

文章编号: 1003-0077 (2011) 00-0000-00

基于卷积神经网络的微博情感倾向性分析*

刘龙飞, 杨亮, 张绍武, 林鸿飞

(大连理工大学 信息检索实验室, 辽宁 大连 116024)

摘要: 微博情感倾向性分析旨在发现用户对热点事件的观点态度。由于微博噪声大、新词多、缩写频繁、有自己的固定搭配、上下文信息有限等原因, 微博情感倾向性分析是一项有挑战性的工作。本文主要探讨利用卷积神经网络进行微博情感倾向性分析的可行性, 分别将字级别词向量和词级别词向量作为原始特征, 采用卷积神经网络来发现任务中的特征, 在 COAE2014 任务 4 的语料上进行了实验。实验结果表明, 利用字级别词向量及词级别词向量的卷积神经网络分别取得了 95.42% 的准确率和 94.65% 的准确率。由此可见对于中文语料而言, 利用卷积神经网络进行微博情感倾向性分析是有效的, 且使用字级别的词向量作为原始特征会好于使用词级别的词向量作为原始特征。

关键词: 深度学习; 情感倾向性分析; 卷积神经网络; 词向量

中图分类号: TP391

文献标识码: A

Convolutional Neural Networks for Chinese Micro-blog Emotional Tendency Identification

LIU Longfei, YANG Liang, ZHANG Shaowu, LIN Hongfei

(Information Retrieval Laboratory, Dalian University of Technology, Dalian, Liaoning 116024,
China)

Abstract: Chinese micro-blog emotional tendency identification aims to discover the user attitude towards hot events. Emotional Tendency identification of Chinese micro-blog messages is challenging because of big noise, too many new words, abbreviations frequently, fixed collocation, the limited contextual information that they normally contain. This paper explores the feasibility of performing Chinese micro-blog sentiment analysis by convolutional neural networks. We try to avoid task-specific features, separately take character level embedding and word level embedding as input, and use convolutional neural networks to discover relevant features to the tasks. For the COAE 4th task corpus, the character level CNN achieves a sentiment prediction (in both binary positive/negative classification) accuracy of 95.42%, and the word level CNN achieves a sentiment prediction accuracy of 94.65%. The results show that the convolutional neural networks model is promising in Chinese micro-blog sentiment analysis. For chinese corpus, character level embedding is slightly better than word level embedding.

Key words: deep learning; emotional tendency identification; convolutional neural networks; word embedding

1 引言

随着社交网络的不断发展人们更愿意通过微博、博客社区来表达自己的观点发表对热点事件的评论从而使通过微博、博客、影评以及产品评价等来了解社交网络用户的情感倾向得到了学术界的广泛关注。根据微博内容进行情感倾向性分析是一个具有挑战性的任务, 近年来引发了学者极大的兴趣^[1]。

已有研究所采用的方法大多数都基于词袋模型, 而这种模型无法捕获到很多有关情感倾

* 收稿日期:

定稿日期:

基金项目: 国家自然科学基金资助项目 (编号: 61277370, 61402075)、国家 863 高科技计划资助项目 (编号: 2006AA01Z151)、辽宁省自然科学基金 (编号: 201202031, 2014020003)、教育部留学回国人员科研启动基金和高等学校博士学科点专项科研基金资助课题 (编号: 20090041110002), 中央高校基本科研业务费专项资金资助。

向性分析的语言现象特征。例如,“反法西斯联盟击溃了法西斯”和“法西斯击溃了反法西斯联盟”这两个词组拥有相同的词袋模型表示方法,而前一个带有积极的感情色彩,后一个带有消极的感情色彩。除此之外,还有很多研究者使用人工标注的数据(情感词典及句法分析等),虽然采用这些方法可以有效的提高情感分析的准确性但由于需要较多的人工标注数据从而限制了这些方法在其他领域以及跨语言的推广^[2]。卷积神经网络模型可以从大量未标注的文本中学习到先验知识,避免依赖于具体任务的人工特征设计,可以在一定程度上解决短文本上下文信息有限的问题。

