

基于权重词向量与改进TextCNN的中文新闻分类

万铮,王芳,黄树成

(江苏科技大学计算机学院,江苏镇江212114)

摘要:为了解决单一深度学习模型提取信息不充分、分类效果不佳的问题,提出一种混合多神经网络的BA-InfoCNN-BiLSTM模型。该模型使用BERT作为词嵌入层,获得词的向量表示,再使用注意力机制让词获得不同权重;然后一方面将其送入改进的文本卷积神经网络(InfoCNN)中获取文本的局部信息特征,另一方面送入双向的长短时记忆网络(Bi-LSTM)中获得文本的全局信息特征;最后将提取到的局部信息和全局信息进行拼接融合,送入softmax函数中进行分类,得到分类结果。通过与其他模型进行对比实验,该模型获得了较好的分类效果,在新浪新闻和搜狐新闻数据集上分别取得了95.07%和84.95%的准确率,在一定程度上解决了单一模型捕获信息不充分的问题。

关键词:文本分类;词嵌入;注意力机制;卷积神经网络;循环神经网络

DOI:10.11907/rjdk.222051

开放科学(资源服务)标识码(OSID):

中图分类号:TP391.1

文献标识码:A

文章编号:1672-7800(2023)009-0059-06



Chinese News Classification Based on Weighted Word Vector and Improved TextCNN

WAN Zheng, WANG Fang, HUANG Shucheng

(School of Computer, Jiangsu University of Science and Technology, Zhenjiang 212114, China)

Abstract: In order to solve the problem of insufficient information extraction and poor classification effect of a single deep learning model, this paper proposes a BA-InfoCNN-BiLSTM model with a hybrid multi-neural network. The model uses BERT as the word embedding layer to obtain the vector representation of the word, and then uses the attention mechanism to obtain different weights for the word, and then sends it to the improved text convolutional neural network on the one hand to obtain The local information features of the text, on the other hand, are sent to the bidirectional long short-term memory network to obtain the global information features of the text, and finally the extracted local information and global information. After the features are spliced and fused, they are sent to the softmax function for classification, and the classification results are obtained. After comparing experiments with other models, this model has achieved good classification results. It has achieved 95.07% and 84.95% accuracy on the Sina news data set and Sohu news data set respectively, which solves the problem of insufficient information captured by single model to a certain extent.

Key Words: text classification; word embedding; attention mechanism; convolutional neural network; recurrent neural network

0 引言

文本分类是自然语言处理领域的一项基础且重要的任务,在新闻推荐、搜索引擎、垃圾邮件检测等方面都有着重要应用。尽管文本分类已经有着多年的发展历史,但仍存在着一些不足之处。目前文本分类仍是自然语言处理领域研究的一个热点问题。文本分类是指通过一定的算

法,给输入的文本分配一个或多个预先设定好的标签^[1]。若只为每个文本分配一个标签,则称为单标签文本分类;若为每个文本分配一个及以上的标签,则称为多标签文本分类。

随着信息技术的快速发展,人们进入了一个信息爆炸的时代,互联网已经成为人们日常生活中获取信息的主要途径之一。其中绝大部分信息都是以文本形式存在的,面对着铺天盖地的文本信息,光靠人力维护是不可能的。那

收稿日期:2022-09-07

基金项目:国家自然科学基金项目(61772244)

作者简介:万铮(1996-),男,江苏科技大学计算机学院硕士研究生,研究方向为自然语言处理;王芳(1971-),女,硕士,江苏科技大学计算机学院副教授、硕士生导师,研究方向为数据挖掘、网络安全;黄树成(1969-),男,博士,江苏科技大学计算机学院教授、硕士生导师,研究方向为人工智能、机器学习。本文通讯作者:万铮。

