

## 短文本分类的 ResLCNN 模型<sup>\*</sup>

王俊丽, 杨亚星, 王小敏

(同济大学 电子与信息工程学院 计算机科学与技术系, 上海 201804)

通讯作者: 杨亚星, E-mail: yang.yaxing@outlook.com



**摘要:** 短文本分类是互联网文本数据处理中的关键任务之一。长短时记忆网络 LSTM(long short-term memory) 和卷积神经网络 CNN(convolutional neural network) 是广泛应用于短文本分类任务的两种深度学习模型。在计算机视觉和语音识别领域的深度学习研究表明, 深层次的神经网络模型具有较好的表达数据特征的能力。受此启发, 面向文本深度学习分类问题, 提出基于 3 层 LSTM 和 CNN 网络结构的 ResLCNN(residual-LSTM-CNN) 深度学习模型。该模型有效结合 LSTM 获取文本序列数据的长距离依赖特征和 CNN 通过卷积操作获取句子局部特征的优势, 同时借鉴残差模型理论, 在第 1 层 LSTM 层与 CNN 层之间加入恒等映射, 构建残差层, 缓解深层模型梯度消失问题。为了探究深层短文本分类中 ResLCNN 模型的文本分类能力, 在多种数据集上将其与 LSTM、CNN 及其组合模型进行对比实验。结果表明, 相比于单层 LSTM 与 CNN 组合模型, ResLCNN 深层模型在 MR、SST-2 和 SST-5 数据集上分别提高了 1.0%、0.5%、0.47% 的准确率, 取得了更好的分类效果。

**关键词:** 深度学习模型; 短文本分类; 长短时记忆网络; 卷积神经网络; 残差网络

中文引用格式: 王俊丽, 杨亚星, 王小敏. 短文本分类的 ResLCNN 模型. 软件学报, 2017, 28(Suppl. (2)): 61-69. <http://www.jos.org.cn/1000-9825/17019.htm>

英文引用格式: Wang JL, Yang YX, Wang XM. ResLCNN model for short text classification. Ruan Jian Xue Bao/Journal of Software, 2017, 28(Suppl. (2)): 61-69 (in Chinese). <http://www.jos.org.cn/1000-9825/17019.htm>

### ResLCNN Model for Short Text Classification

WANG Jun-Li, YANG Ya-Xing, WANG Xiao-Min

(Department of Computer Science and Technology, College of Electronics and Information Engineering, Tongji University, Shanghai 201804, China)

**Abstract:** The short text classification is a key task in the field of Internet text data processing. The long short-term memory (LSTM) and convolutional neural network (CNN) are the two most important deep learning models for short text classification. The research on the deep learning in the field of computer vision and speech recognition shows that the deep level of neural network model has better ability to express data features. Inspired by this, a deep learning model named ResLCNN (residual-LSTM-CNN) is proposed based on the structure of three LSTM layers and one CNN layer for text deep learning classification problem. In this model, the LSTM layer is used to capture long distance dependency features of the sequence data and the CNN layer can extract local features of the sentence by convolution operators. The ResLCNN model combines the advantages of LSTM and CNN effectively. At the same time, based on the residual model theory, the ResLCNN model adds an identity mapping between the first LSTM layer and CNN layer to alleviate the problem of vanishing gradients. In order to explore the ability of ResLCNN model for deep short text classification, some experiments are made on several data sets to compare with LSTM, CNN and their combination models. The result shows that compared with the single

\* 基金项目: 国家高技术研究发展计划(863)(2015IM030300); 上海市科技创新计划(15DZ1101202); 上海市科委项目(14JC1405800); 同济大学中央高校基本科研业务费

Foundation item: National High-Tech R&D Program of China (863) (2015IM030300); Shanghai Science and Technology Innovation Program (15DZ1101202); Shanghai Science and Technology Commission Project (14JC1405800); Fundamental Research Funds for the Central Universities of Tongji University

收稿时间: 2017-06-30; 采用时间: 2017-10-20

LSTM and CNN combination model, the ResLCNN deep model improves the accuracy rate by 1.0%, 0.5% and 0.47% respectively on the data sets of MR, SST-2 and SST-5 and achieves better classification results.

**Key words:** deep learning model; short text classification; long short-term memory networks; convolutional neural networks; residual network

随着互联网技术的快速发展,作为人们了解社会动态的直接信息来源,网络文本信息呈现出大数据信息爆炸式增长的特点.面对如此海量的信息,如何能够快速筛选和分类,进而获取有效信息变得越来越困难.而短文本分类正是互联网文本数据处理中的关键任务之一,可以应用于微博情感分析<sup>[1,2]</sup>、商品评论情感倾向区分<sup>[3]</sup>等.作为短文本分类的关键模型,句子模型是学习文本特征对句子进行表征,对于情感识别、问答系统、翻译等文本处理任务都有重要意义.传统的句子模型使用的是词袋模型,是一种基于向量空间模型表示方法,句子和文档被看做是无序的单词集合,每个特征词之间相互独立.这一类模型不包含词序和语法信息,并且普遍具有维度灾难、稀疏等问题.所以,近年来随着研究的不断深入,对词序敏感的深度神经网络框架成为句子模型的主流框架.