要提取微博文本特征,目前主要做法是对微博进行分词,匹配情感词典,选用其中的情感词或者情感的相关得分作为特征,但是微博属于短文本范畴,噪声大、新词多、缩写频繁、有自己的固定搭配、上下文信息有限,对微博做分词歧义明显,往往得到的是不好的切分。比如:“我发现了一个高大上网站”,在该句中,“高大上网站”如果使用传统分词技术,会被切分为“高大/上/网站”或者“高大/上网/站”,这样的切分无法体现句子的正确语义,甚至后一种切分还将“网站”切分导致丢失评价对象^[3]。为了解决上述问题,本文引入字级别特征,将单个字作为输入特征,通过多个拥有不同大小卷积核的并行卷积层学习微博文本特征。

本文训练了一个包含多个并行卷积层的卷积神经网络,不同卷积层拥有大小不同的卷积核。本文分别将字级别特征和词级别特征作为原始特征进行了实验,利用字级别特征的卷积神经网络取得了 95.42% 的准确率,利用词级别特征的卷积神经网络取得了 94.65% 的准确率。实验表明,对于中文语料而言,利用卷积神经网络进行微博情感倾向性分析是有效的,且将字级别词向量作为原始特征会好于将词级别词向量作为原始特征。

本文的结构如下。第二章介绍了一些相关工作。第三章详细介绍了本文使用的卷积神经网络结构。第四章详细说明了本文的实验设定,介绍了实验结果,并对实验结果进行了详细的讨论。第五章是文章的总结。

2 相关工作

2.1 卷积神经网络

卷积神经网络利用卷积层可以学习局部特征^[4]。在自然语言处理中,CNN 模型在很多方面取得了很多非常好的成绩,例如语法解析^[5]、搜索词检索^[6]、句子建模^[7]及其他传统的自然语言工作^[8]。Cicero dos Santos 等人^[9]提出了 CharSCNN 模型,两个卷积层分别学习词语的构造特征和句子的语义特征。

本文与 Cicero dos Santos 工作的不同之处在于,本文分别利用字级别向量和词级别向量进行实验,而没有学习词语的构造特征。本文与 Cicero dos Santos 工作的另一个不同之处在于,Cicero dos Santos 在利用卷积层学习句子特征时,使用了大小单一的卷积核,而不同大小的卷积核获取的特征是不同的,本文参考了 Kim 等人的工作^[10],利用多个不同大小的卷积核学习句子级别的特征向量,然后对特征向量进行串接,构造新的句子特征向量。

2.2 情感分析

情感分析自从 2002 年由 Bo Pang 提出之后,获得了很大程度的关注,特别是在线评论的情感倾向性分析上获得了很大的发展。本文主要关注无监督的情感分析方法,由于不需要大量标注语料,无监督情感分析方法一直受到许多研究者的青睐,但同时效果也低于有监督的情感分析方法。Turney^[11]首次提出基于种子词(excellent, poor)的非监督学习方法,使用“excellent”和“poor”两个种子词与未知词在搜索网页中的互信息来计算未知词的情感极性,并用以计算真个文本的情感极性。后续的非监督情感分析方法大都是基于生成或已有的情感词典或者相关资源进行情感分析。罗毅^[3]等人通过构建二级情感词典,对不同级别情感词作不同增强,使用 N-gram 获取文本特征,利用构建的情感词典进行微博情感倾向性分析。情感词典的构建过程需要大量的人工标注,在跨领域应用方面有很大的限制。使用 N-gram 模型,当 N 较大时,会导致特征维度太大而难以训练。包含多个并行卷积层的卷积神经网络通过卷积和池化操作,既充分利用了 N-gram 的特征,又不会造成维度灾难。

3 卷积神经网络模型

3.1 句子级别的表示和评分

字级别特征：以单个字作为句子的基本组成单位，对单个字训练词向量。

词级别特征：利用分词器对句子进行分词，以词（包含长度为 1 的词）作为句子的基本组成单位，对单个词训练词向量。

以“中国足球加油！”为例，字级别的句子组成为“中+国+足+球+加+油+！”，词级别的句子组成为“中国+足球+加油+！”。

给定包含 N 个基本单位 $\{r_1, r_2, \dots, r_N\}$ 的句子 x ，本文接下来的工作是计算其句子级别的表示 r_x^{sent} 。字级别句子的基本单位是单个的字，词级别句子的基本单位是分词之后的词。在计算句子级别的特征时，会遇到两个主要的问题：不同句子的长度不同，重要的信息会出现在句子中的任意位置。利用卷积层对句子建立模型，计算句子级别的特征，可以解决上面提到的两个问题。通过卷积操作可以得到句子中每个基本单位（字或者词语）的局部特征，然后对得到的局部特征进行最大化操作，从而得到一个固定长度的句子特征向量。