么如何通过机器自动将这些文本信息进行分类以方便人们更好地获取,成为当下的一个研究难题。人们最先把机器学习方法用于文本分类,如改进的TF-IDF、支持向量机^[2]、朴素贝叶斯^[3]等。传统的机器学习分类方法将整个文本分类问题拆分成特征工程和分类器两部分。特征工程分为文本预处理、特征提取、文本表示3部分,最终目的是把文本转换成计算机可理解的数字,并封装足够用于分类的信息,再进行分类^[4]。虽然这些方法在一定程度上解决了文本分类问题,但仍存在着一些弊端。这些方法过于依赖人工设计的特征,并且对于文本的表示还存在数据稀疏和特征向量纬度过高的问题,对于网络新闻中出现的大量新词不能很好地表示其语义特征。

随着深度学习技术的不断发展,涌现出越来越多性能良好的深度学习模型。这些模型不仅能很好地解决传统机器学习方法存在的数据稀疏和特征向量维度过高的问题,而且准确度也明显提升。但是单一的深度学习模型也有其局限性,如TextCNN只关注到了局部信息,而往往会忽略掉全局语义,造成分类效果不佳。基于此,本文提出一种混合多神经网络的BA-InfoCNN-BiLSTM模型。

1 相关工作

若想要让计算机处理语句或文档,首先需要将这些语句或文档转换成数字,将字或词转换为向量的过程称为词嵌入。最开始采用One-Hot编码,用于判断文本中是否具有该词语。后来发展成根据词语在文本中的分布情况对词进行表示。近年来,随着深度学习的发展,直接推动了词嵌入技术的变革,使得分布式的词语表达得到了大量使用。分布式表示可以克服独热表示的缺点,解决了词汇表示与位置无关的问题。分布式表示通过计算向量之间的距离(欧氏距离、余弦距离)体现词与词之间的相似性。Bengio等^[5]最早使用神经网络来构建语言模型。2013年,Mikolov等^[6]提出一种浅层神经网络模型Word2Vec,其包括Continuous Bag-of-Words^[7]和Skip-Gram^[8]两种模型训练方法,通过分布式假设(如果两个词的上下文是相似的,其语义也是相似的)直接学习词的词向量,同时为了减少输出层的计算量,使用层次softmax和负采样对其进行优化。但该方式只考虑了文本的局部信息,未能有效利用整体信息。针对此问题,Pennington等^[9]提出全局词向量(Global Vectors, Glove)模型,同时考虑了文本的局部信息与整体信息。但无论是Word2Vec还是Glove,本质上都是一种静态的词嵌入方式,无法解决一词多义的问题。2018年,谷歌提出的BERT模型解决了一词多义的问题^[10]。BERT模型通过联合调节所有层中的左右上下文来预训练未标记的文本深度双向表示,此外还通过组装长句作为输入,增强了对长距离语义的理解。

在捕获文本特征方面,Hochreiter等^[11]提出的长短时

记忆神经网络解决了梯度爆炸和梯度消失问题;Kalchbrenner等^[12]提出动态卷积神经网络模型处理长度不同的文本,将卷积神经网络应用于NLP;Kim^[13]提出文本分类模型TextCNN,该模型结构更简单,利用多个大小不同的卷积核提取文本中的特征,然后对这些不同粒度的特征进行池化操作,从而得到更准确的局部特征;陈珂等^[14]利用多通道卷积神经网络模型,从多方面的特征表示学习输入句子的文本信息;Long等^[15]将双向长短时记忆网络与多头注意力机制相结合对社交媒体文本进行分类,克服了传统机器学习中的不足。本文在前人研究的基础上,通过融入前文信息对传统的TextCNN作出了改进。

2 模型设计

本文提出的BA-InfoCNN-BiLSTM模型通过在词嵌入层后加入注意力机制进行残差连接来提升重要词的比重,再通过改进的卷积神经网络与双向长短时记忆网络分别提取局部和全局特征,最后将特征进行融合后用于分类。模型整体结构如图1所示,由输入层、BERT嵌入层、权重词向量层、改进的卷积层、双向长短时记忆网络层和输出层组成。

