基于多重卷积循环网络舆情分析方法的研究

张 瑜
（西安航空职业技术学院 陕西 西安 710089）

摘要：针对规模化、精确化网络舆情分析的需求，文中对文本情感的分析方法进行了研究。通过结合深度学习中卷积神经网络（CNN）和循环神经网络（RNN）的优点，提出了多重卷积循环网络（CRNN）。该网络既保留了CNN深层次、拟合能力强的特性，又引入RNN中的长短记忆单元（LSTM），提升网络对于长文本序列的分析能力。基于该网络，其对网络舆情的分析方法流程进行了设计。仿真结果表明，所提出的方法在标准数据集NLPCC2013上，准确率、召回率和F_1值相较于RNN、CNN网络分别可以提升约6%、2%和2%。

关键词：舆情分析；情感分析；深度学习；卷积神经网络；循环神经网络
中图分类号：TP311 **文献标识码：**A **文章编号：**1674-6236（2020）18-0092-05
DOI：10.14022/j.issn1674-6236.2020.18.020

Research on public opinion analysis method based on multiple convolution cycle network

ZHANG Yu
（*Xi'an Vocational and Technical College of Aeronautics and Astronautics*, *Xi'an* 710089, *China*）

Abstract: To meet the needs of large-scale and accurate network public opinion analysis, this paper studies the text sentiment analysis method By combining the advantages of Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) in deep learning, a multiple Convolutional Recurrent Neural Network (CRNN) is proposed. The network not only retains the characteristics of CNN's deep level and strong fitting ability, but also introduces the Long Short-Term Memory unit (LSTM) in RNN, which improves the network's analysis ability for long text sequence. Based on the network, this paper designs the analysis method flow of network public opinion. The simulation results show that the accuracy, recall rate and F_1 value of the proposed method can be improved by 6%, 2% and 2% respectively compared with RNN and CNN on the standard data set NLPCC213.

Key words: public opinion analysis; sentiment analysis; deep learning; CNN; RNN

随着互联网技术在居民生活中的深入使用，各大互联网平台成为了舆论产生和传播的主要场所。有效、精准、快速的分析网络舆情信息成为了互联网时代的重要课题之一[1-3]。网络舆情归根到底是文本信息，对网络舆情的分析即对文本信息的分析。有效的文本分析可以提取文本中的情感倾向，从而把控舆情导向[4]。近年来，语义分析技术在机器理论的推动下有了长足的发展。逻辑回归（LR）、支持向量机（SVM）、朴素贝叶斯（NB）等传统的分类算法均可以实现自然语言的处理[5-8]。但这些传统方法在文本分析的精度与效率上均有所欠缺，无法满足规模化、精确化的网络舆情分析需求[9]。

基于以上分析，本文对深度学习网络进行了深入研究，重点研究卷积神经网络（CNN）与循环神经网络（RNN）在文本情感分析时的应用特点。随后，利用CNN网络层次深拟合能力强的特点，结合循环神经网络对长序列信息的处理能力，构建新的多重

收稿日期：2019-11-25 **稿件编号：**201911191
基金项目：2019年陕西高校辅导员工作研究课题（2019FKT35）
作者简介：张 瑜（1992—），女，陕西西安人，硕士，助教。研究方向：网络舆情分析。

卷积循环网络(RCNN)。该网络包括:卷积层和循环层,其中卷积层用于一级特征的提取;循环层使用长短期记忆单元,用于二级特征的提取。最后,算法经过在标准语料集上进行了情感分类的测试,模型相关指标较CNN与RNN网络有较大的提升。

1　理论基础

1.1　卷积神经网络

近年来,在图像识别领域,卷积神经网络(CNN)有着广泛的应用。卷积神经网络是一种典型的前馈神经网络,包含卷积运算[10-11]。其通常具有较深的层次结构,当前在工业界应用最为广泛的神经网络结构是LeCun等提出的LeNet-5。在此基础上,发展出多种卷积神经网络结构,最常见的CNN结构如图1所示[12]。

