Journal of Computer Applications
计算机应用, 2020, 40(8): 2189 - 2193
ISSN 1001-9081
CODEN JYIIDU
2020-08-10
http://www.joca.cn

文章编号:1001-9081(2020)08-2189-05
DOI:10.11772/j.issn.1001-9081.2019122114

基于特征聚合的假新闻内容检测模型

何韩森,孙国梓*
(南京邮电大学 计算机学院,南京 210023)
(*通信作者电子邮箱 sun@njupt.edu.cn)

摘　要:针对假新闻内容检测中分类算法模型的检测性能与泛化性能无法兼顾的问题,提出了一种基于特征聚合的假新闻检测模型CCNN。首先,通过双向长短时循环神经网络提取文本的全局时序特征,并采用卷积神经网络(CNN)提取窗口范围内的词语或词组特征;然后,在卷积神经网络池化层之后,采用基于双中心损失训练的特征聚合层;最后,将双向长短时记忆网络(Bi-LSTM)和CNN的特征数据按深度方向拼接成一个向量之后提供给全连接层,采用均匀损失函数uniform-sigmoid训练模型后输出最终的分类结果。实验结果表明,该模型的F1值为80.5%,在训练集和验证集上的差值为1.3个百分点;与传统的支持向量机(SVM)、朴素贝叶斯(NB)和随机森林(RF)模型相比,所提模型的F1值提升了9~14个百分点;与长短时记忆网络(LSTM)、快速文本分类(FastText)等神经网络模型相比,所提模型的泛化性能提升了1.3~2.5个百分点。由此可见,所提模型能够在提高分类性能的同时保证一定的泛化能力,提升整体性能。

关键词:特征聚合;卷积网络;循环网络;均匀损失;泛化性能
中图分类号:TP391.1; TP301.6　　**文献标志码**:A

Fake news content detection model based on feature aggregation

HE Hansen, SUN Guozi*
(*School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing Jiangsu* 210023, *China*)

Abstract: Concerning the problem that detection performance and generalization performance of the classification algorithm model in fake news content detection cannot be taken into account at the same time, a model based on feature aggregation was proposed, namely CCNN (Center-Cluster-Neural-Network). Firstly, the global temporal features of the text were extracted by bi-directional long and short term recurrent neural network, and the word or phrase features in the range of window were extracted by Convolutional Neural Network (CNN). Secondly, the feature aggregation layer based on dual center loss training was selected after the CNN pooling layer. Finally, the feature data of Bi-directional Long-Short Term Memory (Bi-LSTM) and CNN were stitched into a vector in the depth direction and provided to the fully connected layer. And the final classification result was output by the model trained by uniform loss function (uniform-sigmod). Experimental results show that the proposed model has an F1 value of 80.5%, the difference between training and validation sets is 1.3%. Compared with the traditional models such as Support Vector Machines (SVM), Naïve Bayes (NB) and Random Forest (RF), the proposed model has the F1 value increased by 9% - 14%; compared with neural network models such as Long Short Term Memory (LSTM) and FastText, the proposed model has the generalization performance increased by 1.3% - 2.5%. It can be seen that the proposed algorithm can improve the classification performance while ensuring a certain generalization ability, so the overall performance is enhanced.

Key words: feature aggregation; convolutional network; recurrent network; uniform loss; generalization performance

0　引言

随着互联网时代社交媒体的持续发展,假新闻以前所未有的速度覆盖了人们日常生活的各个方面。2016年美国总统大选期间,在南欧小城韦莱斯,这个人口仅5.5万的小镇竟然拥有至少100个支持特朗普的网站,其中很多充斥着各种耸人听闻的假新闻,靠吸引流量来赚取广告费[1]。对假新闻生产者而言,他们通过制造假新闻,诱导用户点击阅读并赚取流量,以低廉的成本获取了巨额利益,他们制造的假新闻也不可避免地对社会产生了负面影响。虚假新闻网站制造假新闻,各大知名互联网平台则对其不加审核地转载。文献[2]的研究结果表明,虚假新闻网站的总流量有一半来自Facebook的首页推荐,Facebook对信誉评估良好的网站引流效果甚微,它们来自Facebook的流量只占到总流量的五分之一。虚假新闻的泛滥一方面源自于媒介平台对新闻信息无选择式的传播,另一方面则受到人群对新闻认知局限性的影响。由于知识获取的有限性,人们本身的知识结构不足以完全实现对假新闻的鉴别。舆论调查机构YouGov的调查表明:近一半的人