在包含 N 个基本单位 $\{r_1, r_2, \dots, r_N\}$ 的句子 x 中，卷积层对每个大小为 k 的连续窗口进行矩阵向量操作。本文假定向量 $z_n \in \mathbb{R}^{dk}$ 是以句子中第 n 个基本单位为中心的前后各 $(k-1)/2$ 个基本单位的词向量的串接：

$$z_n = (r_{n-(k-1)/2}, \dots, r_{n+(k-1)/2})^T \quad (1)$$

卷积层计算计算句子特征向量 $r_x^{sent} \in \mathbb{R}^{cl_u}$ 的第 j 个元素的过程如下：

$$[r^{sent}]_j = \max_{1 \leq n \leq N} [Wz_n + b^1]_j \quad (2)$$

其中， $W \in \mathbb{R}^{cl_u \times dk}$ 是该卷积层的权重矩阵。如图 1 所示，用该权重矩阵计算给定句子中每个基本单位（字或词）的窗口大小为 k 的局部信息，对句子中所有基本单位的窗口取最大值，就抽取了一个在窗口大小为 k 的条件下长度为 cl_u 的句子特征向量。图 1 中窗口大小 k 为 3。

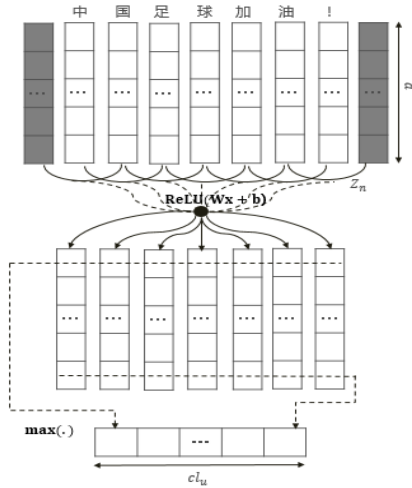


图 1 基于卷积方法抽取句子级别特征

Fig.1 Convolutional approach to sentence-level feature extraction

卷积窗口的大小 k 不同，获取的局部信息也不同。为此，本文利用并行的多个卷积层，学习不同 N-gram 的信息。每个卷积层经过最大池化操作后都会生成一个固定长度的句子特征向量，本文将所有卷积层生成的句子特征向量进行串接，得到一个新的句子特征向量^[9]。

包含多个不同窗口的并行卷积层的架构如图 2 所示。

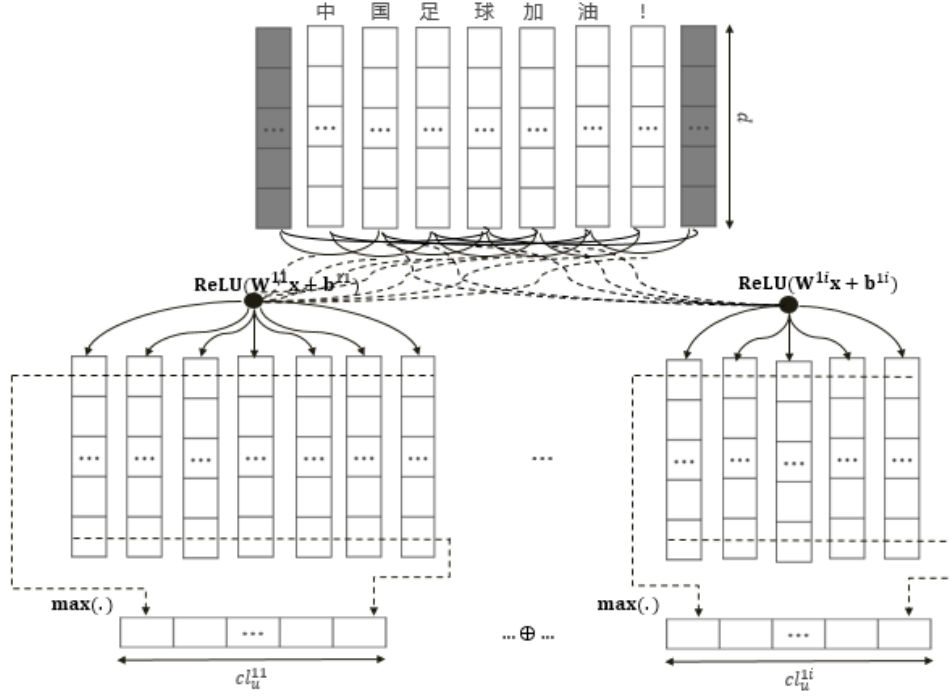


图 2 通过基于不同窗口大小的卷积方法利用多个不同大小的卷积核抽取句子级别特征

Fig.2 Convolutional approach to sentence-level feature extraction with different filter widths