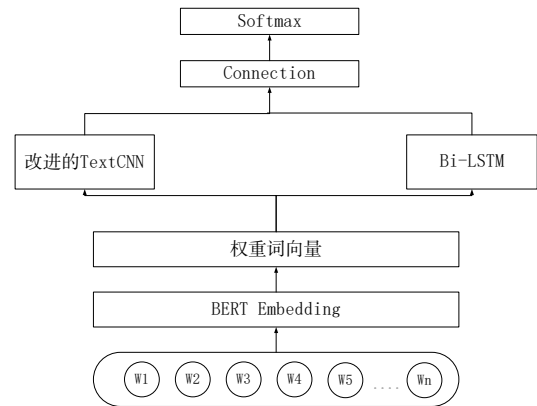


Fig. 1 BA-InfoCNN-BiLSTM model structure

图1 BA-InfoCNN-BiLSTM模型结构

2.1 词嵌入层

本模型嵌入层的目的是将文本转化为词向量,首先需要解决的问题就是分词。对于英文文本,单词与单词之间本就以空格隔开,所以英文文本不需要进行额外的分词操作。与英文文本不同的是,中文文本是由字构成,字本身就可以表达出一定的含义,而字与字之间又可以组成词,表达出新的语义。如今两种比较流行的分词方式是:一是像英文文本分词那样,以字为粒度,直接将文本中的字映射为一个向量,这种做法虽然方便,但是往往会割裂文本中字与字所组成的词的意思;二是利用像Jieba这样的分词工具先对文本进行分词,再将得到的词转化为词向量,但这种方式存在着更严重的弊端,因为分词的好坏会在很大程度上影响最后的分类结果。

如今的分词工具对于陌生词的分词效果较差,在专业

名词上更是严重依赖于用户构建的词典,并且对长词的分词效果较差。而新闻标题中往往会产生大量新词,严重影响分词的正确性,从而干扰最后的分类效果。因此,模型使用基于字粒度的词嵌入方式来弥补该弊端。BERT 的中文版本正是以字为单位进行嵌入的,十分适合作为嵌入层。BERT 的两大功能分别是预训练和微调。预训练有两大任务:一是掩码语言模型,即随机遮掩一部分词,然后让模型预测这些词;二是下一句预测,即判断两个句子之间是否有上下文关系来增强模型对句子的理解能力。微调则是在进行下游任务时,模型不断调整其参数的过程,但由于 BERT 的结构是由 12 个 Transformer 编码器构成,计算量较大,十分消耗时间,所以本实验过程中并没有选择进行微调,而只是使用在大规模语料上预训练过的 BERT 模型参数完成字到词向量的转换。设有文本 $T = \{t_1, t_2, t_3, \dots, t_n\}$, 文本长度为 n , 将其送入 BERT 模型, 得到该文本的词向量矩阵 $E = \{e_1, e_2, e_3, \dots, e_n\}$ 。矩阵大小是 $n \times d$, 其中 d 是每个字的维度。然后将 BERT 生成的矩阵 E 作为注意力层的输入。

2.2 权重词向量层

本模型在 BERT 之后引入注意力机制。注意力机制最早是由 Bahdanau 等提出的,用于模拟人脑的注意力模型,最早用于图像处理方面。Vaswani 等^[16]提出的 Transformer 便是基于自注意力机制获得单词间的长距离依赖关系。本模型之所以在嵌入层后引入注意力机制,是由于在嵌入层中只使用了 BERT 在其他语料上预训练得到的词向量。但是为了避免大量运算,在实验过程中并没有进行微调,没有发挥 BERT 中自注意力机制的作用,而在新的语义环境中,每个词在新闻标题中的重要程度也会有所不同。所以在得到词向量之后,需要通过注意力机制对字词权重重新进行分配,以体现不同词对文本全局语义特征的重要程度。注意力分数计算如式(1)所示。其中, W_a 是可训练参数, b_a 是偏置项, \tanh 是激活函数, v_a 是可学习的上下文向量。 a_i 是经过 softmax 函数后得到的 e_i 的权重(见式(2)), 然后将每个词向量加权后进行残差连接得到 s_i (见式(3)), 最后将得到加权的词嵌入矩阵 $S = \{s_1, s_2, s_3, \dots, s_n\}$ 分别送入改进的卷积层和 Bi-LSTM 层。