收稿日期:2019-12-17;**修回日期**:2020-03-19;**录用日期**:2020-03-24。　**基金项目**:国家自然科学基金资助项目(61502247);数学工程与先进计算国家重点实验室开放课题(2017A10);信息网络安全公安部重点实验室开放课题(C17611)。

作者简介:何韩森(1992—),男,安徽池州人,硕士研究生,主要研究方向:自然语言处理、大数据;　孙国梓(1972—),男,安徽滁州人,教授,博士,主要研究方向:网络空间安全、电子数据取证。

(49%)认为自己可以区分假新闻,但测试结果显示,只有4%的人可以通过标题识别假新闻。同时,人们还会出于愚弄讽刺他人或获取经济利益等个人目的去刻意传播假新闻。

假新闻以猎奇的标题、虚构的事件吸引读者眼球,不但影响了个人对社会事件的准确认知,阻碍了人们对真实新闻的获取,还会对社会和经济的稳定带来负面影响。但是,人工的方法往往难以保证海量新闻内容的质量和真实性,更遑论筛选和阻止假新闻的传播。人工鉴别假新闻的方法同时存在效率低下、具有时滞性等问题。为此,引入人工智能技术,有助于快速有效地减少假新闻在互联网平台的传播,也为社交媒体平台的假新闻自动检测技术的提升提供了可能。

Ajao等[3]提出了一个框架,该框架基于卷积神经网络(Convolutional Neural Network, CNN)和长短时记忆(Long Short-Term Memory, LSTM)神经网络的混合模型来检测和分类来自推特帖子的假新闻消息。由于LSTM神经网络在上下文特征的提取方面表现不佳,而上下文的语义信息可以为假新闻检测任务提供更好的语义特征,所以引入CNN可以较好地弥补LSTM在上下文特征提取方面的缺陷。本文的主要工作有:在Liar数据集[4]上增加了一些新的politifact网站新闻数据,并结合Kaggle数据集和Buzzfeed数据集整理了一个短文本的二分类假新闻数据集;同时提出了一种基于特征聚合的假新闻检测模型CCNN(Center-Cluster-Neural-Network)。

1 相关工作

假新闻检测也属于文本分类的范畴,Kim[5]在2014年便提出了利用CNN对文本进行分类的算法,即TextCNN,该算法利用多个不同大小的卷积核来提取句子中的关键信息,从而更好地捕捉局部相关性;文献[6]中利用CNN从新闻主题的角度检测假新闻,表明CNN可以捕获虚假和真实新闻语料在语法修辞上的差异性,从而实现假新闻的识别工作;而文献[7]中则认为应该结合语言特征分析并采用事实检查的方法来提高假新闻检测分类器的性能。

1.1 现有假新闻数据集

从新闻本身的文本内容来看,有50个单词左右的短文本型新闻,如Liar数据集。该数据集于2017年发布,训练、验证、测试集总计约12 000条记录,而且还提供了一些新闻元数据(类似于作者、作者的党派,还有所在洲等信息);标签划分为6个等级,从几乎为真到几乎为假,内容主要与美国政治相关。Kaggle假新闻检测数据集则包含URLs、Headline、Body三个特征,标签为是否为假,其中假新闻1 872条,真新闻2 137条。另外一种是长文本类型的数据集,如华盛顿大学假新闻数据集,平均单词数量在500左右,记录总数较多,总计约6万条记录。