不同于传统的方法,神经网络句子模型是基于词向量来训练的.词向量最早是由 Hinton<sup>[4]</sup>提出的,将词表达为低维度的向量,可以用欧氏距离来衡量向量的距离.而后, Mikolov<sup>[5]</sup>采用 Skip-gram 模型来预测上下文训练词向量,共现信息被有效地用于表示单词,可以更好地表达出语义或句法特征.因此,词向量具有两大特性:一是语义相近的词向量距离相近;二是词向量线性运算也具有意义,例如:

$$\text{vec}(\textit{Athlete}) + \text{vec}(\textit{Football}) \approx \text{vec}(\textit{Football\_Player})$$

另外,神经网络句子模型能够自己学习如何抓取特征,这些特征比人为处理的特征在后期的自然语言处理任务中表现得更好. Bengio 在讲述深度学习理论动因时,将神经网络分为两部分:特征层(features)和映射层(mapping layer).特征层是学习提取输入数据的特征;映射层是将获取的特征映射到输出层,用于分类或者递归任务.基于深度学习的句子模型关键是能够通过多个特征层学习到更复杂的特征.卷积神经网络 CNN(convolutional neural network)和循环神经网络 RNN(recurrent neural network)是应用到句子模型的两种典型深度学习框架.

卷积神经网络通过不同大小的卷积核来获取句子中不同的词组特征,并利用池化操作提取不同长度、不同位置的重要词组特性.同时,采用权重共享的策略在空间上获取较好的局部特征,在计算机视觉、语音识别领域取得了领先的研究成果<sup>[6]</sup>,随后在语义分析、查询检索、句子建模等文本处理任务中都取得了较好的实验效果.

循环神经网络是另一个广泛应用在文本分类任务中的神经网络框架.它在时间结构上存在共享特性,能够顺序处理任意长度的句子.但是 RNN 存在梯度消失和梯度爆炸的问题.为了解决该问题,能够更好地记忆信息,长短时记忆网络 LSTM(long short-term memory)应运而生,并在机器翻译、情感分类、答案排序、事件获取等方面都表现良好.

CNN 和 RNN 是比较典型的两种获取特征的神经网络结构.CNN 利用卷积核在空间上获取局部特征,RNN 能够在时间序列上获取依赖特征,各有优势.目前,在一些任务中,已经有结合 CNN 与 RNN 的相关研究,通过训练卷积神经网络获取空间上的特征,并利用长短时记忆网络结合序列方向上的特征,在空间结构和时间序列特征比较重要的任务中表现优异,如图片描述、语音识别、语言模型、句子文档建模等.

深度学习的优势在于通过多个特征层得到更复杂、更高级的特征.但是,面临的一个挑战是梯度消失问题.Donahue 等人<sup>[7]</sup>的研究证实多层 LSTM 由于梯度消失问题出现了性能退化,效果不如两层 LSTM 好.主要因为误差无法有效地反向传播到上层网络,致使上层网络的参数没能充分地优化.残差网络就是为改进深层神经网络梯度消失而提出的.He 等人<sup>[8]</sup>对残差网络的深入研究得出同深度的残差网络效果更好,并且计算量更小.

综上所述,考虑利用长短时记忆网络和卷积神经网络,结合文本时序和空间两个层次获取特征,加深网络层以学习到更复杂的特征.本文提出了 ResLCNN(residual-LSTM-CNN)模型,包含 3 层 LSTM 和 CNN,同时借鉴残差网络理论,在第 1 层 LSTM 层与 CNN 层之间加入恒等映射,构建残差层,缓解深层模型梯度消失问题,优化深度网络的性能.本文将 ResLCNN 模型应用在多种分类任务中,并与 LSTM、CNN 及其组合模型进行对比实验.

本文第 1 节介绍目前卷积神经网络、循环神经网络以及残差网络的相关工作.第 2 节对主要模型进行了详

细的描述,包括长短时记忆网络、卷积神经网络以及本文提出的 ResLcNN 模型.第 3 节展示了实验数据和训练参数设置,并进行了模型对比分析和句子长度分析.第 4 节总结本文的研究工作.

## 1 相关工作

基于神经网络模型的深度学习技术在自然语言处理中取得了巨大成功,例如:语言模型、句子和文档模型等,在机器翻译、词性标注、情感分类等任务中表现优异.词向量是语言模型的附加产物,能较好地表达语义和语法信息,可以应用在很多任务模型中.目前,Word2Vec<sup>[5]</sup>(谷歌词向量)和 GloVe<sup>[9]</sup>(全局词向量)应用最为广泛.