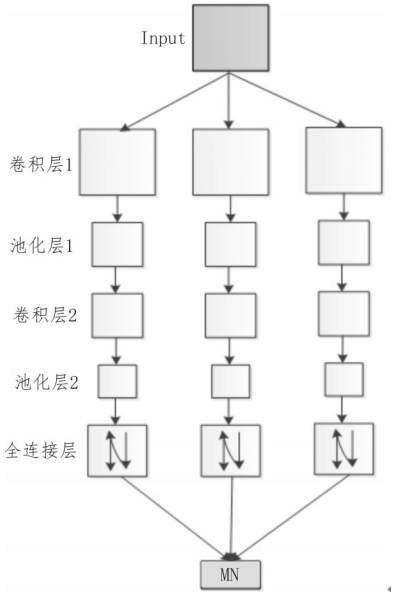

图1　卷积神经网络

图1给出的卷积神经网络包含:卷积层、池化层和全连接层。

1.1.1　卷积层

卷积层主要用于在输入数据中提取特征向量。卷积层主要通过卷积运算相结合各个神经元结构,其特征在于局部结合与全局共享,对于卷积层输入X^l和X^{l+1}:

$$X^{l+1}(i,j)=[X^l \otimes W^l](i,j)+b = \sum_{k=1}^{K_l}\sum_{x=1}^{f}\sum_{y=1}^{f}[X_k^l(s_0 i+x,s_0 j+y)w_k^{l+1}(x,y)]+b \tag{1}$$

其中:

$$(i,j)\in\{0,1,\cdots,L_{l+1}\}$$

$$L_{l+1}=\frac{L_l+2p-f}{s_0}+1$$

式中,W是卷积核,X是二维的特征数据,$X(i,j)$表示特征数据第i行,第j列的数据项;l代表卷积层的序号,L代表数据维度的尺寸,s_0和p表示卷积时的相关参数,b代表偏置。具体的计算示意图如图2所示。

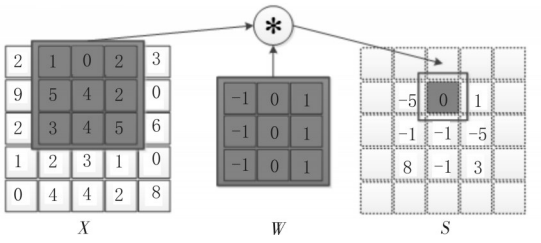

图2　卷积运算示意图

1.1.2　池化层

池化层是对卷积层输出的特征进行特征过滤与选择的过程。池化过程依托于特定的池化函数,本文中用到的池化函数如下:

$$A_k^l(i,j)=\left[\sum_{x=1}^{f}\sum_{y=1}^{f}A_k^l(s_0 i+x,s_0 j+y)^p\right]^{\frac{1}{p}} \tag{2}$$

$$A_k^l=\lambda L_1\left(A_k^l\right)+L_\infty\left(A_k^l\right),\ \lambda\in[0,1] \tag{3}$$

式(2)为L_P池化,当p=1,为平均池化;当p趋近于无穷,为最大值池化。

式(3)为随机池化,该方法可以在指定的池化区域内随机选择特定值,实现神经网络的正则化,避免出现过拟合现象。

1.1.3　全连接层

全连接层位于神经网络的最末端,连接输出层,起到扁平化数据的作用,将特征数据映射到对应的标记空间,经输出层得到整个网络的输出结果。

1.2　循环神经网络

循环神经网络(RNN)在解决序列信息的问题时有着广泛的应用,其网络结构与传统的神经网络有着明显的不同[13-16]。

如图3所示,传统神经网络隐藏层的神经单元为独立的,仅依靠图3中的虚线相结合。而RNN隐藏层的神经元除了处理来自上一层神经元的输入信息,还需处理同一层神经元间的状态信息,即图3中的实线携带信息。在RNN中,随着时间的推移,同