1.2 现有假新闻分类模型

针对假新闻检测,传统的自然语言处理方法主要从三个方面考虑:基于知识验证、基于上下文、基于内容的风格。基于知识验证,即事实检查,Etzioni等[8]通过从网络提取的内容与相关文档的声明匹配来识别一致性,但该方法受制于网络数据的可信度、质量等挑战。Rashkin等[9]在政治事实检查和假新闻检测的背景下对新闻媒体的语言进行了分析研究,对带有讽刺、恶作剧和宣传的真实新闻进行了比较,证明了事实检查确实是一个具有挑战性的任务。

基于上下文主要依赖新闻所带有的元数据信息和在社交媒体网络中的传播方式。Wang[4]的研究表明,在加入作者等元数据信息之后,对于假新闻检测的性能会有所提升;Zhao等[10]发现在传播过程当中,大多数人转发真实新闻是从一个集中的来源,而虚假新闻则会通过转发其他转发者发布的内容来传播,虚假新闻在社交网络的传播过程中具有明显的多点开花式特征。

基于内容风格是从新闻本身的内容探究假新闻可能具有的特征,主要有语言特征、词汇特征、心理语言学特征等,这包括句子的单词数量、音节、词性、指代、标点、主题等。例如Castillo等[11]从内容中提取基本的语义和情感特征,并从用户中提取统计特征的特征分类模型。Volkova等[12]从语义的角度考虑,例如对文本进行情感打分。而与此相关的子任务是立场检测,Bourgonje等[13]提出检测关于文章的声明与文章本身之间关系的立场检测。

目前这些研究当中,鲜有研究关注到假新闻检测模型的分类性能和泛化性能,因此,本文的研究工作基于此提出了基于特征聚合的假新闻检测模型。

2 CCNN模型设计

CCNN模型结构如图1所示,主要由两个部分组成:一个是由卷积层构成的特征提取和特征聚类模块,一个是由双向记忆神经网络来采集时序特征的模块。CCNN模型整体流程如下:

1)数据获取和标签标注:本文所用的假新闻数据集详见3.1节。

2)文本预处理:先分词(jieba,结巴分词),再将自然语言的文本进行数字向量化并对齐所有句子长度,然后使用预训练词向量矩阵化。

3)特征提取:通过CNN对输入文本进行的卷积操作,不同大小的卷积核能提取到不同种类的特征,本文选择3种大小的卷积核分别提取到对应窗口大小的词汇特征,在经由池化降维处理之后,再采用基于双中心损失函数将特征进行聚类。与此同时进行的还有使用双向循环神经网络(Recurrent Neural Network, RNN)对输入文本进行全局时序特征的采集。然后将以上卷积神经网络和循环神经网络两个部分采集到的不同特征作特征拼接融合,融合之后的特征作为输入文本最后的分类特征。

4)分类结果:利用全连接层和改进之后的均匀损失函数训练模型分类器,并得出最终结果。

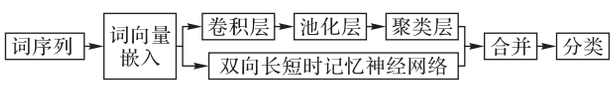

图1 CCNN模型

Fig. 1 Model of CCNN

2.1 CNN网络结构

卷积操作就是利用滤波器对输入层中的每个窗口产生新的特征,传统卷积神经网络在进行卷积操作后会使用池化层,通过对同一滤波器不同尺寸窗口产生的特征取最大值或者取平均值,以此达到减少参数数量的目的,所以,CCNN模型的特征提取层中也设有相应的卷积层和池化层。卷积神经网络的框架结构如图2所示。

与之前池化之后再连接池化层或者输出层不同的是,CCNN模型会在池化层之后连接一个特征聚类层,通过基于双中心损失函数训练的特征聚类层,使得类间距离最大化和类内距离最小化。

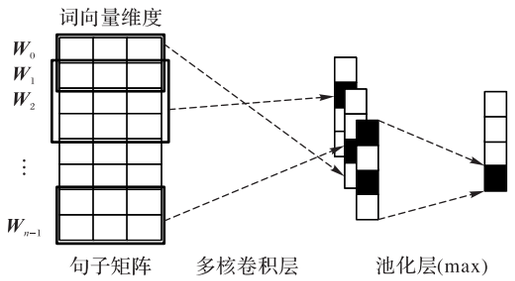

图2　CNN特征提取

Fig. 2　CNN feature extraction

2.2　Bi-LSTM网络结构

在卷积神经网络通过卷积操作来提取一些窗口特征的同时,也采用双向长短时记忆(Bi-directional Long Short Term Memory, Bi-LSTM)网络来对输入的句子进行序列特征的提取。

