

## 基于多通道卷积神经网络的中文微博情感分析

陈珂<sup>1</sup> 梁斌<sup>2</sup> 柯文德<sup>1</sup> 许波<sup>1</sup> 曾国超<sup>1</sup>

<sup>1</sup>(广东石油化工学院计算机科学与技术系 广东茂名 525000)

<sup>2</sup>(苏州大学计算机科学与技术学院 江苏苏州 215000)

(chenke2001@163.com)

## Chinese Micro-Blog Sentiment Analysis Based on Multi-Channels Convolutional Neural Networks

Chen Ke<sup>1</sup>, Liang Bin<sup>2</sup>, Ke Wende<sup>1</sup>, Xu Bo<sup>1</sup>, and Zeng Guochao<sup>1</sup>

<sup>1</sup>(Department of Computer Science and Technology, Guangdong University of Petrochemical Technology, Maoming, Guangdong 525000)

<sup>2</sup>(School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu 215000)

**Abstract** Neural network-based architectures have been pervasively applied to sentiment analysis and achieved great success in recent years. However, most previous approaches usually classified with word feature only, which ignoring some characteristic features on the task of sentiment classification. One of the remaining challenges is to leverage the sentiment resources effectively because of the lack of length of Chinese micro-blog texts. To address this problem, we propose a novel sentiment classification method for Chinese micro-blog sentiment analysis based on multi-channels convolutional neural networks (MCCNN) to capture the characteristic information in micro-blog texts. With the help of the part of speech vector, the model could promote the full use of sentiment features through different part of speech tagging. Meanwhile, the position vector helps the model indicate the degree of importance of every word in the sentence, which impels the model to focus on the important words in the training process. Afterwards, a multi-channels architecture based on convolutional neural networks will be used to learn more feature information of micro-blog texts, and extract more hidden information through combining different vectors and original word embedding. Finally, the experiments on COAE2014 dataset and micro-blog dataset reveal better performance than the current main stream convolutional neural networks and traditional classifier.

**Key words** sentiment analysis; deep learning; convolutional neural networks (CNN); multi-channels; natural language processing

**摘要** 近年来,深度学习在情感分析任务中的应用得到了越来越多的关注.针对以文本词向量作为输入的卷积神经网络无法充分利用情感分析任务中特有的情感特征信息,以及难以有效表示每个词语在句子中的重要程度等问题,提出一种基于多通道卷积神经网络(multi-channels convolutional neural

收稿日期:2017-01-23;修回日期:2017-08-21

基金项目:国家自然科学基金项目(61272382,61672174);广东省自然科学基金项目(2016A030307049,2016A030307028);广东省科技计划项目(2014A010104016,2015B090903084)

This work was supported by the National Natural Science Foundation of China (61272382, 61672174), the Natural Science Foundation of Guangdong Province of China (2016A030307049, 2016A030307028), and the Science and Technology Planning Project of Guangdong Province (2014A010104016, 2015B090903084).

networks, MCCNN)的中文微博情感分析模型. 该模型针对情感分析任务中特有的情感信息来构建文本输入矩阵,使模型在训练过程中有效获取输入句子的情感特征信息. 同时,该模型通过将不同特征信息结合形成不同的网络输入通道,使网络模型在训练过程中从多方面的特征表示来学习输入句子的情感信息,有效表示出每个词语在句子中的重要程度,获取更多的隐藏信息. 最后在 COAE2014 数据集和微博语料数据上进行实验,取得了比普通卷积神经网络、结合情感信息的卷积神经网络和传统分类器更好的性能.

关键词 情感分析;深度学习;卷积神经网络;多通道;自然语言处理

中图法分类号 TP391

随着微博在社交网络上的兴起,越来越多的用户通过微博发表观点和表达情感. 如何利用机器学习和自然语言处理来分析微博用户的情感倾向已得到了越来越多研究人员的关注,并成为自然语言处理领域的研究热点之一<sup>[1]</sup>. 此外,微博文本句子的长度偏短和口语化词语的频繁使用,使得微博的情感分类相比普通文本情感分类更具有难度和挑战性.

