**基于卷积神经网络和词语邻近特征的情感分类模型[[1]](#footnote-1)**

计 算 机 工 程

Computer Engineering

**··**

**文章编号：1000—3428(201 )01—00—0**

**文献标识码：A**

**中图分类号：**

吕 超，杨 超，李仁发

（湖南大学信息科学与工程学院，长沙 410082）

**摘要：**目前，基于卷积神经网络（CNN）的方法在情感分类任务中已经取得了不错的效果，此类方法使用词向量作为网络的输入，但是在卷积过程中每个词向量只能表征单个单词，并不蕴含上下文信息，这不利于信息传递的连续性，并且卷积操作在局部范围内可能会打乱词向量的序列性。针对这个问题，本文提出一种基于词语邻近特征（Adjacent Feature）的卷积神经网络模型，在卷积过程中让每个词向量携带邻近词语的特征，这样既保证信息传递的连续性也保证了词向量在局部范围内的序列性。实验结果表明，在COAE2014（2分类）、COAE2015（3分类）的情感分类任务上的准确率分别达到了89.43%和85.61%，说明本文提出的方法确实可行、有效。

**关键词：**情感分类；卷积神经网络；词向量；连续性；序列性；邻近特征

**Sentiment Classification Model Based On Convolutional Neural Network And Word Adjacent Features**

Lv Chao, Yang Chao, Li Renfa

（School of Information Science and Engineering, Hunan University, Changsha 410082, China）

**【Abstract】**At present, the methods based on Convolutional Neural Network(CNN) has achieved good performance in the task of sentiment classification. This method mainly uses word vectors as the input of the network, but in the convolutional process, the word vector can only characterize a single word, but does not contain contextual information, this is not conducive to the continuity of information transmission, and convolution may disrupt the sequence of word vectors in the local context. Aiming at this problem, this paper proposes a convolutional neural network model based on word adjacent features, which allows each word vector to carry the characteristics of its adjacent words in the convolution process, so that we can both ensure the continuity of information transmission and the sequence of word vectors in the local scope. The experimental result shows that the prediction accuracy rate on the COAE2014 sentiment classification task (2 classification) and COAE2015 sentiment classification task(3 classification) is 89.43% and 85.61% respectively, which indicates that the proposal model is actually feasible and effective.

**【Key words**】sentiment classification; convolutional neural network; word vector; continuity; sequentiality, adjacent feature

# 1 概述

随着微博、微信等社交媒体的兴起，网络上带有评论性的文本资源的数量与日俱增。面对这些大量非结构化文本，迫切需要通过技术手段对文本的情感倾向进行分类判断，这将有助于达到多方面的目的，例如获取大众对于有关政策的意见、相关产品的满意程度和各种社会热点事件的看法等[1]。但是微博文本篇幅较短、存在错别字和语法错误、充满了大量网络词汇和口语词汇，这些问题都给情感分类任务提出了新的挑战，引发了国内外学者极大的研究兴趣[2]。

目前，基于卷积神经网络的情感分类方法得到了广泛关注。卷积神经网络通过多个并行的卷积操作获取文本的局部特征[3]，经过池化层再将筛选出的局部特征拼接然后传递给全连接层进行分类。然而大多数的卷积神经网络模型在卷积操作和特征拼接环节中，很大程度上忽略了文本的有序性即词语之间的序列性，因此它不能保证信息传递的连续性。例如，“我/喜欢/自然/语言/处理/但是/它的/学习/门槛/很高”，假设卷积核的尺寸是5，则将被卷积的分片依次是“我/喜欢/自然/语言/处理”、“喜欢/自然/语言/处理/但是”、“自然/语言/处理/但是/它的”、“语言/处理/但是/它的/学习”、“处理/但是/它的/学习/门槛”、“但是/它的/学习/门槛/很高”，那么得到的局部信息可能是“喜欢”、“自然”、“语言”、“但是”、“门槛”、“很高”，也可能是“自然”、“语言”、“它的”、“处理”、“学习”、“很高”，前一种明显表达的是负面情感，而后一种则很可能被误判为正面情感，因为“很高”带有积极情感。现实生活中，人们也常使用一些脱离语境的极性词来评估情感，比如“大”、“小”、“高”、“低”等，然而这些词在不同的语境中有着不同的含义，因此我们无法仅凭某些极性词就能分辨出整个文本的情感倾向，而必须要考虑语境。为了解决上述问题，本文提出一种基于词语邻近特征的卷积神经网络模型，即在卷积过程中考虑每个词语的前后邻近词的特征，这样可保证即可保证信息传递的连续性。本文分别在COAE2014（2分类）、COAE2015（3分类）任务4的数据集上进行了实验，准确率分别达到了89.43%、85.61%，并且优于未考虑邻近特征的基准神经网络模型。