Bi-LSTM由前向LSTM与后向LSTM组合而成。LSTM建模的一个问题就是无法编码从后到前的信息,而在假新闻检测任务中,语句中情感词、程度词、否定词之间交互对于分类的结果有着重要的作用,通过Bi-LSTM可以更好地捕捉双向的语义依赖。且Bi-LSTM能捕捉长距离的信息,使得特征视野不再局限在窗口中,可以与卷积神经网络形成互补,其LSTM单元结构框架如图2所示。

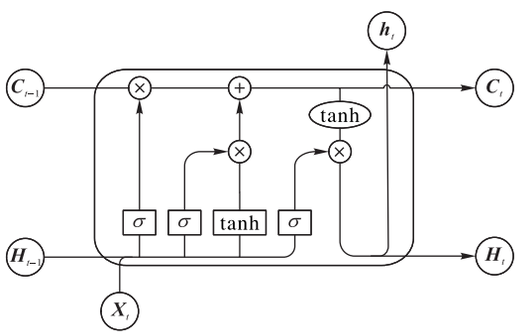

图3　t时刻LSTM结构

Fig. 3　Structure of LSTM at time t

LSTM模型是由t时刻的输入$\boldsymbol{X}_t$、细胞状态$\boldsymbol{C}_t$、临时细胞状态$\tilde{\boldsymbol{C}}_t$、隐层状态$\boldsymbol{h}_t$、遗忘门$\boldsymbol{f}_t$、记忆门$\boldsymbol{i}_t$和输出门$\boldsymbol{o}_t$组成。LSTM通过对细胞状态中已有的信息进行选择性遗忘和记忆对后续时刻计算有用的新信息,使得有用的信息得以传递,而丢弃无用的信息。在每个时间步上都会计算并输出当前时刻的隐层状态$\boldsymbol{h}_t$,其中遗忘、记忆和输出是通过上个时刻的隐层状态$\boldsymbol{h}_{t-1}$和当前的输入$\boldsymbol{X}_t$经由遗忘门$\boldsymbol{f}_t$、记忆门$\boldsymbol{i}_t$、输出门$\boldsymbol{o}_t$来计算控制的。对应的计算公式如式(1)所示:

$$\begin{cases} \boldsymbol{f}_t = \sigma(\boldsymbol{W}_f \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{X}_t] + \boldsymbol{b}_f) \\ \boldsymbol{i}_t = \sigma(\boldsymbol{W}_i \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{X}_t] + \boldsymbol{b}_i) \\ \tilde{\boldsymbol{C}}_t = \tanh(\boldsymbol{W}_C \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{X}_t] + \boldsymbol{b}_C) \\ \boldsymbol{C}_t = \boldsymbol{f}_t * \boldsymbol{C}_{t-1} + \boldsymbol{i}_t * \tilde{\boldsymbol{C}}_t \\ \boldsymbol{o}_t = \sigma(\boldsymbol{W}_o [\boldsymbol{h}_{t-1}, \boldsymbol{X}_t] + \boldsymbol{b}_o) \\ \boldsymbol{h}_t = \boldsymbol{o}_t * \tanh(\boldsymbol{C}_t) \end{cases} \tag{1}$$

将每一个时间步对应的隐层状态收集起来便可得到一个隐层状态序列$\{\boldsymbol{h}_0, \boldsymbol{h}_1, \cdots, \boldsymbol{h}_{n-1}\}$,$n$为句子长度。LSTM得到的是一个正向序列,对于文本也可以说是从左往右的状态序列,而Bi-LSTM在此基础上对文本从右往左进行逆序编码,进而获得逆向的隐层状态序列。最后,将正向和逆向的隐层状态序列进行拼接,继而得到最终的文本向量序列。对于本任务来说,只需取双向最后一个隐层状态向量进行拼接来表示本文的文本,因为它已经包含了正向和逆向的所有信息了。