传统的情感分类技术主要有基于规则和基于机器学习 2 类:1)基于规则方法主要采用根据经验或者专家意见得到的统计特征、情感词典和模板来对文本进行情感分类,需要大量的人工干预<sup>[2-3]</sup>;2)基于机器学习的方法通过人工标注一部分数据建立训练集,对训练集的数据进行特征提取和学习来构建一个分类模型,最后利用训练得到的分类模型来对未知标签的测试数据进行分类预测,自动实现情感极性的判断<sup>[4-5]</sup>. 其中基于机器学习方法在过去的研究中得到很多学者的重点关注,总结出了很多不同分类模型的组合方法. 这些方法已广泛应用于情感分析领域,并已取得了很好的成果. 但这类方法通常需要依赖复杂的特征工程,以及需要结合例如依存关系分析等外部知识.

近年来,随着深度学习在自然语言领域的研究和发展,越来越多的研究员开始用深度学习来解决情感分类问题<sup>[6]</sup>. 例如 Kim<sup>[7]</sup> 用卷积神经网络(convolution neural networks, CNN)对电影评论进行情感分类;Kalchbrenner 等人<sup>[8]</sup>用卷积神经网络解决 Twitter 的极性判断问题;Wang 等人<sup>[9]</sup>用长短期记忆(long-short term memory, LSTM)网络对文本情感极性进行分析. 这类基于深度神经网络的模型在无需大量人工特征的情况下取得了比传统分类器更好的效果. 一些学者针对微博短文本特有的特征信息来构建分类模型. 例如针对 Twitter 文本的大量特征信息,Vo 等人<sup>[10]</sup>提出了使用多样化特征的 Twitter 文本情感分类;针对现有的大量情

感信息,陈钊等人<sup>[11]</sup>提出了将情感特征加入卷积神经网络的方法;针对获取更多的语义信息,刘龙飞等人<sup>[12]</sup>提出了结合不同粒度的卷积神经网络模型. 虽然这类方法针对微博文本情感分类做了改进并取得了比传统方法更好的效果,但这类方法对情感特征的利用依赖于人工整理的情感词典,无法充分利用微博文本的情感信息. 同时,这类方法使用单一的特征表示,对输入向量的初始值依赖较大,且难以正确表示每个词在句子中的重要程度.

针对上述问题,本文提出一种将词语的词性进行向量化操作作为网络模型输入的方法,该方法通过将情感词典中的词语重新进行词性标注,并将不同的词性取值映射为一个多维的连续值向量,从而可以有效将输入文本的情感特征信息加入网络模型,使情感信息得到充分利用. 为了更准确地表示每一个词语在输入句子中的重要程度,本文将句子中不同词语的位置取值进行向量化操作,结合输入句子的词向量和词性向量形成不同的通道作为卷积神经网络的输入,可以使网络模型在训练过程中通过多种文本表示来学习句子的情感特征信息,挖掘更多的隐藏特征信息. 同时,本文将不同的文本表示组合形成不同的通道作为卷积神经网络的输入,可以使模型通过多方面信息学习不同输入特征之间的联系,有效降低网络对特征向量初始值的依赖性,在向量随机初始化的实验中也有不错的分类效果. 此外,本文提出的多通道卷积神经网络(multi-channels convolutional neural networks, MCCNN)模型在一次学习过程就可以完成对不同特征的学习和参数调整,有效降低了模型的训练时间代价. 在中文倾向性分析评测数据集(COAE2014)和微博语料数据集(micro-blog dataset, MBD)上与文献<sup>[11-12]</sup>提出的深度网络模型相比取得了更好的性能,最后比较张志琳等人<sup>[13]</sup>提出的基于传统分类模型的多样化

特征分类方法也取得了更好的分类效果,从而验证了本文方法的有效性。

本文主要贡献有 4 个方面:

- 1) 提出了一种 MCCNN 模型用在情感分析任务中,该模型将情感分析任务中不同的特征信息和卷积神经网络结合,有效提高了情感分类的正确率;
- 2) 提出了一种将不同词性映射为多维连续值的方法,该方法通过多维的词性向量,可以使网络有效识别不同词语种类对情感分类正确率的影响程度;
- 3) 将句子中不同词语的位置特征用连续值向量的形式加入网络模型,在训练过程中通过对位置向量的调整,可以有效获取不同词语在句子中的重要程度;
- 4) 在多个数据集上的对比实验表明,本文提出的 MCCNN 模型在情感分析任务中能有效识别不同类型句子的情感极性,取得更好的分类效果。