$$u_i = v_a \tanh(W_a e_i + b_a) \quad (1)$$

$$a_i = \frac{\exp(u_i)}{\sum_{i=1}^n \exp(u_i)} \quad (2)$$

$$s_i = e_i + a_i * e_i \quad (3)$$

2.3 改进的卷积层

在注意力机制之后引入改进的 TextCNN 来提升模型对特征的捕捉能力。TextCNN 能够通过使用不同大小的卷积核实现对 N-Gram 特征的提取,从而获取到不同层级的语义特征。但其短板是 TextCNN 通过卷积只能获得文本的局部依赖关系,而忽视了远距离语义的影响。所以针对该问题,本模型对 TextCNN 的卷积层进行了一些改进。

从整体上而言,对语义的理解是以从左到右的顺序进行的,所以进行卷积操作的词之前的文本信息是十分重要的。为解决 TextCNN 只关注局部信息的问题,在进行卷积操作的过程中,通过不断融入前文信息来提升模型性能。InfoCNN 过程如图 2 所示。

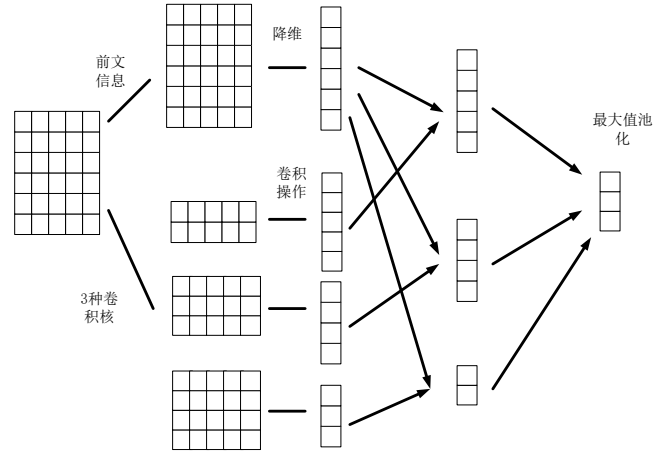


Fig. 2 InfoCNN process

图 2 InfoCNN 过程

首先根据词向量矩阵 $S = \{s_1, s_2, s_3, \dots, s_n\}$ 生成其前文语义矩阵 $R = \{r_0, r_1, r_2, \dots, r_n\}$, 如式(4)所示:

$$r_i = \sum_{j=1}^i s_j \quad (4)$$

其中, r_0 为零向量, 然后用全连接层进行降维, 得到前文信息向量 $G = \{g_0, g_1, g_2, \dots, g_n\}$ 。接着再用窗口大小为 2、3、4 的卷积核 W 进行卷积操作, 每次卷积操作得到特征 c_i , 提取局部特征的公式如式(5)所示:

$$c_i = f(W_h * S_{i:i+h-1} + b_h) \quad (5)$$

其中, h 为卷积核 W_h 滑动窗口的大小, 卷积核 W_h 的大小是 $h \times d$, d 是词向量维度。 $S_{i:i+h-1}$ 为从 S 中第 i 行到 $i+h-1$ 行的局部文本矩阵, b_h 为偏置项, f 代表非线性激活函数。最后, 结合提取的局部特征和前文信息特征, 最终得到的卷积结果 u_i 如式(6)所示:

$$u_i = 0.8c_i + 0.2g_{i-1} \quad (6)$$

最后, 在得到的结果 U 中, 采用最大值池化策略获取每个通道的最大值, 将这些值送入最后的输出层。

2.4 BiLSTM 层

由于神经网络特别适合处理序列数据, 已被成功应用于自然语言处理等众多时序问题中。为了能有效解决传统神经网络的梯度消失或爆炸问题, 本模型使用 Bi-LSTM 对 BERT 和注意力机制得到的权重词向量矩阵进行特征提取。长短时记忆网络结构如图 3 所示。

