 F1 分数提高了 0.62%, 在 Twitter 数据集上模型效果提高了 1.02%; 与 CNN-BiLSTM 模型相比, BiLSTM-DRCNN 模型在 4 个数据集上的平均 F1 分数提升

了 0.97%, 在 o2o 数据集上模型效果提升了 2.27%。因 IMDB 数据集为公开数据集, BiLSTM-DRCNN 模型相较于 SOTA 上的 CNN-LSTM 和 CNNu 模型, 在准确率指标上平均高出 0.3%, 比 S-LSTM 模型高出 2.05%, 收敛速度更快。

BiLSTM-CNN 模型在 4 个数据集上的模型的 F1 分数比 CNN-BiLSTM 模型的 F1 分数高 0.39%。因为先使用 BiLSTM 可以使序列很好地学习到上下文信息后再通过 CNN 进行特征提取; 而 CNN-BiLSTM 模型存在 CNN 处理后的语块信息已经缺

少部分序列的完整性,再使用 BiLSTM 便无法很好地抓取上下文信息。在 4 个数据集上,BiLSTM-DRCNN 模型的收敛速度都比其他模型更快。因为不同 RCB 层之间的权重不共享,故研究不同 Dropout 比例下模型拟合能力的强弱尤为重要,可以看出模型在 Dropout 比例为 0.3 时效果最优,过高过低的 Dropout 比例都会影响模型的拟合能力。

本研究设置了 3 组消融实验,通过改变模型的组合方式或代替模型中部分网络结构等方法来验证 BiLSTM-DRCNN 模型的有效性。首先,BiLSTM-DRCNN 模型中 DRCNN 用 LSTM 代替后组成 BiLSTM-LSTM 模型来检测 DRCNN 的提取能力。其次,将 BiLSTM-DRCNN 模型中的注意力机制剔除组成 BiLSTM-DRCNN 模型,验证引入注意

力机制是否会对模型的实验结果产生影响。最后,将 BiLSTM 部分用 LSTM 代替后组成 LSTM-DRCNN,如此可以验证使用 BiLSTM 接收双向信息是否比 LSTM 接收单向信息对模型结果产生影响。实验结果如表 4 所示。

从表 4 可见,模型在 4 个不同的数据集上均取得了最好的 F1 分数。与 BiLSTM-LSTM 模型相比,BiLSTM-DRCNN 模型在 4 个数据集上的平均 F1 分数提高了 1.41%,在 Twitter 数据集上提高了 2.42%。与 BiLSTM-DRCNN 模型相比,BiLSTM-DRCNN 模型总体提升了 1.26%,体现使用注意力机制对模型结果的有较大影响。从使用 LSTM 代替本文模型中 BiLSTM 部分的 LSTM-DRCNN 模型实验结果可以看出, BiLSTM-DRCNN 总体提升了 0.75%,其中在 IMDB 数据集上总体提升了 1.31%。

表 4 不同模型在 4 种数据集上的消融实验结果

Tab.4 Results of ablation experiments of different models on 4 datasets

模型	o2o 上的实验结果			Cov19 上的实验结果			IMDB 上的实验结果			Twitter 上的实验结果		
	P/%	R/%	F/%	P/%	R/%	F/%	P/%	R/%	F/%	P/%	R/%	F/%
BiLSTM-LSTM	88.93	85.48	87.24	92.11	82.9	87.25	90.54	86.45	88.55	91.87	87.25	89.04
BiLSTM-DRCNN	75.37	67.89	71.38	74.77	68.78	71.55	74.54	69.29	71.67	74.66	69.97	72.31
LSTM-DRCNN	87.41	88.52	87.70	89.70	85.84	87.73	81.91	92.76	86.89	86.80	89.68	88.20
BiLSTM-DRCNN	78.59	76.92	77.75	76.96	79.03	78.14	78.26	81.02	79.59	78.77	81.62	80.17

3.7 模型训练时间分析

对算法训练时间进行统计分析,如表 5 所示,在模型对比试验中选择的模型进行筛选,共取在 4 种数据集上 10 个 epoch 的总计训练时间做对比,剔除了不包含 BiLSTM 网络结构的模型,因为 BiLSTM 的输入为时序输入,在训练耗时上远远超过 CNN 网络,对比的意义不显著。

表 5 各模型在 4 个数据集上训练 10 个 epoch 的总共训练时间

Tab.5 Total time for each model to train 10 epochs on 4 datasets

模型	t/s
BiLSTM	564
ATT-BiLSTM	588
Fusion model	668
BiLSTM-CNN	640
BiLSTM-DRCNN	706

从表 5 可以看出训练最快的为 BiLSTM 模型,加入了注意力机制后训练时间延长了 24 s,提出的 BiLSTM-DRCNN 虽然耗时最长,但其收敛速度却优于其他模型,如图 6 所示,可以看出 BiLSTM-DRCNN 对比另外 2 个模型复杂度相当的深度学习模型,均可以更快的收敛到最优解附近。