2.3　特征聚集

通常的假新闻检测模型(二分类)中,在提取完特征之后就将其送入分类器进行分类判别,并且采用sigmoid训练模型,但这时的特征往往并不能很好地区分这些类别,所以本文采用了类似K最近邻(K-Nearest Neighbor, KNN)的方法在特征被送入分类器之前做了进一步的处理,使得特征具有了一定的聚类特性。采用文献[14]中的加聚类惩罚项作为模型的损失函数,函数定义如式(2)所示。

$$loss = -\frac{1}{n}\sum_{i=1}^{n}(y_i \operatorname{lb}\hat{y}_t + (1-y_i)\operatorname{lb}(1-\hat{y}_t)) + \lambda\left\|\boldsymbol{x} - \boldsymbol{c}_{yi}\right\|^2 \tag{2}$$

其中:y_i为正确的样本;$-\frac{1}{n}\sum_{i=1}^{n}(y_i \operatorname{lb}\hat{y}_t + (1-y_i)\operatorname{lb}(1-\hat{y}_t))$就是二分类的交叉熵;$\lambda\left\|\boldsymbol{x} - \boldsymbol{c}_{yi}\right\|^2$就是额外的惩罚项,$\boldsymbol{c}_{yi}$则是每个类别定义的可训练中心。所以,前者用于扩大不同类别中心的距离,而后者则用于缩小当前类别的样本到该类中心的距离。

2.4　均匀损失

选择逻辑回归函数(sigmoid)加交叉熵(cross entropy)这样的搭配训练模型时,往往会使分类器只学习到了正例样本的特征,增加过拟合的风险。因此,为使分类器不至于单纯地去拟合真实的样本分布,增加了一个任务,让它同时也去拟合一下均匀分布,故采用的损失函数如式(3)所示。

$$loss = -(1-\varepsilon)\operatorname{lb}\left(\frac{e^{z_1}}{e^{z_1}+e^{z_2}}\right) - \frac{\varepsilon}{2}\left[\operatorname{lb}\left(\frac{e^{z_1}}{e^{z_1}+e^{z_2}}\right) + \operatorname{lb}\left(\frac{e^{z_2}}{e^{z_1}+e^{z_2}}\right)\right] \tag{3}$$

其中:ε为自定义的变量;e为自然数;$[z_1, z_2]$是模型输出的预测结果。

3　实验与结果分析

3.1　实验数据

由于推特、微博等社交工具的文本长度限制在140词,本文从公开的假新闻数据集和相关假新闻验证网站上搜集整理数据,得到一个短文本假新闻二分类数据集,并命名为pkb假新闻数据集。pkb假新闻数据集的主要来源有:politifact网站、kaggle假新闻竞赛数据集和Buzzfeed数据集。对于politifact网站上的数据,选取其中4个类别,分别是true、false、barely-true和pants-on-fire,后3个类别统一归为假新闻一类;kaggle假新闻竞赛数据集和Buzzfeed数据集按照原有数据集的真假新闻标签分别获取真假新闻数据。pkb假新闻数据集全部为不超过140个词的短文本新闻,标签为0(假,对应表1中的负例列)或1(真,对应表1中的正例列)总计13 070条数据,具体划分见表1。

表1　实验中使用的数据集(pkb)

Tab. 1　Datasets used in the experiment(pkb)

数据集	负例数	正例数
训练集	4 899	5 570
验证集	606	688
测试集	612	695

3.2 实验对比

3.2.1 评价指标

为保证实验的公平及可对比性，针对pkb假新闻数据集0（真新闻）和1（假新闻）两个类别，根据样例真实类别与CCNN预测类别组合划分为真阳性（*TP*）、真阴性（*TN*）、假阳性（*FP*）和假阴性（*FN*）四种类型，混淆矩阵如表2所示。