该网络有 3 个门: 一是遗忘门, 用来控制上一时刻 C_{t-1} 保存到当前时刻 C_t 的特征信息, 如式(7)所示; 二是输入门, 其控制了此时网络的输入 X_t 保存到当前时刻 C_t 的特征信息, 计算方式如式(8)、式(9)所示; 三是输出门, 用来控制当前时刻 C_t 的输出值 Y_t , 计算方式如式(10)一式(12)

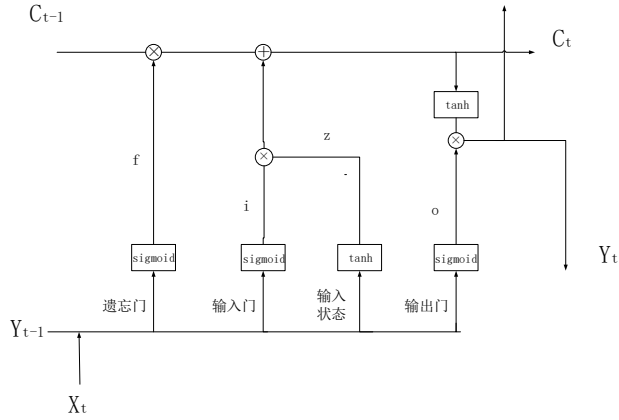


Fig. 3 LSTM structure

图3 LSTM结构

所示:

$$f_t = \sigma(W_f[Y_{t-1}, X_t] + b_f) \quad (7)$$

$$i_t = \sigma(W_i[Y_{t-1}, X_t] + b_i) \quad (8)$$

$$z_t = \tanh(W_z[Y_{t-1}, X_t] + b_z) \quad (9)$$

$$o_t = \sigma(W_o[Y_{t-1}, X_t] + b_o) \quad (10)$$

$$C_t = f_t * C_{t-1} + i_t * z_t \quad (11)$$

$$Y_t = o_t * \tanh(C_t) \quad (12)$$

但由于长短时记忆网络当前时刻的输出信息是由前一时刻的输出信息和当前时刻的输入信息共同决定的,即当前时刻的输出信息只考虑了该时刻与该时刻之前的信息,而没有考虑该时刻之后的信息,没有充分利用上下文信息。为了解决这一问题,Graves等^[17]提出双向长短时记忆网络。双向长短时记忆网络通过正向和逆向的LSTM获得第 t 时刻正向隐藏层状态向量 Y_{1t} 与逆向隐藏层状态向量 Y_{nt} ,并将 Y_{1t} 和 Y_{nt} 拼接起来作为最终的隐藏层状态向量 Y_t ,该向量包含了上下文信息。计算方式如式(13)所示:

$$Y_t = [Y_{1t} \oplus Y_{nt}] \quad (13)$$

2.5 输出层

模型最后的输出层是把改进的卷积层得到的结果与Bi-LSTM层得到的结果进行拼接融合,然后引入全连接层进行降维,之后用Dropout方法让降维后的特征向量以一定的概率失活,从而避免出现过拟合现象。最后送入softmax函数^[18]进行分类,得到最终的预测结果。