## 1 相关工作

### 1.1 情感分析

文本情感分析是通过对文本上下文内容的学习来判定文本的情感极性,是自然语言处理领域的一个重要分支。2002 年 Pang 等人<sup>[14]</sup>提出情感分析之后,有越来越多的研究员开始关注情感分析,也有越来越多的学者使用机器学习方法来解决文本情感判定。常用的基于机器学习的情感分类方法主要包括支持向量机、决策树、朴素贝叶斯等。在情感分析任务中,这类方法通过对数据集的信息提取和特征构建,取得了很好的分类效果。基于机器学习的方法通常结合文本的一元词特征、二元词特征、词性特征、情感特征等,将文本映射为多维向量,并通过分类模型学习特征信息。常用的特征映射方法有 One-hot 表示方法。该方法把文本的词条映射为多维向量,但忽略了词语之间的关系和文本的语义关系。Bengio 等人<sup>[15]</sup>在 2003 年提出了用多维的实数向量来表示文本的词条,考虑了词与词之间的位置关系和语义信息,总结出用神经网络建立统计语言模型的框架,同时学习词向量的权重和概率模型的参数;2013 年 Mikolov 等人<sup>[16]</sup>实现了用 CBOW 模型和 Skip-gram 模型计算词向量的方法,该方法很好地度量词与词之间的相似性,得到了很多学者的关注和使用。

对于短文本的情感分析,自 2009 年 Go 等人<sup>[17]</sup>首次提出微博情感分析以来,对例如微博的短文本情感极性判定也受到了越来越多人的重点关注。情

感词典对情感分类有很大的影响,常用的短文本情感分类技术是基于短文本的词向量信息和情感特征信息来构建模型。例如 Vo 等人<sup>[18]</sup>加入表情特征自动构建文本的情感词典来对 Twitter 文本进行情感分析。该方法有效利用了 Twitter 文本中不同表情隐藏的情感信息,通过对表情符号的学习,使模型充分利用输入文本的情感信息,有效提高了情感分类的性能。Tang 等人<sup>[19]</sup>通过情感种子扩充特定领域情感词对用户评论进行情感分类。该方法利用算法对特定领域数据集中的数据进行情感词的提取和扩充,使输入语料的情感词得到有效利用。此类方法都很好地结合了短文本特有的特征信息进行情感极性的判断,使情感分类有更好的效果。

### 1.2 卷积神经网络

卷积神经网络是深度学习的重要网络之一,它通过卷积层和池化层学习输入数据的局部特征<sup>[20]</sup>,是基于人工神经网络而提出的一种前馈神经网络。自 20 世纪 60 年代,Hubel 和 Wiesel<sup>[21]</sup>提出卷积神经网络以来,经过几十年的发展,如今 CNN 已经广泛应用于各种领域当中。由于卷积神经网络无需对数据进行复杂的预处理,并可以学习大量的特征信息,所以卷积神经网络在很多领域都取得了不错的成果。特别是在模式识别领域,卷积神经网络无需对图像进行复杂的预处理,直接把图像作为网络的输入,让卷积神经网络得到了学术界和工业界的广泛关注。

近年来,随着深度学习的发展和深入研究,卷积神经网络已被越来越多的学者应用到自然语言处理领域当中。目前,卷积神经网络已广泛应用于语义匹配<sup>[22]</sup>、词序列预测<sup>[23]</sup>和情感分析<sup>[24]</sup>等领域。文献<sup>[7]</sup>提出了将 CNN 模型应用到句子分类当中,利用多窗口的卷积核对输入文本进行卷积操作来提取局部特征。最后在多种数据集上进行实验,有效地提高了句子分类的性能,取得了比以往方法更好的效果。Chen 等人<sup>[25]</sup>通过动态多样化池化层卷积神经网络提取句子中的事件。该方法对 CNN 的池化层进行改进,可以有效获取输入文本更多的局部特征,保留更多的特征信息,有效获取输入文本中的事件信息。文献<sup>[12]</sup>以字级别和词级别的文本表示作为 CNN 的输入来学习句子的特征信息,通过不同级别的文本表示可以得到不同的特征集合。最后利用多卷积核对特征集合进行卷积操作,提取更丰富的情感特征信息,取得了比传统 SVM 方法更好的分类效果。文献<sup>[11]</sup>提出了一种将卷积神经网络和情感序列特征结合应用在情感极性分类的 WFCNN 模型。该模