一个神经元可以处理不同的数据序列。

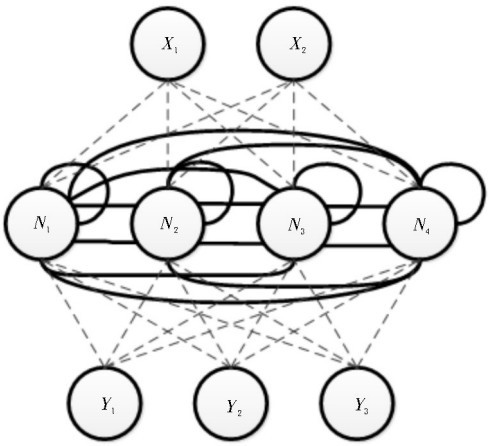

图3　循环神经网络

RNN网络隐藏层的神经元又被称为记忆单元。记忆单元可以保存输入层不同时刻传来的输入状态，其内部的逻辑结构如图4所示。

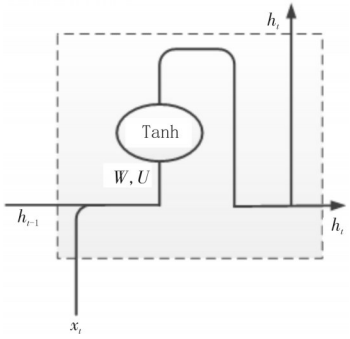

图4　标准RNN的记忆单元

图4中的状态转移方程可以表述为：

$$\boldsymbol{h}_t = Tanh(\boldsymbol{W}\boldsymbol{x}_t + \boldsymbol{U}\boldsymbol{h}_{t-1} + b) \tag{4}$$

其中，$\boldsymbol{x}_t$代表输入矩阵，$\boldsymbol{h}_t$代表输出矩阵。$\boldsymbol{W}$代表权重矩阵，$\boldsymbol{U}$代表自连接矩阵。该模型代表的标准RNN记忆单元，对于较短的序列信息的处理能力较强。但当序列信息较长时，随着时间的推移，到后期时，前期信息便会缺失。因此，本方法引入另一种记忆单元结构，该结构被称为长短时（LSTM）记忆单元，通过引入门限值，使得记忆单元在状态转移时忽略掉一些信息项，以延长信息传递的时间长度。该记忆单元结构，如图5所示。

从图5可以看出，在LSTM记忆单元结构中，引入3个Sigmoid和乘法器组成的遗忘门。该结构的引入可以控制信息在记忆单元结构中传递时遗忘的信息量。该记忆单元的状态转移计算方法如下：

输入序列经过第一个遗忘门时：

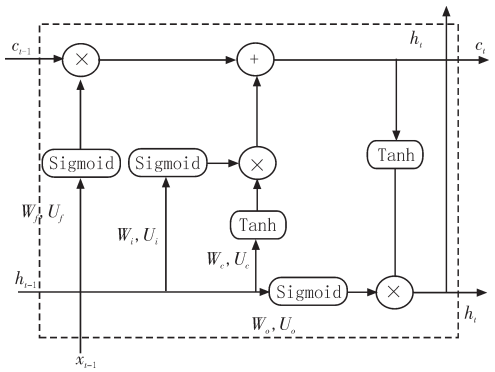

图5　LSTM记忆单元

$$i_t = Sigmoid(W_i x_t + U_i h_{t-1} + b_i) \tag{5}$$

输入序列经过第二个遗忘门时：

$$i_t = Sigmoid(W_i x_t + U_i h_{t-1} + b_i) \tag{6}$$

隐藏层的状态c_t：

$$c_t = f_t \times c_{t-1} + i_t \times Tanh(W_c x_c + U_c h_{t-1} + b_c) \tag{7}$$

最后，计算输出前的遗忘门门限O_t并利用该门限得到输出h_t：

$$\begin{aligned} O_t &= Sigmoid(W_o x_o + U_o h_{t-1} + b_o) \\ h_t &= O_t \times Tanh(c_t) \end{aligned} \tag{8}$$