3 实验与分析

3.1 实验数据集

为了验证本模型在新闻主题文本分类任务上的有效性,本文使用两个以新闻为主题的数据集进行实验,如表1所示。

(1)新浪新闻数据集。新浪新闻数据集中的数据来自于新浪新闻2018—2022年间产生的新闻标题。通过收集这些新闻标题,然后经反复筛选压缩及过滤后,整理生成新浪新闻数据集。该数据集共包含20万条短文本,分为

Table 1 Data set information

表1 数据集信息

数据集	类别	平均长度	总数	训练集	验证集	测试集
新浪新闻	10	19	200 000	180 000	10 000	10 000
搜狗新闻	12	17	34 218	22 699	5 755	5 764

电子竞技、地产、体育、股市、科学、财经、时事、教育、政治、明星10个类别,每类包含2万条数据。

(2)搜狗新闻数据集。通过网络开源搜狗新闻数据集进行数据清洗,去除部分缺少标签的数据,并去除新闻内容,只保留新闻主题。数据集包含旅游、电子竞技、地产、军事、体育、股市、科学、财经、时事、教育、政治、明星共12个类别。

3.2 实验设置

(1)实验环境。本实验在PyCharm上进行代码编写,编程语言选择Python 3.7版本,深度学习框架选择Pytorch 1.1版本,CPU型号为AMD EPYC 7302 16-Core Processor,内存为252GB,GPU型号为GeForce RTX 3080,显存为10GB。

(2)实验参数。由于模型使用BERT的中文版本进行词嵌入,所以词向量的维度设为768。卷积层中使用3种大小不同的二维卷积核,卷积核的高度分别为2、3和4,卷积核宽度与词向量维度相同,每种卷积核的数量为256。BiLSTM层中的隐藏单元个数为128,dropout的参数大小设置为0.1。每次训练的批次batch_size大小为128,学习率大小为0.0005,每句话的最大长度为32,epoch数为3。

3.3 实验结果与分析

本文将BA-InfoCNN-BiLSTM模型与当前较流行的几种分类方法进行了比较。

(1)TextCNN。由Kim^[13]提出的TextCNN在CNN概念的基础上,让卷积核宽度与词向量维度保持一致进行特征提取,然后拼接最大值池化后的特征,最后送入softmax函数进行分类。

(2)TextRNN。由Liu^[19]提出的TextRNN在LSTM概念的基础上,取单向LSTM最后一个时间步的隐藏层状态向量作为新闻标题的语义表示,然后将该向量送入softmax函数中进行分类。

(3)DPCNN。由Johnson等^[20]提出的一种通过增加卷积神经网络的深度来获取长距离语义关系的模型。

(4)FastText。Facebook于2016年开源的一种文本分类方法,FastText在保证与CNN和RNN等深层网络同等准确率的基础上,提升了训练速度。

(5)Att-BiLSTM。通过在双向长短时记忆网络后引入注意力机制,对双向长短时记忆网络提取的特征分配不同权重,从而突出重要词的作用。

(6)BERT。BERT是一个多层双向的Transformer Encoder模型,主要分为两个阶段:预训练和微调。在预训练阶段,模型会在大量没有标注的语料上进行训练;在微调阶段,模型会对预训练得到的参数进行初始化,然后在进

行下游任务过程中对参数进行调整。

本文使用准确率作为评估指标,实验结果如表 2、图 4 所示。

Table 2 Accuracy

表 2 准确率

模型	新浪新闻	搜狐新闻
TextCNN	91.13	83.14
TextRNN	91.05	83.17
Att-BiLSTM	90.84	81.23
FastText	92.13	82.36
DPCNN	91.14	78.05
Bert	94.41	84.24
BA-InfoCNN-BiLSTM	95.07	84.95