最初,卷积神经网络由 LeCun 提出并应用于图像处理<sup>[10]</sup>,随后被 Collobert 等人用在自然语言处理任务中<sup>[11]</sup>.2014 年,Kalchbrenner 等人<sup>[12]</sup>提出了用于句子建模的动态卷积神经网络模型 DCNN(dynamic convolutional neural network),卷积层数和池化层都随着句子长度而变化.随后,Kim<sup>[13]</sup>提出了单卷积层的卷积神经网络句子模型,卷积运算以单词为单位.它们是很多基于卷积神经网络的句子模型变体的来源.Wang 等人<sup>[14]</sup>在 Kalchbrenner 的 DCNN 模型上,增加了多比例的语义单元层,通过筛选有意义的语义单元,缓解短文本信息稀疏和模糊问题,改进短文本分类效果.Ma 等人<sup>[15]</sup>基于 Kim 的模型,结合树结构模型提出了基于依赖的卷积神经网络模型 DBCNN(dependency-based convolutional neural network),在部分任务上取得了领先的效果.同时,Mou 等人<sup>[16]</sup>受循环神经网络结构启发,为了更好地获取句子结构信息,提出了一种结合树结构的句子模型 TBCNN(tree-based convolutional neural network),改进了卷积方式,不仅在短文本分类任务中表现良好,还在程序语言处理中取得较好的实验结果<sup>[17]</sup>.而 Yin 等人<sup>[18]</sup>的句子分类模型 MVCNN(multichannel variable-size convolutional neural network)中,充分利用词向量信息,将包括 Word2Vec、GloVe 等在内的 5 种词向量作为卷积神经网络的 5 个通道输入,并结合了 Kalchbrenner 动态池化思想,取得不错的效果.此外,Kim 等人<sup>[19]</sup>通过改进卷积神经网络 Kernel,缓解了短文本分类中的稀疏问题,优化了分类结果.

循环神经网络能够将不同长度的句子作为输入,并在时间轴上存储记忆信息.基于建模时序数据的特性,循环神经网络广泛应用在自然语言处理领域任务中.LSTM 早在 1997 年就被 Hochreiter 和 Schmidhuber 提出.2014 年,Ilya 等人<sup>[20]</sup>利用 LSTM 来处理统计机器翻译任务,并取得了显著的成果.此后,LSTM 迅速在自然语言处理领域流行,并衍生出变体,如双向长短时记忆模型 Bi-LSTM<sup>[21]</sup>,结合树结构的变种 Tree-LSTM<sup>[22]</sup>等.

考虑到卷积神经网络和长短时记忆模型各自的优缺点,将 CNN 与 LSTM 联合优势互补的研究非常可行.不仅在图像描述<sup>[23]</sup>和语音识别<sup>[24]</sup>领域,自然语言处理领域的 CNN 和 LSTM 的结合研究也逐步增多.Chiu 等人<sup>[25]</sup>在命名体识别任务中,结合 CNN 和双向 LSTM 来获取字符和单词的表征.Kim<sup>[26]</sup>采用同样的理念,将其用在语言模型中,以字符表征作为输入,获取单词的表征,不同的是,Kim 使用 Highway Network 连接卷积神经网络和长短时记忆模型,在多种形态丰富的语言上表现良好.对于文本分类任务,Zhou 等人<sup>[27]</sup>提出 C-LSTM(convolutional based long short-term memory)模型,将 CNN 卷积得到的特征作为 LSTM 模型的输入进行句子表达和文本分类.此外,Zhang 等人<sup>[28]</sup>的 DSCNN(dependency sensitive convolutional neural networks)模型则先利用 LSTM 模型预训练单词和句子的向量表达,随后输入到 CNN 模型进行分类,得到了更好的实验效果.

众所周知,深层神经网络能够增强特征表达和概括的能力,但也正是由于网络的深度逐渐增加,使得训练越来越困难.为解决此问题,Srivastava 等人<sup>[29]</sup>提出了 Highway Network,网络中一些层增加了跳跃连接,并通过门控结合原网络层数据,作为接下来网络层的输入.2016 年,He 等人<sup>[8]</sup>使用恒等映射来使得深层网络拟合出低层网络与目标函数间的残差函数.同时,他们对残差网络中恒等映射也进行了深入研究<sup>[30]</sup>.受此启发,本文提出了深层文本分类模型 ResLcNN 模型,结合长短时记忆网络和卷积神经网络,同时增加恒等映射构建残差层,来缓解梯度消失问题.

## 2 模型框架

### 2.1 长短时记忆网络

长短时记忆网络是为解决循环神经网络长距离传输中的梯度消失和梯度爆炸问题而提出的一种改进模型.不同于初始循环神经网络单元,LSTM 的每个单元有 4 个神经网络层,分别是输入门  $i_t$ 、遗忘门  $f_t$ 、输出门  $o_t$ 、记忆单元  $C_t$ ,它们被以一种特别的方式结合起来,得出隐藏层输出  $h_t$ .