表2 假新闻检测结果混淆矩阵

Tab. 2 Confusion matrix of fake news detection result

真实结果	预测结果	
	正例（假新闻）	负例（真新闻）
正例（假新闻）	*TP*（预测为假新闻，实际为假新闻）	*FN*（预测为真新闻，实际为假新闻）
负例（真新闻）	*FP*（预测为假新闻，实际为真新闻）	*TN*（预测为真新闻，实际为真新闻）

在假新闻检测任务中，更希望尽可能地检测出假新闻，同时可以顾及用户体验而不致于大量的真实新闻被误伤，所以综合考虑*acc*、*P*、*R*和*F*1作为评价模型性能的标准，计算公式如下：

$$acc = (TP + TN)/(TP + TN + FP + FN)$$

$$P = TP/(TP + FP)$$

$$R = TP/(TP + FN)$$

$$F1 = 2 \times P \times R/(P + R)$$

3.2.2 参数选择

根据文献[8,15-16]等实验发现，在假新闻检测基础模型的选择时，一般有支持向量机（Support Vector Machines，SVM）、朴素贝叶斯（Naïve Bayes，NB）、随机森林（Random Forest，RF）等，故本文也选择这三种模型作为基础模型。

SVM[17]是一种监督学习算法，其决策边界是对学习样本求解的最大边距超平面，适合于小样本的学习，训练速度快且具有较好的鲁棒性；NB是在贝叶斯算法的基础上进行了相应的简化，即假定给定目标值时属性之间相互条件独立，梁柯等[18]的研究表明NB算法具有计算复杂度小、数据缺失对算法影响度小、适合大量数据计算及易于理解等优点；RF是一个包含多个决策树的分类器，其输出是由个别树输出类别的众数决定的，王奕森等[19]研究表明RF对噪声和异常值有较好的容忍性，对高维数据分类问题具有良好的可扩展性和并行性，解决了决策树性能瓶颈的问题。

基础模型选用线性判别分析（Linear Discriminant Analysis，LDA）降维之后的特征数据作为特征输入，神经网络模型除FastText采用自训练的词向量外，其余模型均采用预训练词向量+微调（fine-tune）的模式，对应的预训练词向量采用全局向量（glove），维度为100维。FastText的词组窗口大小设置为3，其余卷积神经网络的卷积核大小设置为[2，3，4]。

3.2.3 结果分析

各模型的性能指标结果如表3所示；测试集与验证集的差值结果见图4和图5，图4对应传统的机器学习模型，图5对应神经网络模型。传统的机器学习模型通过人工提取特征，具有较强的主观性且无法学习到深层次的潜在特征和对应关系。从表3的结果数据中可以观察到，在各项性能指标中，除LSTM外的神经网络模型普遍要优于传统的机器学习模型；再结合模型的泛化能力综合看，在图5中，CCNN的结果较好，各项性能指标退化较少且曲线较为平和；在图4中，随机森林的曲线波动要比其他两个模型稳定，并且提升的幅度也占优，但整体性能不如神经网络模型。

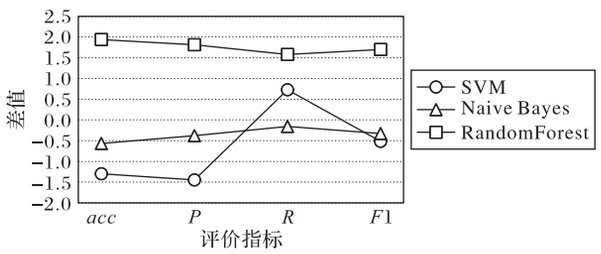

图4 传统的机器学习模型测试集与验证集结果差值

Fig. 4 Result differences between test set and validation set of traditional machine learning model

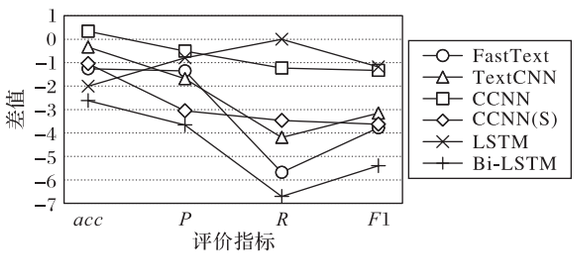

图5 神经网络模型测试集与验证集结果差值

Fig. 5 Result differences between test set and validation set of neural network model