其中, W^{1i}, b^{1i} 是模型需要学习的参数, 卷积单元的数量 cl_u^{1i} 是由用户指定的超参数。上下文窗口的大小 k^i 是由用户指定的超参数。 $\max(\cdot)$ 表示最大化操作。图中 W^{11} 用来学习窗口大小为 3 的给定句子的特征向量。

如图 2 所示, 本文在卷积层之后, 加入了 ReLU 层, 将 ReLU 作为激活函数。通过加入 ReLU 层可以加速随机梯度下降的收敛速度^[10]

将所有窗口生成的句子特征向量串接后得到的新特征向量如下:

$$r^{sent} = r^{sent1} \oplus r^{sent2} \oplus \dots \oplus r^{senti} \quad (3)$$

其中, \oplus 表示串接操作, r^{senti} 表示由上下文窗口大小为 k^i 的卷积核通过卷积核最大化操作之后得到的句子特征向量。

最后, 表示句子 x 的全局特征的向量 r_x^{sent} 被传递给包含两个全连接层的神经网络进行处理, 计算该句子属于每个情感标签 $\tau \in T$ 的得分:

$$s(x) = W^3 h(W^2 r_x^{sent} + b^2) + b^3 \quad (4)$$

其中, 矩阵 $W^2 \in \mathbb{R}^{hl_u \times (cl_u^1 + cl_u^2 + \dots + cl_u^i)}$, 矩阵 $W^3 \in \mathbb{R}^{|T| \times hl_u}$, 向量 $b^2 \in \mathbb{R}^{hl_u}$, 向量 $b^3 \in \mathbb{R}^{|T|}$ 是模型需要学习的超参数。激活函数 $h(\cdot)$ 使用的正切函数。隐藏层单元数目 hl_u 是有用户指定的超参数。

3. 2 模型训练

微博情感倾向性分析本质上是一个基于主题的文本分类问题, 将微博短文本做两类分类, 最终归纳到正面和负面两种情感类别中。

本文的模型是通过最小化训练集 D 上的负对数似然函数 (negative log-likelihood) 进行

训练。给定一个句子 x ，参数集合为 θ 的模型对每个情感标签 $\tau \in T$ 计算一个得分 $s_{\theta}(x)_{\tau}$ 。为了将这些得分转换为给定句子的情感标签和模型参数集 θ 的条件概率分布，我们在所有情感标签 $\tau \in T$ 的得分进行 Softmax 操作：

$$p(\tau|x, \theta) = \frac{e^{s_{\theta}(x)_{\tau}}}{\sum_{\forall i \in T} e^{s_{\theta}(x)_i}} \tag{5}$$

对上述公式取对数：

$$\log p(\tau|x, \theta) = s_{\theta}(x)_{\tau} - \log(\sum_{\forall i \in T} e^{s_{\theta}(x)_i}) \tag{6}$$

本文使用随机梯度下降（SGD）最小化负似然函数：

$$\theta \mapsto \sum_{(x,y) \in D} -\log p(y|x, \theta) \tag{7}$$

其中， x （变量变为斜体）表示训练预料 D 的一条句子， y 表示该句子的情感标签。

4 实验

4. 1 情感分析数据集

本文采用 COAE2014 任务 4 微博数据集，该数据集共 40000 条数据，其中官方公布了 5000 条微博的极性。由于没有标准的训练集和测试集，本文利用该 5000 条数据，进行 10 倍交叉验证。利用数据集提供的 40000 条数据训练词向量。

表 1 COAE2014 数据样例
Tab.1 COAE2014 Data sample

积极	消极
奥迪车质量值得 信赖 。	奥迪和人寿保险一样都是 垃圾 ！再不买！
三星手机很 好用 ，比较人性化。	三星手机真 垃圾 ，电池一点也不给力。关键时候 手机自动关机。
儿子住院自付部分，保险居然全给报销了，这个保险上的一点也不 亏 啊，嘿嘿嘿。	保险这玩意啊，就是他忽悠你交上，然后不想教得时候退回来 所剩无几 的。