其中记忆单元  $C_t$  是 LSTM 的关键.在整个模型的运行过程中,它就像一条贯穿其中的传送带,仅有一些线性变换,累积历史信息.简单来说,每一步 LSTM 根据输入  $x_t$ 、 $h_{t-1}$ 、 $C_{t-1}$  计算出当前时刻的  $h_t$ 、 $C_t$ ,计算公式如下:

$$\left. \begin{aligned} i_t &= \sigma(W_{xi} \cdot x_t + W_{hi} \cdot h_{t-1} + b_i) \\ f_t &= \sigma(W_{xf} \cdot x_t + W_{hf} \cdot h_{t-1} + b_f) \\ o_t &= \sigma(W_{xo} \cdot x_t + W_{ho} \cdot h_{t-1} + b_o) \\ \tilde{C}_t &= \tanh(W_{xc} \cdot x_t + W_{hc} \cdot h_{t-1} + b_c) \\ C_t &= f_t \times C_{t-1} + i_t \times \tilde{C}_t \\ h_t &= o_t \times \tanh(C_t) \end{aligned} \right\} \quad (1)$$

公式(1)中, $\sigma(\cdot)$ 和  $\tanh(\cdot)$ 分别代表的是 sigmoid 函数和双曲正切函数.输入门  $i_t$ 、遗忘门  $f_t$ 、输出门  $o_t$  这 3 个门是用来控制信息量的,相当于过滤器的作用.上一时刻的记忆单元  $C_{t-1}$  经过遗忘门  $f_t$  得到从上一时刻累积流入的信息,当前输入信息  $\tilde{C}_t$  经过输入门  $i_t$  得到当前更新流入的信息,两者相加获得当前时刻记忆单元  $C_t$ .而当前时刻隐藏层的输出  $h_t$  是从  $C_t$  计算来的,将  $C_t$  通过非线性的  $\tanh()$  函数处理后,再由输出门  $o_t$  过滤.

总的来说,LSTM 是根据输入门  $i_t$ 、遗忘门  $f_t$ 、输出门  $o_t$  过滤信息,并通过线性自连接的记忆单元来累积信息,将其作为中间物来计算当前时刻隐藏层输出  $h_t$ .

### 2.2 卷积神经网络

卷积神经网络连接着 LSTM 层,在不同大小滤波器的作用下对输入进行卷积运算.卷积运算是指一个滤波器  $W \in R^{h \times k}$  作用在  $h$  个向量维度为  $k$  单词上,获得一个新的特征.例如,特征  $c_i$  是由窗口中单词  $x_{i:i+h-1}$  通过式(2)计算得到的.

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

其中, $b \in R$  是偏差因子  $f$  是非线性函数,例如双曲正切函数. $x_{i:i+h-1}$  表示句子中第  $i$  个单词到第  $i+h-1$  个单词组成的词向量矩阵.滤波器也可以称为卷积核,对长度为  $n$  的句子进行局部卷积,作用在滑动窗口中的单词  $\{x_{1:h}, x_{2:h+1}, \dots, x_{n-h+1:n}\}$  上,得到一个特征图:

$$c = [c_1, c_2, \dots, c_{n-h+1}] \quad (3)$$

其中, $c \in R^{n-h+1}$ .

通过卷积层对单词进行卷积运算,将低级特征组合形成更高级的特征表示,可以得到更高级的特征信息,这可以直观地理解为从单词到词组的特征学习,比如学习词组 neural network 的含义.然后通过池化层(max-over-time pooling),针对卷积得到的若干特征值,保留最有价值的特征.

### 2.3 ResLCNN 模型

借鉴 LSTM 和 CNN 在特征获取方面各自的优势,本文将结合 LSTM 和 CNN 进行建模,采用多层 LSTM 优化词向量,给词向量添加时序信息,再通过 CNN 层获取空间组合特征,从而更有效地进行分类.

目前,文本分类句子模型通常是浅层网络,即一层或两层 CNN 或 LSTM,相对于深层神经网络特征表达能力较差.但是深层神经网络模型效果受梯度消失影响.在 He 等人<sup>[8]</sup>研究中,34 层的 CNN 比 18 层的 CNN 效果差.同样,在 Donahue 等人<sup>[7]</sup>研究中也提到,双层的 LSTM 模型效果是最好的,多层的 LSTM 效果出现退化现象.为解决此退化问题,He 等人<sup>[8]</sup>提出了残差网络.受此启发,为了提高模型性能,本文提出了 ResLCNN 模型,如图 1 所示.

ResLCNN 模型以 Word2vec 和 GloVe 词向量构成的句子矩阵作为输入,第 1 层 LSTM 根据隐藏层和记忆单

元获取文本时序特征,随后残差层包含两层 LSTM,接着 CNN 层使用多个卷积核进行卷积运算,并提取最优卷积特征通过 softmax 分类.ResLCNN 模型为了获取更复杂的特征,将 3 层 LSTM 结合 CNN 构建深层网络,同时也在第 1 层 LSTM 输出与 CNN 层输入之间添加恒等映射,构建残差层缓解训练模型阶段低网络层梯度消失问题。

设  $M_i$  代表第  $i$  层的输入,其中  $i \in \{1,2,3,4\}$ 。 $F$  表示残差函数,权重为  $W$ 。残差层运算如下:

$$M_4 = F(M_2, W) + M_2 \tag{4}$$

残差层的输入为  $M_2$ ,输出直接作为第 4 层 CNN 的输入  $M_4$ ,进而得出公式(4)。公式(4)反向求导的过程为

$$\frac{\partial \varepsilon}{\partial M_2} = \frac{\partial \varepsilon}{\partial M_4} \frac{\partial M_4}{\partial M_2} = \frac{\partial \varepsilon}{\partial M_4} \left( 1 + \frac{\partial}{\partial M_2} F(M_2, W) \right) \tag{5}$$