2　方法实现

上文中对CNN与RNN网络的相关理论进行了介绍。在本节中，结合两种网络，提出多重卷积循环神经网络用于网络舆情的文本情感分析，并给出算法的仿真结果。

2.1　实验设计

将对本文提出的多重卷积循环网络（CRNN）的设计方法进行阐述。其中，卷积层和循环层的计算法方法在上文中已进行了详细的阐述。CRNN方法的整体结构，如图6所示。

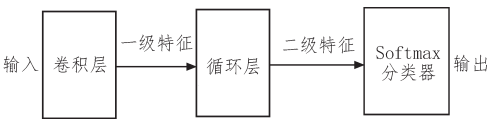

图6　CRNN算法流程

在进行文本情感分析时，需要搜集相关的语料资源，并对这些语料资源进行预处理、文本向量化等操作。最后，利用向量化后的文本完成模型的训练与测试，具体的流程如图7所示。

在语料选择上，本文选择的情感分析语料及其规模，如表1所示。

经预处理后的语料，使用Python中的Jieba工具

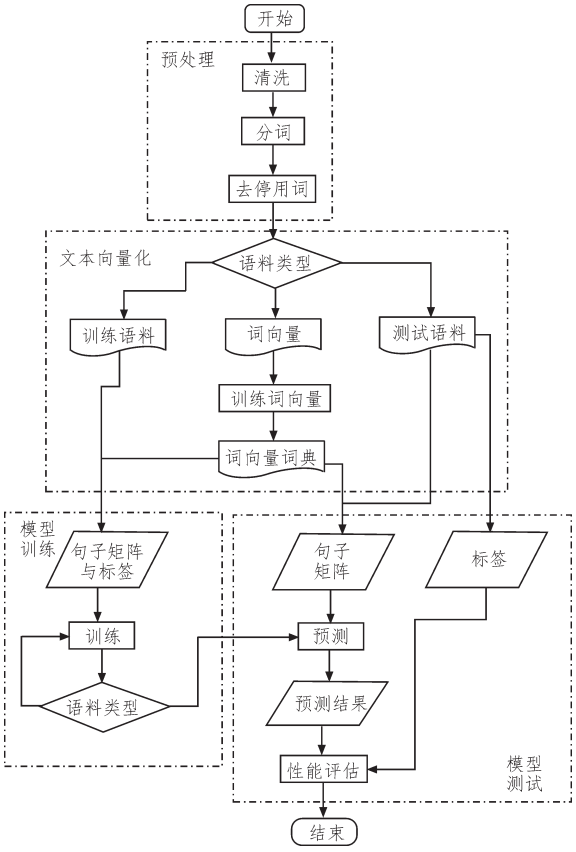

图7　方法总流程

表1　情感分析语料参数

		NLPCC2013-train	NLPCC2013-test
规模	微博	4 000	10 000
	句子	13 000	32 000

完成分词处理。该工具可以生成每条语料所有可能成词情况，然后构成有向的无环图。最后，使用动态规划算法基于词频完成有效的分词。

在构造词向量词典时，使用Word2vec进行训练。调用Word2vec时的相关参数，如表2所示。

表2　Word2vec训练参数设置

参数名称	参数值	描述
Window	5	上线文窗口
Hs	1	采样方式Hierarchical softmax
Sg	1	使用skip-gram模型
Sample	1e-3	无
Negative	0	无

模型训练时，使用Keras框架中的fit()方法，训练中的相关参数使用默认参数，Optimizer使用SGD，学习率设置为0.000 6。

模型测试时，使用的测试指标有：准确率P、召回率R、F_1和MSE。其中，F_1和MSE的定义方法如下：

$$F_1=\frac{2\times P\times R}{P+R} \tag{9}$$

$$MSE=\frac{\sum_{i=1}^{N}(\mathrm{target}_i-\mathrm{output}_i)^2}{N} \tag{10}$$

2.2　仿真结果

在实验中，CRNN网络的性能会随着卷积核数量的变化而改变。因此，在对比分析CRNN网络与RNN、CNN的性能前，必须确定合适的卷积核大小。

表3给出了模型测试时，卷积核长度为5时，各项指标随着卷积核数目的变化情况。

表3　模型指标随卷积核大小的变化

大小	P	R	F_1	MSE
25	24.4%	21.7%	19.5%	0.189
50	23.9%	24.9%	23.5%	0.183
100	27.3%	28.0%	26.2%	0.178
150	28.3%	27.5%	27.1%	0.179
200	31.1%	29.4%	28.9%	0.168
300	29.0%	28.7%	27.9%	0.174
350	29.6%	29.3%	28.9%	0.175

从表3可以看出，当卷积核的大小＜200时，模型的各项指标随卷积核的大小增加而优化；当卷积核的大小为200时，模型的精度P，召回率R达到最优值；当卷积核的大小超过200时，模型指标并未继续优化。对于CRNN网络，由于节点数量限制，只可以保留有限的特征。当卷积核数量超过节点数量的表征能力时，会造成模型的失真，从而影响模型的性能。

根据表3的实验结果，引入相同规模的RNN与CNN网络，并在该数据集上作进一步的实验。此时，卷积核的数量都设置为200，RNN网络中隐藏层的神经元数设置为400，卷积核长度设置为5。经模型参数调试，3个模型可以得到表4所示的测试结果。