4. 2 卷积神经网络

词向量在卷积神经网络模型中具有非常重要的作用。词向量在训练过程中，可以获取句法和语义信息，这在情感分析中是至关重要的^[13]。最近的一些工作表明通过使用无监督学习得到的词向量，可以极大的提高模型的准确率^[14,15,16]。在本文的实验中利用 word2vec 工具^[17]，进行无监督的词向量学习。word2vec 实现了 CBOW（continuous bag-of-words）和 SG（skip-gram）两种结构，用于计算词语的向量表示。

4. 2. 1 字级别的词向量

本文以字作为句子的基本单位，为每个字训练一个词向量。在运行 word2vec 工具过程中，本文设定出现次数超过 5 次的字会被加入字典中。最终得到了一个包含 5200 条目的字典。对于没有出现在字典中的字符的词向量会被随机初始化。训练过程中的参数设置如表 2 所示。

4. 2. 2 词级别的词向量

本文首先利用分词工具对语料进行分词。分词之后，以词作为句子的基本单位，为每个词训练一个词向量。在运行 word2vec 工具过程中，本文设定出现次数超过 5 次的词会被加

入字典中。最终得到了一个包含 19020 条目的字典。对于没有出现在字典中的词语的词向量会被随机初始化。训练过程中的参数设置如表 2 所示。

表 2 字级别和词级别 word2vec 可调参数的设置

Tab.2 Character-level and word-level word2vec hyper-parameters

可调参数	值
词向量维度	300
迭代次数	1
算法选择	skip-gram
上下文窗口	10
采样值	1e-3

4. 2. 3 超参数

本文对多个卷积神经网络模型进行了实验。训练过程中采用 Adadelta 更新规则^[18]，对乱序的微批次样本中进行随机梯度下降（SGD）。模型的其他参数如表 3 所示。

表 3 模型参数设置

Tab.3 Model hyper-parameters

参数	参数名	字级别	词级别
d	词向量维度	300	300
k^1	上下文窗口大小	2	2
cl_u^{11}	卷积单元数量	100	100
k^2	上下文窗口大小	3	3
cl_u^{12}	卷积单元数量	100	100
k^3	上下文窗口大小	5	5
cl_u^{13}	卷积单元数量	100	100
hl_u	隐藏层单元数量	300	300

4. 3 对比实验

本文与传统的词袋模型进行了对比，将 N-gram 词袋向量作为输入，利用线性核 SVM 进行了微博情感倾向性分类。本文对三种类型的 N-gram 词袋模型进行了测试：bow1 $N \in \{1\}$ ，bow2 $N \in \{1, 2\}$ ，bow3 $N \in \{1, 2, 3\}$ 。其中，bow1 是传统的词袋向量，bow2 向量中的每一个元素表示 uni-gram 或 bi-gram 特征，而 bow3 向量中的每一个元素表示 uni-gram、bi-gram 或者 tri-gram 特征，其中的特征值使用 TF-IDF 方法计算，并使用了 libSVM 进行 SVM 的实验^[19]。

4. 4 结果与讨论

本文对多个模型进行了实验。

- **SVM bow1:** 向量特征是 uni-gram 特征，利用 SVM 分类。
- **SVM bow2:** 向量特征是 uni-gram 和 bi-gram 特征，利用 SVM 分类。
- **SVM bow3:** 向量特征是 uni-gram、bi-gram 和 tri-gram 特征，利用 SVM 分类。
- **CNN-word-rand-static:** 将词级别的词向量进行随机初始化。在实验过程中，词向量保持不变，只学习模型的其他参数。
- **CNN-word-rand-non-static:** 将词级别的词向量进行随机初始化。在实验过程中，词向量会被微调，同时学习模型的其他参数。

- **CNN-word-static:** 利用 word2vec 训练出的词级别的词向量进行试验。在实验过程中，词向量保持不变，只学习模型的其他参数。
 - **CNN-word-non-static:** 利用 word2vec 训练出的词级别的词向量进行试验。在实验过程中，词向量会被微调，同时学习模型的其他参数。
 - **CNN-character-static:** 利用 word2vec 训练出的字级别的词向量进行实验。在实验过程中，词向量保持不变，只学习模型的其他参数。
 - **CNN-character-non-static:** 利用 word2vec 训练出的字级别的词向量进行试验。在实验过程中，词向量会被微调，同时学习模型的其他参数。
- 各个模型的实验结果如表 4 所示。