公式(5)中,在括号中第 2 项  $\frac{\partial}{\partial M_2} F(M_2, W)$  不可能为 -1 的条件下,不管梯度  $\varepsilon$  多小,第 1 项可以直接把深层的梯度传递到浅层,从而梯度消失问题得到缓解。

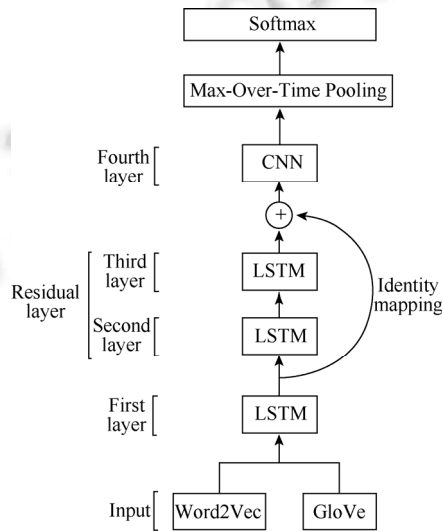


Fig.1 The ResLCNN model framework

图 1 ResLCNN 模型框架

### 3 性能评测与分析

#### 3.1 数据

影评数据 MR (movie review data)是由 Pang 等人<sup>[31]</sup>提供的情感分类数据集.该数据集包括相同数量的正面和负面评论,共 10 662 条,每条评论都是一条短句.实验中训练集、验证集和测试集的比例是 8:1:1.

另一个广泛使用的情感分类数据集是斯坦福情感数据树库 SST (stanford sentiment treebank)由 Socher 等人<sup>[32]</sup>提供.SST-5 表示将评论分为 5 种标签:非常负面、负面、中立、正面、非常正面.实验时,训练集、验证集、测试集数量分别为 8 544、1 101、2 210.除去中立的评论,数据集还可以分为负面和正面两种,用 SST-2 表示.实验时有 6 920 条训练集、872 条验证集以及 1 821 条测试集.

此外,ResLCNN 模型将用于问题分类任务,使用的是问句类型数据集 TREC(text retrieval conference)<sup>[33]</sup>.其中,句子和问题被分为 6 个类型:缩写、实体、描述、人物、地点、数值.其中,测试数据 500 条,5 452 条训练数据中 10%作为验证集.

#### 3.2 训练参数设置

两种预训练的词向量 Word2vec 和 GloVe 组成了本文模型的双通道输入.3 层 LSTM 的隐藏层大小都为

300, CNN 的卷积核大小为 3, 数量为 100. 由 Krizhevsky<sup>[34]</sup> 对非线性激活函数 ReLu 和普通 Sigmoid 系函数做的对比测试得出, ReLu 函数使得学习周期大大缩短. 综合速率和效率, 本文采用非线性激活函数 ReLu. 对于正则化, 采用了局部正则化(batch normalization)和 Dropout 操作<sup>[35]</sup>, Dropout 用在 LSTM 输出层, 取值为 0.5. 没有使用约束各层的参数. 本文使用 Adadelta 梯度下降算法, 最小化预测和真实类别标签的交叉熵损失函数. 在验证集准确率不再升高、损失不再下降时, 停止训练. 对于各数据集, 模型参数设置相同, 所有实验数据都采用 10 次实验的平均结果.

实验是基于 TensorFlow 框架, 使用 Python 实现的. 运行在 Linux 平台上, 双 GPU, 型号为 GTX1080. 以 MR 数据集为例, ResLCNN 模型每分钟处理 3 864 条文本.

### 3.3 模型对比分析

在 MR、TREC、SST-2 以及 SST-5 这 4 种数据集上, 将 ResLCNN 模型与若干以 CNN 和 LSTM 为基本层进行组合的模型进行了实验分析和比较. 实验结果见表 1. 性能指标是测试集的分类准确率, 以百分号为单位.

表 1 中, 第 1 部分是单层 CNN 和 LSTM. 第 2 部分是 CNN、LSTM 组合层. CNN+LSTM 与 LSTM+CNN 都表示单层 CNN 与单层 LSTM 的组合模型, 不同的是前者是 CNN 层在前, 后者是 LSTM 层在前. 而  $i$ LSTM+CNN 则表示  $i$  层 LSTM 后接一层 CNN,  $i \in \{2, 3\}$ . 第 3 部分是本文提出的 ResLCNN 模型及其变种. ResLCNN+res 是 6 层深度网络模型, 它在 ResLCNN 模型基础上在 CNN 层之前再添加一层残差层, 残差块中包括 2 层 LSTM.