型介绍了一种将情感分类任务中特有的情感信息和词向量拼接作为网络模型输入的方法,可以有效利用输入文本的情感特征信息. 本文提出的 MCCNN 模型和文献[11]的不同之处在于,本文使用多维连续值向量的形式来表示词语的词性特征,相比文献[11]使用二值取值来表示词语特征,模型在训练过程中可以通过更细微的参数调整来学习句子的情感信息,取得更好的情感分类效果.

## 2 情感分析模型

在文本情感分类任务中,文本的词特征,尤其是情感词,可以直接影响分类的性能<sup>[26]</sup>. 本文通过对微博文本中的情感词进行重新标注和向量化,作为卷积神经网络的输入,提出一种结合情感词的卷积神经网络模型(sentiment words convolution neural networks, SWCNN),验证本文提出将情感词重新标注和向量化在情感分析任务中的有效性. 此外,由

于微博文本的长度普遍较短,包含的特征信息有限. 为了把卷积神经网络更好地应用到微博的情感倾向分析任务中,本文在结合情感词的基础上加入句子中词语的位置信息,提出了一种结合多种特征的 MCCNN 模型. 该模型把不同的特征信息结合形成不同的通道作为卷积神经网络的输入,从而使模型获取更多的隐藏信息,有效表示每个词语在句子中的重要程度,取得更好的情感分类效果.

### 2.1 任务定义

对于长度为  $n$  的句子  $s = \{w_1, w_2, \dots, w_n\}$ , 其中  $w_i$  为句子  $s$  中的第  $i$  个词条,情感分析的任务是根据句子  $s$  的词序列所隐含的特征信息来判断句子  $s$  的情感极性. 基于文献[7]提出的模型,卷积神经网络也可以和其他特征结合作为网络的输入来构建模型. 一种简单的拼接结构如图 1 所示. 通过把词向量和不同特征的拼接作为卷积神经网络的输入,能使网络模型在训练过程中针对不同的特征信息来学习和调整模型的参数,获取更多的隐藏信息.

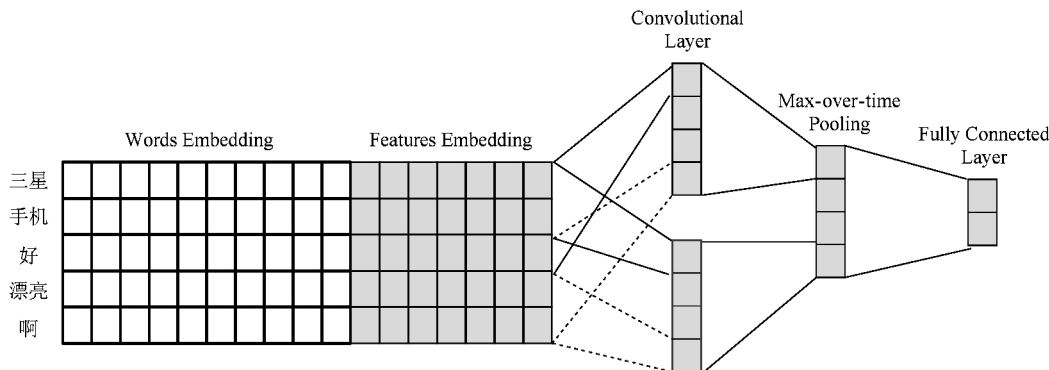


Fig. 1 The structure of features combination

图 1 特征拼接模型结构

### 2.2 构建特征

本文将通过加入情感分类任务中最重要的情感特征来阐述卷积神经网络和其他特征结合的方法. 介绍在输入文本内容特征的基础上,结合情感分析特征的情感词特征和词语在句子中的位置特征来构建网络模型的输入矩阵,并利用不同的输入通道来接收不同特征信息的组合,使模型在训练过程中学习更丰富的情感特征信息,有效识别短文本句子的情感极性.