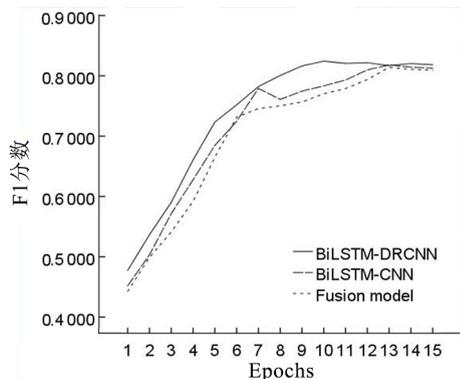


图 6 收敛速度分析图

Fig.6 Convergence rate analysis

4 结语

本研究提出一种融合双向长短期记忆网络与对向循环卷积神经网络的 BiLSTM-DRCNN 模型,将其应用于情感分析任务中。通过和多种神经网络模型在情感分类任务实验分板对比,验证了 BiLSTM-DRCNN 模型具有更好的性能。所提出的 BiLSTM-RCNN 模型可以将文本数据经过 BiLSTM 与注意力机制处理,借助 BiLSTM 的双向机制使得每个词都包含了上下文信息,消除传统 LSTM 无法捕获下文信息的缺点,并且将处理过

后的序列输入到 DRCNN 中,利用 DRCNN 的内循环机制来充分提取局部情感特征,进一步获取句子的情感极性,有效处理了边缘词汇和局部情感词汇。BiLSTM-DRCNN 模型改变深度学习网络的权重共享模式,各循环结构之间使用权重私有化的方法来使模型学习到更丰富的情感特征信息。权重私有化使模型具有更多参数,相比使用 CNN 和注意力机制的模型,该结构可以更好的提升模型情感分类能力。同时单个循环结构中采用了多卷积的模式,特征提取相比 CNN 更强。

参考文献:

- [1] BOIY E, MOENS M F. A machine learning approach to sentiment analysis in multilingual Web texts[J]. *Information Retrieval*, 2009, 12(5): 526-558.
- [2] ZHANG L, WANG S, LIU B. Deep learning for sentiment analysis: a survey[J]. *WIREs Data Mining and Knowledge Discovery*, 2018, 8(4): e1253.
- [3] KIM S M, HOVY E. Identifying and analyzing judgment opinions[C]//*Proceedings of the Main Conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*. Morristown, NJ, USA: Association for Computational Linguistics, 2006: 200-207.
- [4] KIM Y. Convolutional neural networks for sentence classification[C]//*Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2014: 1408-1417.
- [5] 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.
- [6] WANG J, YU L C, LAI K R, et al. Dimensional sentiment analysis using a regional CNN-LSTM model[C]//*Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2016: 225-230.
- [7] 李洋, 董红斌. 基于 CNN 和 BiLSTM 网络特征融合的文本情感分析[J]. *计算机应用*, 2018, 38(11): 3075-3080.
- [8] LAI S W, XU L H, LIU K, et al. Recurrent convolutional neural networks for text classification [J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2015, 29(1): 2267-2273.
- [9] 刘全, 梁斌, 徐进, 等. 一种用于基于方面情感分析的深度分层网络模型[J]. *计算机学报*, 2018, 41(12): 2637-2652.
- [10] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. *Neural Computation*, 1997, 9(8): 1735-1780.
- [11] HINTON G E, SRIVASTAVA N, KRIZHEVSKY A, et al. Improving neural networks by preventing co-adaptation of feature detectors[J]. *Computer Science*, 2012, 3(4): 212-223.
- [12] WANG Y Q, HUANG M L, ZHU X Y, et al. Attention-based LSTM for aspect-level sentiment classification[C]//*Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2016: 606-615.
- [13] ZHAO Z W, WU Y Z. Attention-based convolutional neural networks for sentence classification[C]//*Interspeech 2016. ISCA: ISCA*, 2016: 705-709.
- [14] ZHOU P, SHI W, TIAN J, et al. Attention-based bidirectional long short-term memory networks for relation classification [C]//*Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2016: 207-212.
- [15] YIN W P, SCHÜTZE H, XIANG B, et al. ABCNN: attention-based convolutional neural network for modeling sentence pairs[EB/OL]. 2015: arXiv: 1512.05193[cs.CL]. <https://arxiv.org/abs/1512.05193>.
- [16] BIN Y, YANG Y, SHEN F M, et al. Describing video with attention-based bidirectional LSTM[J]. *IEEE Transactions on Cybernetics*, 2019, 49(7): 2631-2641.
- [17] 赵宏, 王乐, 王伟杰. 基于 BiLSTM-CNN 串行混合模型的文本情感分析[J]. *计算机应用*, 2020, 40(1): 16-22.