表 4 不同模型的准确率
Tab.4 Accuracy of different models

模型	COAE
SVM bow1	89.36
SVM bow2	91.74
SVM bow3	92.13
CNN-word-rand-static	91.94
CNN-word-rand-non-static	93.06
CNN-word-static	93.80
CNN-word-non-static	94.65
CNN-character-static	94.29
CNN-character-non-static	95.42

4. 4. 1 Random vs. word2vec

通过比较 CNN-word-rand-static (93.06%) 和 CNN-word-static (93.80%) 的准确率，可以发现，利用预训练的词向量作为原始输入进行情感倾向性分析的准确率要高于利用随机初始化的词向量作为原始输入进行的情感倾向性分析的准确率。原因在于利用 word2vec 工具训练出的词向量包含了上下文语义信息，因此在进行句子情感分析时可以得到更好的效果。实验表明，利用卷积神经网络模型进行自然语言处理时，无监督方式预训练的词向量是十分重要的。

4. 4. 2 Static vs. Non-static

通过比较 CNN-character-static (94.29%) 和 CNN-character-non-static (95.42%)，CNN-word-static(93.80%)和 CNN-word-non-static(94.65%)，CNN-word-rand-static(91.94%) 和 CNN-word-rand-non-static (93.06%) 的准确率，可以发现，在卷积神经网络的训练过程中对预训练的词向量进行微调，可以进一步提升模型的准确率。实验表明，在模型训练过程中对词向量进行微调，可以让预训练的词向量更加适应于专门的语料，从而进一步提高准确率。

4. 4. 3 Character-level vs. Word-level

通过比较 CNN-word-static (93.80%) 和 CNN-character-static (94.29%)，CNN-word-non-static (94.65%) 和 CNN-character-non-static (95.42%) 的准确率，可以发现，使用字级别词向量作为原始特征要好于使用词级别词向量作为原始特征。实验表明，对于中文语料而言，使用字级别词向量作为原始特征会好于使用词级别词向量作为原始特征。结合实验结果及分析原因主要在于：字级别特征的粒度比词级别的粒度小，字级别词向量相比于词级别词向量可以学习到更加具体的特征。

表 5 展示了利用 word2vec 工具训练的字级别词向量相加得到的词级别词向量相似度与直接用 word2vec 工具训练得到的词级别词向量相似度之间的比较。通过对比可以得出利用字级别词向量相加得到的词级别词向量之间的相似度要高于直接用 word2vec 工具训练得到的词级别词向量之间的相似度。

但是也存在一些词语组合不存在这种情况。例如，“三元”与“牛奶”之间的相似度为 0.783，因为“三元”是一个牛奶品牌，所以两者之间的相似度较高，而“三”+“元”与“牛”+“奶”之间的相似度要小于 0.783。出现这种情况的可能原因是由于语料中关于“三元”的内容不多，以至于字级别的词向量没有较好地学到相关的信息。

表 5 词向量相似度比较
Tab.5 Vector similarity comparison

词语	词语	余弦相似度
手机	三星	0.668
手+机	三+星	0.727
车险	保险	0.637
车+险	保+险	0.687
酸奶	牛奶	0.805
酸+奶	牛+奶	0.851

另一个原因在于：传统的分词技术往往对微博造成歧义的分词，比如：“我发现了一个高大上网站”，在该句中，“高大上网站”如果使用传统分词技术，会被切分为“高大/上/网站”或者“高大/上网/站”，这样的切分无法体现句子的正确语义，甚至后一种切分还将“网站”切分导致丢失评价对象。而将字级别特征作为输入，通过并行的卷积层可以学习到不同 N-gram 的信息，例如“高大上”（N=3）、“高大上网站”（N=5）。本文分别用 Jieba 分词工具和 ICTALAS 分词工具进行分词，得到了相近的实验结果。