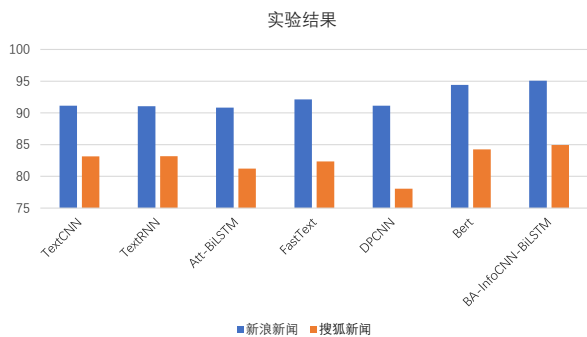


Fig. 4 Experimental results

图 4 实验结果

BA-InfoCNN-BiLSTM 模型在新浪新闻数据集和搜狐新闻数据集上分别获得了 95.07% 与 84.95% 的准确率。与前 6 个模型相比,BA-InfoCNN-BiLSTM 模型取得了最好的效果。与 6 个模型中效果最好的 Bert 模型相比,BA-InfoCNN-BiLSTM 模型在新浪新闻数据集上的准确率提升了 0.66%,在搜狐新闻数据集上的准确率提升了 0.71%,从而证明了 BA-InfoCNN-BiLSTM 模型通过在词嵌入后加入注意力机制补充词的重要程度,再分别捕获多粒度下的局部信息和全局语义信息,可以有效提升模型的准确率。

3.4 消融实验

为了验证 BA-InfoCNN-BiLSTM 模型中不同组件对于模型的有效性,本文通过消融实验进行验证。BA-BiLSTM 为原模型中去掉了改进 TextCNN 层的模型,仅将 BiLSTM 最后一个时间步的隐藏状态向量作为全局语义特征用于分类输出。BERT-InfoCNN-BiLSTM 为原模型中去掉了注意力机制的模型,在嵌入层后使用改进的 TextCNN 层和 BiLSTM 层分别捕捉局部与全局特征,将两种特征融合后输出。BA-InfoCNN 为原模型中去掉了 BiLSTM 层的模型,使用改进的 TextCNN 层捕捉多个粒度下的局部语义特征并将其用于分类,同时将输出改为直接输出。BA-CNN-BiLSTM 为了去掉原模型中对 TextCNN 的改进部分,使用融合后的结果用于分类。消融实验结果如表 3 所示。

首先将 BA-BiLSTM 的实验结果与本文提出的 BA-InfoCNN-BiLSTM(以下简称 BAIB)进行对比,在两个数据集

Table 3 Ablation experiment results

表 3 消融实验结果

模型	新浪新闻	搜狐新闻
BA-BiLSTM	94.67	84.33
BA-InfoCNN	94.86	84.76
BERT-InfoCNN-BiLSTM	94.91	84.82
BA-CNN-BiLSTM	94.73	84.61
BA-InfoCNN-BiLSTM	95.07	84.95

上 BAIB 的效果都明显优于 BA-BiLSTM,说明提取局部信息对分类结果有一定影响。将 BA-InfoCNN 的实验结果与 BAIB 相比,BAIB 的效果要优于 BA-InfoCNN,说明 BiLSTM 能够有效提取全文信息特征,提升模型效果。BAIB 去除注意力机制之后的效果也不如 BAIB,说明使用注意力机制凸出词在句子中的重要性对于提升分类效果也是很有必要的。最有意义的是,将没有改进的 BAA-CNN-BiLSTM 融合模型与 BAIB 进行比较,发现融入前文信息的卷积网络分类更准确,模型对文本语义的理解更充分。

4 结语

本文提出的文本分类模型 BA-InfoCNN-BiLSTM 通过融合改进的卷积神经网络和循环神经网络,解决了传统的单一深度学习网络提取信息不充分、分类效果差的问题。相比于其他融合模型,本模型直接在词嵌入后加入注意力机制,生成权重词向量,突出重要词对整体语义的影响,然后分别送入卷积神经网络和循环神经网络,同时对卷积神经网络进行了改进。在进行卷积操作过程中融入部分前文信息,让卷积神经网络不再仅关注局部信息。最终的实验结果表明,该方法对分类的准确率有一定提升。接下来为了使模型得到进一步优化,可以从以下方面入手:考虑到文本进行分类时,文中存在较多干扰信息以及一些专业性较强的名词,可以在词向量动态训练过程中加入对抗扰动,以进一步提升生成的新闻文本词向量的鲁棒性以及表征能力。