本文的组织结构如下：第二节介绍情感分类的相关工作；第三节介绍本文提出的情感分类模型；第四节通过对比实验证明本文方法的有效性，并对实验结果进行讨论；第五节对本文的主要贡献做出梳理和总结。

# 2 相关工作

近年来，深度学习（Deep Learning）模型在计算机视觉[4]和语音识别[5]邻域取得了巨大成果。深度学习模型通常是具有多层结构的神经网络，强调从大规模的数据集学习各种现实事物的特征表示， 并且这些特征能够被计算机直接应用于各种计算模型中。当前在自然语言处理领域，大多数深度学习方法通过神经网络模型学习单词的词向量表达[6]，然后在此基础上再做进一步处理，例如语义分析[7]，搜索词检索[8]，句子分类[9]等等。目前，主流的神经网络模型是递归神经网络（如RNN、LSTM）和卷积神经网络（如CNN、DCNN）。递归神经网络将文本看作是有序的词语序列，并且考虑了词语间的关联性，可以充分利用上下文信息从而更好地进行语言建模。RNTN模型[10]利用情感树库在二元化的句法树结构上进行语义合成，文献[11]在此模型基础上，通过对依存句法树进行变换，完成句中指定评价对象的情感极性分析。文献[12]通过添加语义合成函数的pooling层，使得 RNTN可以自适应地选择最适合当前节点的语义合成函数，从而大幅提高情感分类效果。但是这些类RNN模型仅考虑了文本的有序性而忽略了词语在语义上的结构性，并且该类模型在训练过程中存在着梯度爆炸和梯度消失的问题[13]，训练过程难以收敛。由于微博文本的不规范性，各种句法分析结果通常不准确，因而CNN模型成为微博情感分类的首选模型。文献[9]提出一种DCNN模型，通过学习句子结构来提高情感语义合成性能，文献[14]提出CharSCNN模型，为英语单词增加了字符表示层的特征，可以提高对于英文Twitter中不规则单词的表示能力。文献[15]首次尝试将CNN应用到句子分类（Sentence Classification）任务中，通过词向量训练工具word2vector生成的词向量训练一个简单的3层卷积神经网络，在二分类和多分类数据集上均取得了很好的分类性能，说明预先训练的词向量是一类通用的特征向量，可以在不同的分类任务中直接使用。另外文献[15]中采用的MCNN是CNN的一种，能够对多个窗口大小内的词向量执行语义合成操作，在多个不同的情感数据集上也取得了很好的分类性能。文献[16]基于MCNN，提出一种称为EMCNN的模型，将基于表情符号的情感空间映射与深度学习模型MCNN结合，有效增强了MCNN捕捉情感语义的能力。文献[17]提出一种LSTM-CNN模型，首次尝试从子句的角度来分析文本的情感极性，利用LSTM（长短时记忆网络）归纳某个子句的上下文信息，然后将上下文信息拼接在子句的词向量表达形式的两侧，进而再使用CNN对子句进行特征提取、分类，该模型在影评数据集上取得了比传统分类方法更好的性能。

上述模型都是基于传统CNN模型的变体，它们在做卷积操作时依然无法保证词向量之间的序列性以及信息传递的连续性。本文在参考了文献[15]的工作后，分别提出三种基于词语邻近特征的卷积神经网络模型LAF-CNN、RAF-CNN、LRAF-CNN，这些模型在卷积过程中让每个词向量携带邻近词语的特征，保证了信息传递的连续性，并且在局部范围内确保词语的序列性不会被打乱，在一定程度上具备了RNN模型的归纳能力。相比传统的CNN模型，本文提出的LAF-CNN、RAF-CNN、LRAF-CNN在卷积过程中兼顾了词语的上下文特征和词语的序列性，提高了深层语义情感分析的效果。