4. 4. 4 Bag-of-N-gram vs. CNN

通过比较 SVM bow3（92.13%）和 CNN-character-non-static（95.42%）的准确率，可以发现，在微博情感倾向性分析中，卷积神经网络模型要优于传统的词袋模型。通过比较 SVM bow1（89.36%）、SVM bow2（91.74%）和 SVM bow3（92.31%）的准确率，可以发现，bi-gram 和 tri-gram 特征让准确率有了明显提升。而当 N 较大时，会造成维度灾难，导致模型难以训练。而包含多个并行卷积层的卷积神经网络，可以利用不同大小的卷积核学习不同 N-gram 的信息，通过池化操作降低维度，从而使得模型的准确率得以提高。

5 总结

本文探讨了利用卷积神经网络进行微博情感倾向性分析的可行性，并利用卷积神经网络模型取得了优于传统词袋模型的准确率，以此证明了卷积神经网络在微博情感倾向性分析中的可行性。本文利用字级别词向量及词级别词向量的卷积神经网络分别取得了 95.42% 的准确率和 94.65% 的准确率，实验结果可见，对于中文语料而言，利用卷积神经网络进行微博情感倾向性分析是有效的，且使用字级别的词向量作为原始特征会好于使用词级别的词向量作为原始特征。

参考文献

- [1] Pang B, Lee L. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales[C]//Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2005: 115-124.

- [2] 梁军, 柴玉梅, 原慧斌, 等. 基于深度学习的微博情感分析[J]. 中文信息学报, 2014, 28(5): 155-161.
- [3] 罗毅, 李利, 谭松波, 等. 基于中文微博语料的情感倾向性分析[J]. 山东大学学报 (理学版), 2014, 49(11): 1-7.
- [4] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [5] Yih W, He X, Meek C. Semantic parsing for single-relation question answering[C]//Proceedings of ACL. 2014.
- [6] Shen Y, He X, Gao J, et al. Learning semantic representations using convolutional neural networks for web search[C]//Proceedings of the companion publication of the 23rd international conference on World wide web companion. International World Wide Web Conferences Steering Committee, 2014: 373-374.
- [7] Blunsom P, Grefenstette E, Kalchbrenner N. A convolutional neural network for modelling sentences[C]//Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 2014.
- [8] Collobert R, Weston J, Bottou L, et al. Natural language processing (almost) from scratch[J]. The Journal of Machine Learning Research, 2011, 12: 2493-2537.
- [9] dos Santos C N, Gatti M. Deep convolutional neural networks for sentiment analysis of short texts[C]//Proceedings of the 25th International Conference on Computational Linguistics (COLING), Dublin, Ireland. 2014.
- [10] Kim Y. Convolutional neural networks for sentence classification[J]. arXiv preprint arXiv:1408.5882, 2014.
- [11] Turney P D. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews[C]//Proceedings of the 40th annual meeting on association for computational linguistics. Association for Computational Linguistics, 2002: 417-424.
- [12] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012: 1097-1105.
- [13] Collobert R. Deep learning for efficient discriminative parsing[C]//International Conference on Artificial Intelligence and Statistics. 2011 (EPFL-CONF-192374).
- [14] Luong M T, Socher R, Manning C D. Better word representations with recursive neural networks for morphology[J]. CoNLL-2013, 2013, 104.
- [15] Zheng X, Chen H, Xu T. Deep Learning for Chinese Word Segmentation and POS Tagging[C]//EMNLP. 2013: 647-657.
- [16] Socher R, Bauer J, Manning C D, et al. Parsing with compositional vector grammars[C]//In Proceedings of the ACL conference. 2013.
- [17] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[C]//Advances in Neural Information Processing Systems. 2013: 3111-3119.
- [18] Zeiler M D. ADADELTA: an adaptive learning rate method[J]. arXiv preprint arXiv:1212.5701, 2012.
- [19] Chang C C, Lin C J. LIBSVM: A library for support vector machines[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2011, 2(3): 27.

作者简介：



刘龙飞（1989—），男，硕士研究生，主要研究领域为机器学习、情感计算。
Email: liudragonfly@mail.dlut.edu.cn;



杨亮（1986—），男，博士研究生，主要研究领域为情感分析、自然语言理解。
Email: yangliang@mail.dlut.edu.cn;



张绍武（1967—），男，博士，副教授，主要研究领域为情感计算和搜索引擎。
Email: zhangsw@dlut.edu.cn。



林鸿飞（1962—），通信作者，男，博士，教授，主要研究领域为搜索引擎、文本挖掘、情感计算和自然语言理解。Email: hflin@dlut.edu.cn。