参考文献:

- [1] TAI K S, SOCHER R, MANNING C D. Improved semantic representations from tree-structured long short-term memory network [C]// Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing ACL, 2015: 1556-1566.
- [2] WANG Z Q, SUN X, ZHANG D X, et al. An optimal SVM-based text classification algorithm [C]// Proceedings of 2006 International Conference on Machine Learning and Cybernetics, 2006: 1378-1381.
- [3] ZHANG L G, JIANG L X, LI C Q. et al. Two feature weighting approaches for naive Bayes text classifiers [J]. Knowledge-Based Systems, 2016, 100: 137-144.
- [4] JIANG M H, LI S M, ZHENG H H, et al. Rumor detection model based

- on improved location coding [J]. *Computer Science*, 2022, 49 (8) : 330–335.
- [5] BENGIO Y, DE M R, FLAMMIA G, et al. Global optimization of a neural network–hidden Markov model hybrid[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 1992(2):252–259.
- [6] MIKOLOV T, BOJANOWSKI P, GRAVE E. Enriching word vectors with subword information[J]. *Transactions of the Association for Computational Linguistics*, 2017(5):15–16.
- [7] XIONG Z, SHEN Q, XIONG Y, et al. New generation model of word vector representation based on CBOW or Skip-Gram[J]. *Computers, Materials and Continua*, 2019, 60(1):259–273.
- [8] ZHANG J, LI Z, QU D. An improved recurrent neural network language model with context vector features [C]// Beijing: Proceedings of 2014 IEEE 5th International Conference on Software Engineering and Service Science, 2014.
- [9] PENNINGTON J, SOCHER R, MANNING C D. GloVe: global vectors for word representation [C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 2014:1532–1543.
- [10] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [C]// Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, 2019:4171–4186.
- [11] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. *Neural Computation*, 1997, 9(8):1735–1780.
- [12] KALCHBRENNER N, GREFFENSTETTE E, BLUNSON P. A convolutional neural network for modelling sentences[DB/OL]. <https://arxiv.org/abs/1404.2188v1>.
- [13] KIM Y. Convolutional neural networks for sentence classification [C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 2014:1746–1751.
- [14] CHEN K, LIANG B, KE W D, et al. Chinese micro-blog sentiment analysis based on multi-channels convolutional neural networks [J]. *Journal of Computer Research and Development*, 2018, 55(5):945–957.
陈珂, 梁斌, 柯文德, 等. 基于多通道卷积神经网络的中文微博情感分析[J]. *计算机研究与发展*, 2018, 55(5):945–957.
- [15] LONG F, ZHOU K, OU W H. Sentiment analysis of text based on bidirectional LSTM with multi-head attention [J]. *IEEE Access*, 2019, 7: 141960–141969.
- [16] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [DB/OL]. <https://arxiv.org/abs/1706.03762v2>.
- [17] GRAVES A, JAITLY N, MOHAMED A. Hybrid speech recognition with deep bidirectional LSTM [C]// Olomouc: 2013 IEEE Workshop on Automatic Speech Recognition and Understanding, 2013.
- [18] WANG S, JIANG Y, HOU X E, et al. Cerebral micro-bleed detection based on the convolution neural network with rank based max pooling [J]. *IEEE Access*, 2017(5): 1676–1683.
- [19] LIU J. Convolutional recurrent neural networks for text classification [J]. *Journal of Database Management*, 2021, 32(4):65–82.
- [20] JOHNSON R, ZHANG T. Deep pyramid convolutional neural networks for text categorization [J]. *Journal of Physics: Conference Series*, 2017(1): 562–570.

(责任编辑:黄健)