神经网络通过接收文本的向量化输入来学习输入句子的特征信息,在文本分类任务中,句子中词语的内容隐含着句子最重要的特征信息. 本文以词为

单位来表示句子,将每一个词映射为一个多维的连续值向量,可以得到整个数据集词集合的词向量矩阵  $E \in R^{m \times |V|}$ , 其中  $m$  为每个词的向量维度,  $|V|$  为数据集的词条集合大小. 对于长度为  $n$  的句子  $s = \{w_1, w_2, \dots, w_n\}$ , 句子中每一个词语  $w_i$  可以映射为一个  $m$  维向量,即  $e_i \in R^m$ .

本文利用普通的 Hownet<sup>①</sup>情感词集合,对输入句子重新进行词性标注. 如表 1 所示,通过将句子中的特殊词语赋予特定的词性标注,可以让模型充分利用对情感分类有重要作用的词语,例如积极和消极情感词、否定词、程度副词等,从而在训练过程中注重学习这些词语的特征信息. 本文除了考虑句子

① <http://download.csdn.net/detail/monkey131499/9491884>

中的情感词,同时也对否定词和程度副词重新进行词性标注.例如“喜欢”是积极情感词,而“不喜欢”则是消极情感词,所以否定词会使句子隐含和情感词相反的情感极性.对于不同的词性标注,通过向量化操作,将每一种词性标注映射为一个多维的连续值向量  $tag_i \in \mathbb{R}^k$ ,其中  $tag_i$  为第  $i$  个词性向量,  $k$  为词性向量维度.网络模型在训练过程中可以针对不同的词性标注来对词性向量的各分量进行微调,从而可以学习到更细微的特征信息.

Table 1 POS Tagging

表 1 词性标注

Part of Speech	Tagging
Positive Sentiment Words	Pos
Negative Sentiment Words	Neg
Adverbs	Adv
Negative Words	Inver

因为微博的字数限制,微博文本的长度普遍比较简短,句子所含的情感信息有限,所以词条在微

博中的位置也是微博文本的一个重要特征.同一个词在不同位置出现,可能包含不同的信息.计算句子  $s$  中第  $i$  个词条  $w_i$  的位置值:

$$p(w_i) = i - len(s) + maxlen, \tag{1}$$

其中,  $p(w_i)$  为  $w_i$  在句子  $s$  中的位置值,  $i$  为词条  $w$  在句子  $s$  中的位置,  $len(s)$  为句子  $s$  的长度,  $maxlen$  为输入的句子最大长度.和词性向量操作一样,本文把每个位置值映射到一个  $l$  维向量,即  $position_i \in \mathbb{R}$ ,其中  $position_i$  为第  $i$  个位置值的向量.

2.3 结合词性特征的卷积神经网络

为了验证本文提出的将词性映射为多维向量方法的有效性,将输入句子的词性向量和句子内容层面的词向量结合作为卷积神经网络的输入,提出一种 SWCNN 模型.如图 2 所示,SWCNN 模型主要由输入层、卷积层、池化层和全连接层组成.输入层接收输入句子的特征矩阵;卷积层利用卷积核对输入的基本单位进行卷积操作提取特征;池化层对卷积层提取到的特征做采样处理,以过滤的形式保留重要的特征;全连接层通过提取到的特征信息输出待分类句子的分类结果.