表3 各模型的性能指标对比 单位：%

Tab. 3 Comparison of performance indexes of various models unit：%

模型	验证集				测试集			
	acc	*P*	*R*	*F*1	*acc*	*P*	*R*	*F*1
SVM	61.67	62.34	70.49	66.17	60.37	60.89	71.22	65.65
Naïve Bayes	58.11	56.95	86.92	68.81	57.54	56.57	86.76	68.48
RandomForest	66.46	67.35	71.66	69.44	68.40	69.16	73.24	71.14
FastText	80.60	78.46	83.60	80.86	79.34	77.11	77.92	77.08
TextCNN	80.68	78.19	86.74	82.12	80.34	76.50	82.55	78.96
CCNN	81.22	78.89	85.43	81.78	81.56	78.38	84.20	80.45
CCNN(S)	81.99	80.52	86.34	83.04	80.95	77.46	82.87	79.41
LSTM	48.83	47.67	99.00	64.45	46.82	46.87	99.00	63.27
Bi-LSTM	78.75	76.25	83.76	79.69	76.13	72.59	77.05	74.29

在验证集和测试集上，CCNN(S)和CCNN分别获得了最优的性能。其中，注意到加入了均匀损失函数的CCNN模型在泛化能力上要优于其他的模型，CCNN(S)为损失函数采用传统的二分类交叉熵损失函数，虽然在验证集上的性能达到了最优，但是在测试集上的结果却没能继续保持。从*F*1的性能指标上来看，以卷积神经网络为核心的TextCNN、CCNN和CCNN(S)模型分类性能要略优于以循环神经网络为核心的LSTM和Bi-LSTM模型，这表明利用CNN提取文本特征的方式

要略优于RNN,比传统的基于统计的人工特征更加有效。

文献[20]的研究认为泛化性能是训练好的模型对于不在训练集内数据的拟合能力,检验模型在实际场景中是否能达到跟训练时一样的效果,本质就是反映模型是否学习到了真实的分类特征,还是对数据产生了过拟合,因此本文采用了不参与训练的测试集作为评估模型泛化能力的方法。。

4　结语

本文提出了一种短文本假新闻检测模型CCNN,结合了CNN和RNN模型的优点,同时利用特征聚类和均匀损失提高了模型的泛化能力。在我们整理的一个短文本的二分类假新闻数据集pkb上的实验结果表明,在提高检测性能的基础上,本文模型CCNN的泛化性能也得到了保证。

在接下来的工作中,需要寻找更大规模、更多特征、更具通用性的多模态假新闻数据集;针对算法模型方面,由于数据集的匮乏,也可以考虑转向半监督甚至无监督的学习研究,同时也可以考虑融合知识图谱的相关工作;由于对模型泛化性能的评估缺乏量化计算,依赖实验经验,模型的泛化能力量化评估的相关工作也需要进一步研究。

参考文献 (References)

[1] 秦宇. 探访马其顿假新闻工厂[N]. 南方都市报,2017-02-26(RB13). (QIN Y. The fake-news complex [N]. Nanfang Metropolis Daily, 2017-02-26(RB13).).

[2] PÉREZ-ROSAS V, KLEINBERG B, LEFEVRE A, et al. Automatic detection of fake news [C]// Proceedings of the 27th International Conference on Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2018: 3391-3401.

[3] AJAO O, BHOWMIK D, ZARGARI S. Fake news identification on twitter with hybrid CNN and RNN models[C]// Proceedings of the 9th International Conference on Social Media and Society. New York: ACM, 2018:226-230.

[4] WANG W Y. "Liar, liar pants on fire": a new benchmark dataset for fake news detection [C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2017: 422-426.

[5] KIM Y. Convolutional neural networks for sentence classification [C]// Proceedings of the 2014 Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2014:1746-1751.