Table 1 Classification accuracy of different models

表 1 模型分类准确率

方法	MR	SST-2	SST-5	TREC
CNN	80.31	83.67	43.05	92.46
LSTM	81.16	84.90	44.71	90.08
CNN+LSTM	79.30	82.37	41.81	89.40
LSTM+CNN	81.29	84.95	45.03	92.64
2LSTM+CNN	77.45	81.39	40.44	89.34
3LSTM+CNN	76.93	80.35	32.85	88.54
ResLCNN	82.29	85.45	45.50	92.08
ResLCNN+res	79.56	83.17	41.45	90.94

对于所有模型, 如果 LSTM 层在最后, 则在分类之前有一个平均池化操作. 其他的参数与 ResLCNN 设置参数相同. CNN+LSTM 模型为了达到更好的分类效果, CNN 卷积核数量设置为 150. 各模型训练时间相差不大, 以 MR 数据集为例, ResLCNN 模型每 epoch 的平均训练时间为 2 分 29 秒, 比 LSTM 和 CNN 分别高 26 秒和 35 秒.

第 1 部分 CNN 与 LSTM 模型的实验数据中, 在 MR、SST-2、SST-5 数据集上, LSTM 模型效果更好. 只有 TREC 数据集上, CNN 模型比 LSTM 模型分类结果高 2.38%. 这是由于 TREC 是问句类型数据集, 类型标签是人物、地点等类别, 相比 LSTM 获取时序依赖特征, CNN 获取的局部词组特征对于分类更有效.

在第 2 部分中, 实验结果显示, 与第 1 部分单层 CNN 与 LSTM 模型相比, LSTM+CNN 模型分类效果明显提高, 但是 CNN+LSTM 的效果反而降低了. 分析 LSTM 和 CNN 的获取特征的原理发现, CNN 旨在通过卷积获取局部特征, 将其结果作为 LSTM 的输入, 有可能影响 LSTM 按照文本序列来获取长依赖特征. 而若先通过 LSTM 模型处理文本, 优化原序列文本的特征表达, 之后再经 CNN 卷积得到局部特征, 这样不影响 LSTM 层和 CNN 层各自发挥其优势, 此时模型更合理, 因而能够有效地提高分类效果.

通过前面的分析, 了解到 LSTM 能根据文本序列增强特征表达能力, 为了获取更复杂更有效的特征, 本文增加了长短时记忆网络层数, 构建了 2LSTM+CNN 与 3LSTM+CNN 模型. 但是随着长短时记忆网络层数的增加, 由于深层网络反向传播梯度消失的原因, 分类效果反而明显下降.

借鉴残差网络的理论, 本文提出的 ResLCNN 模型, 改进了 3LSTM+CNN 模型, 在第 1 层 LSTM 层输出与第 4 层 CNN 层输入之间增加恒等映射, 构建了残差层. MR、SST-2、SST-5 以及 TREC 数据集上的实验结果表明, 与 3LSTM+CNN 模型相比, 分类准确率分别升高 5.36%、5.10%、12.65%、3.54%, 残差层的效果显著. 并且 ResLCNN 模型与 CNN 模型相比, 分别提高 MR、SST-2、SST-5 数据集 1.98%、1.78%、2.45% 的准确率, 与 LSTM 模型相比, 各数据集上结果都有提升. 而且它还比 LSTM+CNN 模型在 MR、SST-2 和 SST-5 数据集上高

1.0%、0.5%、0.47%。这表明深层网络模型能够得到更好的拟合函数,从而更有效地表达数据特征,提升分类效果。

### 3.4 句子长度分析

对不同长度句子的分类效果进行分析。选取的数据是 TREC 和 SST-2 数据集中的测试集,根据长度将句子分为 7 类,如图 2 所示的柱状图,左侧纵轴表示句子数量。

将 LSTM、CNN 以及本文的 ResLCNN 3 种模型在 TREC 和 SST-2 测试集上进行实验,统计 3 种模型在不同句子长度类别上的分类结果(图 2),右侧纵轴表示分类准确率。3 种模型的基本参数都相同。

从图 2 折线图中可以看出,随着句子长度的增加,模型分类准备率呈下降趋势。CNN 模型在句子长度大于 20 后的类别中,分类准确率持续下降。在句子长度大于 30 后,CNN 模型分类效果低于 LSTM 模型。这也印证了 LSTM 模型获取时序长依赖特征的能力。在长度不超过 30 的句子中,ResLCNN 模型在明显高于 LSTM 和 CNN 模型,表现较好。而在大于 30 的长句中,表现虽然比 CNN 模型好,但不如 LSTM 模型。分析其原因,发现 ResLCNN 模型中的 CNN 层在下层,梯度传播不影响,但对于上层的 LSTM 模型,受到反向传播过程中的梯度消失问题影响,获取长时序依赖能力减弱。

总体来说,对不同长度的句子,ResLCNN 模型整体分类效果高于单层 LSTM 与 CNN 模型。这说明,结合卷积神经网络与长短时记忆网络各自特性,能够提升对不同长度句子的分类效果。