表4　CNN、RNN、CRNN网络性能对比

模型	P	R	F_1	MSE
CNN	29.97%	30.43%	29.25%	0.170
RNN	30.10%	29.30%	29.21%	0.172
CRNN	36.64%	32.44%	32.50%	0.163

从表4可看出，CRNN网络在准确率P有约6%的提升，在召回率R和F_1上均有约2%的提升。经过在深度网络中同时引入卷积层和循环层，得到新的深度卷积循环网络可以改善RNN和CNN网络在情感分析时的相应弊端，提升网络的性能。

3 结束语

网络舆情分析是互联网时代的重要课题之一，有效而精准的把控舆情中的情感，及时引导网络中舆论的健康导向。时下，基于各种深度学习的文本情感分析方法被应用到舆情分析中。本文提出的CRNN网络，通过引入卷积层和循环层，可以有效地避免传统的RNN与CNN网络在文本分析时的缺陷，提升深度学习网络的性能指标。

参考文献：

[1] 杜雨萌，张伟男，刘挺.基于主题增强卷积神经网络的用户兴趣识别[J].计算机研究与发展，2018，55(1):188-197.

[2] 刘娇，崔荣一，赵亚慧.基于自联想记忆与卷积神经网络的跨语言情感分类[J].中文信息学报，2018，32(12):118-124.

[3] 牛连强，陈向震，张胜男，等.深度连续卷积神经网络模型构建与性能分析[J].沈阳工业大学学报，2016，38(6):662-666.

[4] 骆正茂.结合卷积神经网络不同层的特征进行包类商品检索[J].计算机应用与软件，2018(1):195-199.

[5] 闫河，王鹏，董莺艳，等.改进的卷积神经网络图片分类识别方法[J].计算机应用与软件，2018，35(12):193-198.

[6] 金占勇，田亚鹏，白莽.基于长短时记忆网络的突发灾害事件网络舆情情感识别研究[J].情报科学，2019，37(5):142-147，154.

[7] 黄磊，杜昌顺.融合群稀疏与排他性稀疏正则项的神经网络压缩情感分析方法[J].北京化工大学学报:自然科学版，2019，46(2):103-112.

[8] 王努努，张伟佳，钮亮.基于ARIMA和BP神经网络模型的舆情情感预测[J].电子科技，2016，29(5):83-87.

[9] 孙靖超，高见，胡啸峰.基于改进注意力模型的网络舆情趋势预测研究[J].情报杂志，2018，37(11):120-125.

[10] 毛雪娇，谈健，姚颖蓓，等.基于长短期记忆神经网络的饱和负荷预测方法及应用[J].水电能源科学，2019(6):192-195.

[11] 黄友文，万超伦，冯恒.基于卷积神经网络与长短期记忆神经网络的多特征融合人体行为识别算法[J].激光与光电子学进展，2019，56(7):243-249.

[12] 龚琴，雷曼，王纪超，等.基于注意力机制的卷积双向长短期记忆模型跨领域情感分类方法[J].计算机应用，2019，39(8):2186-2191.

[13] 丁盼，庞晓平，陈进，等.基于CNN-BiLSTM网络引入注意力模型的文本情感分析[J].武汉工程大学学报，2019(4):386-391.

[14] 罗帆，王厚峰.结合RNN和CNN层次化网络的中文文本情感分类[J].北京大学学报:自然科学版，2018，54(3):459-465.

[15] 唐贤伦，林文星，杜一铭，等.基于串并行卷积门阀循环神经网络的短文本特征提取与分类[J].工程科学与技术，2019(4):125-132.

[16] 郜丽鹏，郑辉.基于PReLUs-Softplus非线性激励函数的卷积神经网络[J].沈阳工业大学学报，2018，40(1):54-59.

(上接第91页)

[12] 凌健中. WirelessHART协议栈的设计与实现[D].成都:电子科技大学，2013.

[13] Liu W，Sheng X U，Zhang X，et al. Design and implementation of wireless sensor network system based on WirelessHART[J]. Chinese Journal of Electron Devices，2017.

[14] Chen G，Cao X，Liu L，et al. Joint scheduling and channel allocation for end-to-end delay minimization in industrial WirelessHART networks[J]. IEEE Internet of Things Journal，2019，6(2):2829-2842.

[15] Samek M.，Quantum platform programmer's manual [EB/OL]. [2006]. North Carolina，Pittsboro，http://www.quantum-leeps.com/doc/QP_mannual.pdf.

[16] 刘红蕾.面向时间不确定模型的复杂事件处理技术研究[D].沈阳:东北大学，2019.