[6] O'BRIEN N, LATESSA S, EVANGELOPOULOS G, et al. The language of fake news: opening the black-box of deep learning based detectors [EB/OL]. [2019-11-01]. https://cbmm. mit. edu/sites/default/files/publications/fake-news-paper-NIPS. pdf.

[7] ZHOU Z K, GUAN H, BHAT M M, et al. Fake news detection via NLP is vulnerable to adversarial attacks [C]// ICAART 2019: Proceedings of the 11th International Conference on Agents and Artificial Intelligence. Setúbal, Portugal: Science and Technology Publications, 2019, 2: 794-800.

[8] ETZIONI O, BANKO M, SODERLAND S, et al. Open information extraction from the Web[J]. Communications of the ACM, 2008, 51(12): 68-74.

[9] RASHKIN H, CHOI E, JANG J Y, et al. Truth of varying shades: analyzing language in fake news and political fact-checking[C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2017:2931-2937.

[10] ZHAO Z, ZHAO J, SANO Y, et al. Fake news propagate differently from real news even at early stages of spreading[EB/OL]. [2019-03-09]. https://arxiv. org/ftp/arxiv/papers/1803/1803. 03443. pdf.

[11] CASTILLO C, MENDOZA M, POBLETE B. Information credibility on twitter [C]// Proceedings of the 20th International Conference on World Wide Web. New York: ACM, 2011: 675-684.

[12] VOLKOVA S, SHAFFER K, JANG J Y, et al. Separating facts from fiction: linguistic models to classify suspicious and trusted news posts on twitter[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2017:647-653.

[13] BOURGONJE P, SCHNEIDER J M, REHM G. From clickbait to fake news detection: an approach based on detecting the stance of headlines to articles [C]// Proceedings of the 2017 Empirical Methods in Natural Language Processing: Natural Language Processing meets Journalism. Stroudsburg: Association for Computational Linguistics, 2017: 84-89.

[14] WEN Y, ZHANG K, LI Z, et al. A discriminative feature learning approach for deep face recognition [C]// Proceedings of the 14th European Conference on Computer Vision, LNCS 9911. Cham: Springer, 2016:499-515.

[15] KARIMI H, ROY P, SABA-SADIYA S, et al. Multi-source multi-class fake news detection [C]// Proceedings of the 27th International Conference on Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2018: 1546-1557.

[16] 李力钊,蔡国永,潘角. 基于C-GRU的微博谣言事件检测方法[J]. 山东大学学报(工学版), 2019, 49(2):102-106, 115. (LI L Z, CAI G Y, PAN J. A microblog rumor events detection method base on C-GRU [J]. Journal of Shandong University (Engineering Science), 2019, 49(2):102-106, 115.)

[17] CORTES C, VAPNIK V. Support-vector networks[J]. Machine Learning, 1995, 20(3): 273-297.

[18] 梁柯,李健,陈颖雪,等. 基于朴素贝叶斯的文本情感分类及实现[J]. 智能计算机与应用, 2019, 9(5): 150-153, 157. (LIANG K, LI J, CHEN Y X, et al. Text emotional classification and realization based on Naïve Bayes [J]. Intelligent Computer and Applications, 2019, 9(5):150-153, 157.)

[19] 王奕森,夏树涛. 集成学习之随机森林算法综述[J]. 信息通信技术, 2018, 12(1):49-55. (WANG Y S, XIA S T. A survey of random forests algorithms [J]. Information and Communications Technologies, 2018, 12(1):49-55.)

[20] ZHANG C, BENGIO S, HARDT M, et al. Understanding deep learning requires rethinking generalization [EB/OL]. [2019-02-26]. https://arxiv. org/pdf/1611. 03530. pdf.

This work is partially supported by the National Natural Science Foundation of China (61502247), the Open Project of State Key Laboratory of Mathematical Engineering and Advanced Computing (2017A10), the Open Project of Key Lab of Information Network Security of Ministry of Public Security (C17611).

HE Hansen, born in 1992, M. S. candidate. His research interests include natural language processing, big data.

SUN Guozi, born in 1972, Ph. D., professor. His research interests include cyberspace security, electronic data forensics.