# 3 基于词语邻近特征的卷积神经网络模型

## 3.1 词语的邻近特征

词语的邻近特征是指与其相邻的词语的特征。假设原始句子为：

 （1）

对应的词向量矩阵为：

 （2）

其中，表示组成句子的各个词语，表示对应的词向量特征。那么在卷积过程中，对于某个词语我们采取3种不同的邻近特征附加策略：附加左邻近特征（Left Adjacent Feature, LAF）、附加右邻近特征（Right Adjacent Feature, RAF）、附加左右邻近特征（Left Right Adjacent Feature, LRAF），它们的作用都是为了增强词语在特定上下文中所表达的语义。如果使用LAF，则卷积过程中使用词向量，即从第一个词向量开始，每一个词向量都会与其左侧的词向量特征进行合并操作，这样做的好处是前一个词向量的信息可以传递给下一个词向量，因此下一个词向量在参加卷积操作时可以提供相对充足的情感信息供卷积层提取到有效的特征。中每一个词向量与其右侧的词向量特征合并，中每一个词向量与其左、右侧的词向量特征合并。这三种策略的直观描述如图1所示，其中表示零向量。



**图1 三种词语邻近特征附加策略**

## 3.2 基于词语邻近特征的卷积神经网络模型

本文提出的基于词语邻近特征的卷积神经网络模型如图2所示。图2（这里使用的是LRAF附加策略）中使用的卷积核有三个尺寸：2、3、4，每个尺寸的卷积核有2种，每个卷积核对句子矩阵进行卷积操作并生成长度可变的特征图，然后对每个特征图进行最大池化（max-pooling）操作，即取出每个特征图中的最大值，由这些单值生成一个固定大小的特征向量，并且这个特征向量以全连接的形式传给Softmax层，Softmax层使用该特征向量作为输入并且计算出各个类别上的概率分布（这里我们假设是二值分类，因此图2中只有2个分类状态）。

我们假设输入层句子**的长度为，表示维的词向量，对应句子**中的第个词语，那么句子**就可以表示成：

**图2 基于词语邻近特征的卷积神经网络模型结构**

 （3）

其中表示串联操作，换句话说，表示词向量的串联。如果使用邻近特征附加策略，则句子会有另外3种表示形式（图2中使用的是）：

 （4）

 （5）

 （6）

其中，运算符为两个向量的合并符号，例如向量，向量，那么有。

卷积层利用大小为的滤波器对输入特征矩阵进行卷积操作，即：

 （7）

其中，代表特征图中第个特征值，为卷积核函数，为滤波器，为滑动窗口大小，为偏置值。表示由第行到第行组成的局部特征矩阵。因此，特征图为：

 （8）

在卷积层之后，加入了ReLU层，将ReLU作为激活函数，可以加速随机梯度下降的收敛速度[16]。

下采样层使用文献[18]提出的max-pooling方法进行特征采样，得到的特征值为：

 （9）

卷积层和下采样层共同构成LRAF-CNN的特征提取层。LRAF-CNN由多个不同尺寸的特征提取层（有多个值）并列组成，每种尺寸的特征提取器各有个，因此全连接层的特征向量为：

 （10）

其中，表示第种类型的滤波器产生的第个特征值。相比普通的CNN模型，本文提出的这种网络结构可以有效的提取出与正负面情感标签相关的词向量特征，并用于最终的情感分类。

下采样层输出的多个单值特征作为全连接层的输入，全连接层将这些单值特征进行拼接，形成特征向量，然后利用Softmax多分类函数输出各个分类的概率大小，从而确定分类结果，并根据训练数据的实际分类标签，采用反向传播算法对模型参数进行梯度更新，即：

 （11）

其中，表示句子的实际情感标签，表示模型的参数集合（即神经元之间的权重矩阵，训练过程中会不断变化，直到收敛）。通过该式，可以计算出给定句子的情感标签和模型参数集合的条件概率分布图，进而得出分类结果。

**3.3 正则化**

我们在网络的倒数第二层引入Dropout策略并使用L2-constraint规则权重约束条件对网络参数进行正则化处理[19]。Dropout通过在前向传播过程中随机丢弃一定比例的神经元（本文的Dropout比例等于0.5）以阻止神经元之间的相互适应。对于倒数第二层的输入，我们不再使用下式计算输出神经元：

 （12）

取而代之的是Dropout策略：

 （13）

其中，为向量内积符号，是遵从Bernoulli随机分布的“掩饰向量”（masking vector），向量中元素值为1的概率是，为0的概率是。梯度只在未进行掩饰处理的神经元中反向传播。