## 4 总结

为探究短文本分类中 LSTM 与 CNN 组合模型的性能以及深层模型分类能力,本文提出了文本深度学习分类 ResLCNN 模型,包含 3 层 LSTM 和 CNN 层,并通过残差层缓解深层网络反向传播梯度消失问题。该模型可以有效结合 LSTM 获取时序依赖特征的能力和 CNN 获取局部特征的优势,获取更复杂的、更有效的特征。在多种数据集上的对比实验结果显示,LSTM+CNN 模型比单层 LSTM、CNN 模型分类效果好,而 ResLCNN 深层模型比 LSTM+CNN 在 MR、SST-2 和 SST-5 数据集上分别提高了 1.0%、0.5%、0.47% 的准确率,具有更好的分类效果。

### References:

- [1] Huang FL, Yu G, Zhang JL, Li CX, Yuan CA, Lu JL. Mining topic sentiment in micro-blogging based on microblogger social relation. Ruan Jian Xue Bao/Journal of Software, 2017,28(3):694-707 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5157.htm> [doi: 10.13328/j.cnki.jos.005157]
- [2] He YX, Sun ST, Niu FF, Li F. A deep learning model enhanced with emotion semantics for microblog sentiment analysis. Chinese Journal of Computers, 2017,40(4):773-790 (in Chinese with English abstract).
- [3] Peng Y, Wan CX, Jiang TJ, Liu DX, Liu XP, Liao GQ. Extracting product aspects and user opinions based on semantic constrained LDA model. Ruan Jian Xue Bao/Journal of Software, 2017,28(3):676-693 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5154.htm> [doi: 10.13328/j.cnki.jos.005154]
- [4] Hinton GE. Learning distributed representations of concepts. In: Proc. of the 8th Annual Conf. of the Cognitive Science Society. 1986. 1-12.

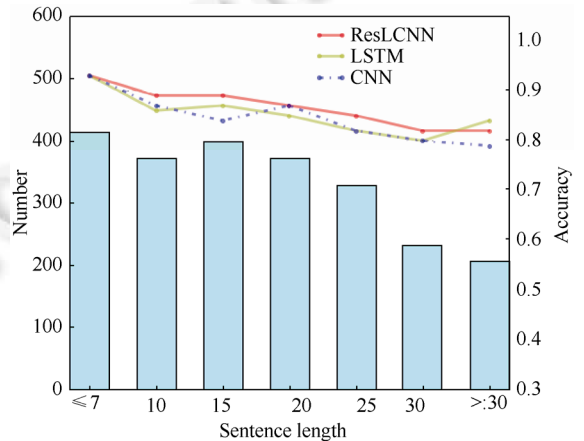


Fig.2 The effect of models on sentence classification with different length

图 2 模型对不同长度句子分类效果

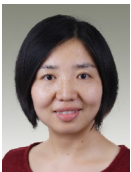
- [5] Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781, 2013.
- [6] Gu JX, Wang ZH, Kuen J, Ma LY, Shahroudy A, Shuai B, Liu T, Wang XX, Wang L, Wang G, Cai JF, Chen T. Recent advances in convolutional neural networks. arXiv preprint arXiv:1512.07108, 2015.
- [7] Donahue J, Hendricks LA, Guadarrama S, Rohrbach M, Venugopalan S, Darrell T, Saenko K. Long-Term recurrent convolutional networks for visual recognition and description. In: Proc. of the 2015 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). Washington: IEEE Computer Society, 2015. 2625–2634.
- [8] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proc. of the 2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). Washington: IEEE Computer Society, 2016. 770–778.
- [9] Pennington J, Socher R, Manning C. Glove: Global vectors for word representation. In: Moschitti A, *et al.*, eds. Proc. of the 2014 Conf. on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg: Association for Computational Linguistics, 2014. 1532–1543.
- [10] Cun YL, Boser B, Denker JS, Henderson D, Howard RE, Hubbard WE, Jackel LD. Handwritten digit recognition with a back-propagation network. *Advances in Neural Information Processing Systems*, 1990,2(2):396–404.
- [11] Collobert R, Weston J, Bottou L, Karlen M, Kavukcuoglu K, Kuksa P. Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, 2011,12:2493–2537.
- [12] Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences. In: Proc. of the 52nd Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2014. 655–665.
- [13] Kim Y. Convolutional neural networks for sentence classification. In: Moschitti A, *et al.*, eds. Proc. of the 2014 Conf. on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg: Association for Computational Linguistics, 2014. 1746–1751.
- [14] Wang P, Xu JX, Xu B, Liu CL, Zhang H, Wang FY, Hao HW. Semantic clustering and convolutional neural network for short text categorization. In: Proc. of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Int'l Joint Conf. on Natural Language Processing of the Asian Federation of Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2015. 352–357.
- [15] Ma M, Huang L, Zhou B, Xiang B. Dependency-Based convolutional neural networks for sentence embedding. In: Proc. of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Int'l Joint Conf. on Natural Language Processing of the Asian Federation of Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2015. 174–179.
- [16] Mou L, Peng H, Li G, Xu Y, Zhang L, Jin Z. Discriminative neural sentence modeling by tree-based convolution. In: Márquez L, *et al.*, eds. Proc. of the 2015 Conf. on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg: Association for Computational Linguistics, 2015. 2315–2325.
- [17] Mou L, Li G, Zhang L, Wang T, Jin Z. Convolutional neural networks over tree structures for programming language processing. In: Schuurmans D, *et al.*, eds. Proc. of the 13th AAAI Conf. on Artificial Intelligence. New York: AAAI Press, 2016. 1287–1293.
- [18] Yin W, Schütze H. Multichannel variable-size convolution for sentence classification. In: Alishahi A, *et al.*, eds. Proc. of the 19th Conf. on Computational Natural Language Learning. Stroudsburg: Association for Computational Linguistics, 2015. 204–214.
- [19] Kim J, Rousseau F, Vazirgiannis M. Convolutional sentence kernel from word embeddings for short text categorization. In: Márquez L, *et al.*, eds. Proc. of the 2015 Conf. on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg: Association for Computational Linguistics, 2015. 775–780.
- [20] Irsoy O, Cardie C. Deep recursive neural networks for compositionality in language. *Advances in Neural Information Processing Systems*, 2014(3):2096–2104.
- [21] Graves A, Jaitly N, Mohamed AR. Hybrid speech recognition with deep bidirectional LSTM. In: Proc. of the 2013 IEEE Workshop on Automatic Speech Recognition and Understanding. Piscataway: IEEE, 2013. 273–278.
- [22] Tai KS, Socher R, Manning CD. Improved semantic representations from tree-structured long short-term memory networks. In: Proc. of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Int'l Joint Conf. on Natural Language Processing of the Asian Federation of Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2015. 1556–1566.