# 4 实验

## 4.1 数据集

本文使用COAE2014、COAE2015的任务4的微博数据集[[2]](#footnote-2)。中文观点倾向性分析评测(Chinese Opinion Analysis Evaluation, COAE)始办于2008年，是国内第一个情感分析方面的权威评测会议。COAE2014任务４（2分类）的微博数据集共包含40000条数据，其中官方公布了7000条微博的极性（负面有3224条、正面有3776条），另33000条是混淆语料，语料涵盖手机、翡翠、保险三个领域。COAE2015任务４（3分类）的微博数据集共包含133201条数据，其中官方公布了20154条微博的极性（负面有6209条、中性有1639条、正面有12306条），另113047条是混淆语料，语料涵盖汽车、电子、手机、美食、娱乐、宾馆等多个领域。两个数据集的概要信息如表1所示。

**表1 COAE2014、COAE2015的数据集概要**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Data** | ***c*** | ***l*** | ***N*** | ***V*** | ***Vpre*** | ***CV*** |
| COAE2014  COAE2015 | 2  3 | 100  102 | 7000  20154 | 13993  19389 | 13451  18384 | 10  10 |

其中，***c***表示目标数据集是几分类，***l***表示该数据集中最长的句子包含多少个词语，***N***表示数据集中有多少条标注数据，***V***表示标注数据的词汇量大小，***Vpre***表示标注数据含有多少词汇已存在预训练词向量中，***CV***等于10表示采用十折交叉验证法划分实验数据。

## 4.2 数据预处理和预训练词向量

在数据预处理方面，本文直接去除了文本中所有标点符号和表情符号，只留下含义较多的中、英文，然后利用ICTCLAS[[3]](#footnote-3)分词工具对实验数据集进行分词。在预训练词向量方面，本文使用新浪微博提供的数据抽取接口抽取了1000万条微博数据，经过去燥和预处理之后，再使用Google开源的词向量训练工具word2vec[[4]](#footnote-4)训练词向量。经过训练后，得到的词向量文件包含80552个词语，在COAE2014、COAE2015任务4的数据集上的词汇覆盖率分别是96.13%、95.33%。

## 4.3 模型对比实验

**CNN：**这是基准模型，使用预训练词向量对所有词语进行初始化，不在预训练词向量中的词语随机初始化（随机初始化范围是[-0.25, 0.25]），并且在训练过程中所有词向量保持不变。

**LAF-CNN：**词向量的初始化方案与CNN相同；训练过程中为每个词语添加左邻近特征，如果左邻近词语不在预训练词向量中，则随机初始化左邻近特征，且在训练过程中所有词向量保持不变。

**RAF-CNN：**词向量的初始化方案与CNN相同；训练过程中，为每个词语添加右邻近特征，如果右邻近词语不在预训练词向量中，则随机初始化右邻近特征，且在训练过程中所有词向量保持不变。

**LRAF-CNN：**词向量的初始化方案与CNN相同；训练过程中，为每个词语添加左、右邻近特征，如果左、右邻近词语不在预训练词向量中，则随机初始化左、右邻近特征，且在训练过程中所有词向量保持不变。

上述四个模型除了邻近特征的附加策略不同之外，超参数设置均相同，超参数如表2所示。模型在训练过程中，采用Zeiler[20]提出的Adadelta Update Rule规则进行随机梯度下降更新参数。

**表2 模型的可调超参数**

|  |  |
| --- | --- |
| **可调参数** | **值** |
| 卷积核函数 | ReLU |
| 词向量维度 | 300 |
| 批处理大小 | 64 |
| 滤波器尺寸 | 3, 4, 5 |
| 每个尺寸的滤波器个数 | 128 |
| 参数更新比例 | 0.5 |
| 权重惩罚系数 | 1.0 |
| 迭代次数 | 200 |

## 4.4 实验结果与讨论

模型对比实验在COAE2014数据集（2分类）和COAE2015数据集（3分类）上的实验结果如表3和表4所示。实验效果评价指标为：查准率、召回率、F值、准确率。本文的基准模型是未考虑任何邻近特征的卷积神经网络。由表4可知，相比基准模型CNN，LAF-CNN、RAF-CNN、LRAF-CNN在查准率、召回率、F1值、准确率方面均有提升，其中准确率分别提高了1.14%、0.57%、3%。由表5可知，相比基准模型CNN，LAF-CNN、RAF-CNN、LRAF-CNN在查准率、召回率、F1值方面浮动比较大，这是因为各个分类的训练集数量不均衡所导致的，但在准确率方面依然有一定提升，分别提高了1.63%、1.09%、1.83%。

## 4.4.1 CNN vs. LAF-CNN、RAF-CNN、LRAF-CNN

从两个数据集上的实验结果来看，当加入邻近特征后，模型在综合性能确实得到了一定提升，这说明在卷积过程中给每个词语添加邻近特征后，一方面可以强化词语在特定上下文中所表达的语义，另一方面相邻词语之间相互关联的关系提高了信息传递的连续性。这两点在一定程度上可确保在卷积过程中即使漏掉了个别情感词语，也不会对最终的分类结果产生多大影响。

**表3 模型对比实验在COAE2014上的查准率、召回率、F1值、准确率**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 模型 | 负面 | | | 正面 | | | 准确率 |
| 查准率 | 召回率 | F1值 | 查准率 | 召回率 | F1值 |
| CNN | 87.30% | 82.72% | 84.94% | 85.75% | 89.63% | 87.65% | 86.43% |
| LAF-CNN | 87.38% | 85.49% | 86.43% | 87.73% | 89.36% | 88.54% | 87.57% |
| RAF-CNN | 88.20% | 83.02% | 85.53% | 86.08% | 90.43% | 88.20% | 87.00% |
| LRAF-CNN | 90.97% | 85.98% | 88.40% | 88.21% | 92.47% | 90.29% | **89.43%** |

**表4 模型对比实验在COAE2015上的查准率、召回率、F1值、准确率**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 模型 | 负面 | | | 中性 | | | 正面 | | | 准确率 |
| 查准率 | 召回率 | F1值 | 查准率 | 召回率 | F1值 | 查准率 | 召回率 | F1值 |
| CNN | 83.02% | 80.13% | 81.55% | 56.00% | 23.46% | 33.07% | 85.64% | 94.32% | 89.77% | 83.78% |
| LAF-CNN | 80.19% | 88.67% | 84.22% | 55.91% | 30.95% | 39.85% | 90.83% | 91.29% | 91.06% | 85.41% |
| RAF-CNN | 76.29% | 88.70% | 82.03% | 64.00% | 9.82% | 17.02% | 89.71% | 92.75% | 91.20% | 84.87% |
| LRAF-CNN | 84.52% | 83.72% | 84.12% | 50.00% | 21.23% | 29.81% | 87.80% | 94.23% | 90.90% | **85.61%** |

## 4.4.2 LAF-CNN vs. RAF-CNN

从两个数据集上的实验结果来看，LAF-CNN的分类准确率均略高于RAF-CNN，这可能跟词语的序列性有关。众所周知，中文文本语义的表达是从左至右、先有上文后有下文，而LAF-CNN恰巧具有这种承上启下的作用。具体是否因为这个原因，还有待进一步验证。

## 4.4.3 LAF-CNN、RAF-CNN vs. LRAF-CNN

从两个数据集上的实验结果来看，LAF-CNN、RAF-CNN的分类准确率均略低于 LRAF-CNN，这跟词语在特定上下文中所蕴含语义有关，由于LAF-CNN只携带了左邻近特征、RAF-CNN只携带了右邻近特征，而LRAF-CNN兼顾左右邻近特征，可以提取到更精确的语义信息，因此效果更好。

# 5 结语

本文为了在卷积过程中增强词语在当前上下文中的语义和保证词语在局部范围内的序列性，提出了一种基于词语邻近特征的卷积神经网络模型，并且使用提出了三种词语邻近特征附加策略。实验结果表明，本文的方法确实可行、有效。

本文在附加邻近特征时，只考虑了当前词语前、后邻近词的特征，如果在计算能力允许的情况下，是否可以扩大邻近词的数量从而获取更精确的上下文的信息，这是本文今后的一个研究方向。另外，本文在预处理语料时，直接去掉了所有标点符号和表情符号，然而例如问号（“？”）、感叹号（“！”）、开心（“^\_^”）、沮丧（“>\_<”）等标点符号和表情符号必然会影响上下文的情感极性转移，这也无疑对模型最后的分类效果产生一定的负面影响，因此考虑由标点符号和表情符号而引起的情感极性转移问题是本文今后的另一个研究方向。

# 参考文献

[1] B Pang, L Lee. Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales[C]// Proceedings of the ACL. Ann, USA: ACM Press, 2005: 115-124.