- [23] Xu K, Ba J, Kiros R, Cho K, Courville AC, Salakhutdinov R, Zemel RS, Bengio Y. Show, attend and tell: Neural image caption generation with visual attention. In: Bach FR, *et al.*, eds. Proc. of the 32nd Int'l Conf. on Machine Learning (ICML). Cambridge: JMLR, 2015. 2048–2057.
- [24] Sainath TN, Vinyals O, Senior AW, Sak H. Convolutional, long short-term memory, fully connected deep neural networks. In: Proc. of the 2015 IEEE Int'l Conf. on Acoustics, Speech and Signal Processing. Piscataway: IEEE, 2015. 4580–4584.
- [25] Chiu JPC, Nichols E. Named entity recognition with bidirectional LSTM-CNNs. In: Nagata M, ed. Proc. of the Trans. of the Association for Computational Linguistics. 2016. 357–370.
- [26] Kim Y, Jernite Y, Sontag D, Rush AM. Character-Aware neural language models. In: Schuurmans D, *et al.*, eds. Proc. of the 13th AAAI Conf. on Artificial Intelligence. New York: AAAI Press, 2016. 2741–2749
- [27] Zhou C, Sun C, Liu Z, Lau FCM. A C-LSTM neural network for text classification. arXiv preprint arXiv:1511.08630, 2015.
- [28] Zhang R, Lee H, Radev D. Dependency sensitive convolutional neural networks for modeling sentences and documents. In: Knight K, *et al.*, eds. Proc. of the 2016 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg: Association for Computational Linguistics, 2016. 1512–1521.
- [29] Srivastava RK, Greff K, Schmidhuber J. Training very deep networks. arXiv preprint arXiv:1507.06228, 2015.
- [30] He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. arXiv preprint arXiv:1603.05027, 2016.
- [31] Bo P, Lee L. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In: Knight K, *et al.*, eds. Proc. of the 43rd Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2005. 115–124.
- [32] Socher R, Perelygin A, Wu J, Chuang J, Manning C, Ng A, Potts C. Recursive deep models for semantic compositionality over a sentiment treebank. In: Proc. of the Conf. on Empirical Methods in Natural Language Processing (EMNLP 2013). 2013. 1631–1642.
- [33] Li X, Roth D. Learning question classifiers. In: Proc. of the 19th Int'l Conf. on Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2002. 1–7.
- [34] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Proc. of the 25th Int'l Conf. on Neural Information Processing Systems. Curran Associates Inc., 2012. 1097–1105.
- [35] Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov R. Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580, 2012.

#### 附中文参考文献:

- [1] 黄发良,于戈,张继连,李超雄,元昌安,卢景丽.基于社交关系的微博主题情感挖掘.软件学报,2017,28(3):694–707. <http://www.jos.org.cn/1000-9825/5157.htm> [doi: 10.13328/j.cnki.jos.005157]
- [2] 何炎祥,孙松涛,牛菲菲,李飞.用于微博情感分析的一种情感语义增强的深度学习模型.计算机学报,2017,40(4):773–790.
- [3] 彭云,万常选,江腾蛟,刘德喜,刘喜平,廖国琼.基于语义约束 LDA 的商品特征和情感词提取.软件学报,2017,28(3):676–693. <http://www.jos.org.cn/1000-9825/5154.htm> [doi: 10.13328/j.cnki.jos.005154]



王 俊 丽 (1978—), 女, 山西祁县人, 博士, 副研究员, 主要研究领域为深度学习, 网络数据分析, 语义网络。



王 小 敏 (1991—), 男, 硕士, 主要研究领域为深度学习。



杨 亚 星 (1992—), 女, 硕士, 主要研究领域为深度学习。