[2] 韦航, 王永恒. 基于主题的中文微博情感分析[J]. 计算机工程, 2015: 238-244.

[3] LeCun, L. Bottou, Y. Bengio, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[4] Krizhevsky A, Sutskever I, Hinton G. Imagenet classification with deep convolutional neural networks[C]//Proceedings of the NIPS. Lake Tahoe, USA: NIPS Press, 2012: 1097-1105.

[5] Graves, Mohamed A, Hinton G. Speech recognition with deep recurrent neural networks[C]//Proceedings of the ICASSP. Vancouver, Canada: IEEE Press, 2013: 6645-6649.

[6] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[C]//Proceedings of the NIPS. Lake Tahoe, USA: NIPS Press, 2013: 3111-3119.

[7] WT Yih，X He，C Meek. Semantic parsing for single-relation question answering[C]//Proceedings of the ACL. Baltimore, USA: Citeseer Press, 2014: 643-648.

[8] Y Shen，X He，J Gao, et al. Learning Semantic Representations Using Convolutional Neural Networks for Web Search[C]//Proceedings of the WWW. New York, USA: ACM Press, 2014: 373-374.

[9] N Kalchbrenner，E Grefenstette，P Blunsom. A Convolutional Neural Network for Modelling Sentences[C]//Proceedings of the ACL. Baltimore‚ USA: arXiv Press, 2014: 1404-2188.

[10] Socher R, Perelygin A, Wu J Y, et al. Recursive deep models for semantic compositionality over a sentiment treebank[C]//Proceedings of the EMNLP. Seattle, USA: Citeseer Press, 2013: 1631-1642.

[11] Dong L, Wei F, Tan C, et al. Adaptive recursive neural network for target-dependent Twitter sentiment classification[C]//Proceedings of the ACL. Baltimore‚ USA: ACL Press, 2014: 49-54.

[12] Dong L, Wei F, Zhou M, et al. Adaptive multi-compositionality for recursive neural models with applications to sentiment analysis[C]//Proceedings of the AAAI. Quebec, Canada: AAAI Press, 2014: 1537-1543.

[13] J Kolen, S Kremer. Gradient flow in recurrent nets: the difficulty of learning long term dependencies[M]. Wiley-IEEE Press, 2001: 237-243.

[14] Dos Santos C N, Gatti M. Deep convolutional neural networks for sentiment analysis of short texts[C]//Proceedings of the COLING. Dublin, Ireland: COLING Press, 2014: 69-78.

[15] Kim Y. Convolutional neural networks for sentence classification[C]//Proceedings of the EMNLP. Doha, Qatar: arXiv Press, 2014: 1408-5882.

[16] 何炎祥, 孙松涛, 牛菲菲等. 用于微博情感分析的一种情感语义增强的深度学习模型[J]. 计算机学报, 2016, 39: 1-19.

[17] Peng Chen, Bing Xu, Muyun Yang, et al. Clause sentiment identification based on convolutional neural network with context embedding[C]//Proceedings of the ICNC-FSKD. Changsha, China: IEEE Press, 2016: 1532–1538.

[18] R Collobert, J Weston, L Bottou, et al. Natural language processing (almost) from scratch[J]. Journal of machine learning research, 2011, 12(1): 2493-2537.

[19] GE Hinton, N Srivastava, A Krizhevsky, et al. Improving neural networks by preventing co-adaptation of feature detectors[J]. Computer Science, 2012, 3(4): 212-223.

[20] MD Zeiler. ADADELTA: AN ADAPTIVE LEARNING RATE METHOD[J]. Computer Science, 2012, 3(4): 325-330.

1. **基金项目：**国家自然科学基金资助项目（No.61173036）；湖南省科技计划（No.2015GK3015）；国家高技术研究发展计划（No.2012AA01A301-01）。

   **作者简介：**吕 超（1987-），男，硕士研究生，主研方向是深度学习、自然语言处理；杨 超，副教授，博士，硕士生导师；李仁发，教授，博士，博士生导师。

   **收稿日期：** **修回日期：** **E-mail：**chao9038@hnu.edu.cn [↑](#footnote-ref-1)
2. http://115.24.12.5/web/index.html [↑](#footnote-ref-2)
3. http://ictclas.nlpir.org/ [↑](#footnote-ref-3)
4. https://code.google.com/archive/p/word2vec/ [↑](#footnote-ref-4)