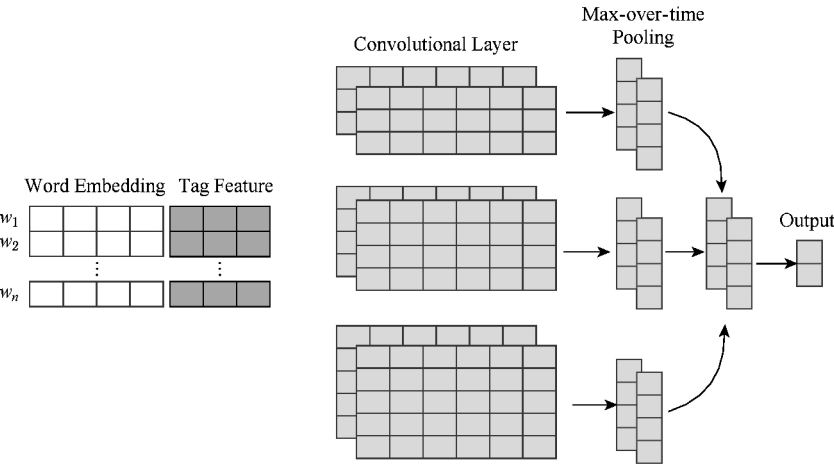


Fig. 2 Model architecture of SWCNN

图 2 SWCNN 模型结构图

本文以词为单位对句子进行卷积操作,对于长度为  $n$  的句子,其特征表示为

$$e_{1:n} = e_1 \quad e_2 \quad \cdots \quad e_n, \tag{2}$$

$$tag_{1:n} = tag_1 \quad tag_2 \quad \cdots \quad tag_n, \tag{3}$$

其中,  $e$  为词向量,  $tag$  为词性特征.为了简化网络模型结构,本文使用简单拼接操作形成特征矩阵  $x \in \mathbb{R}^{n+k}$ ,作为卷积神经网络的输入:

$$x = e \quad tag, \tag{4}$$

其中,  $\quad$  为拼接操作.本文通过把特定情感词映射为多维的词性特征,这可以使网络在训练过程中通过

调整词性特征分量来优化分类模型.实验中,本文对句子的输入设定一个最大长度  $maxlen$ ,对于长度小于  $maxlen$  的句子用  $0$  向量补全.

卷积层可以通过不同的卷积核对输入矩阵进行丰富的局部特征提取,对于长度为  $h$  的卷积核,可以把句子分为  $\{x_{0:h-1}, x_{1:h}, \dots, x_{i+h-1}, \dots, x_{n-h+1:n}\}$ ,然后对每一个分量进行卷积操作,得到卷积特征图:

$$C = (c_1, c_2, \dots, c_{n-h+1}), \tag{5}$$

其中,  $c_i$  是对分量  $x_{i:i+h-1}$  进行卷积操作后提取到的信息.

$$c_i = \text{relu}(W \cdot x_{i:i+h-1} + b), \quad (6)$$

其中,  $W \in \mathbb{R}^{h \times (m+k)}$  为卷积核权重,  $b \in \mathbb{R}$  为偏置. 本文在池化层利用 max-over-time pooling<sup>[27]</sup> 方法对特征信息进行采样, 提取最重要的特征信息:

$$\hat{c} = \max\{C\}, \quad (7)$$

上述  $\hat{c}$  为一个卷积核采样得到的结果, 对于  $d$  个卷积核采样得到的特征信息可以表示为

$$\hat{C} = (\hat{c}_1, \hat{c}_2, \dots, \hat{c}_d), \quad (8)$$

在 MCCNN 模型中, 把池化层采样得到的特征信息作为全连接层的输入, 得到分类结果:

$$y = \text{softmax}(W_f \cdot \hat{C} + b_f), \quad (9)$$

其中,  $b_f \in \mathbb{R}$  为偏置,  $W_f \in \mathbb{R}^d$  为全连接层权重,  $y$  为输出结果.

## 2.4 多通道卷积神经网络

如图 3 所示, 结合多特征的多通道卷积神经网络 MCCNN 模型主要由 6 部分组成:

1) 输入层. 本文主要使用 4 个通道来接收待分

类句子的不同特征组合, 使用不同的通道获取待分类句子更丰富的特征表示.

2) 卷积层. 对于不同的通道, 本文使用多窗口多卷积核的方式进行卷积操作, 获取不同通道输入的局部特征, 形成特征信息图.

3) 池化层. 为了使每个通道中的特征信息都能得到充分的利用, 本文使用不同的池化层对不同的通道进行下采样操作, 获取每个通道中最重要的特征信息.

4) 合并层. 本文采用一个合并层合并从不同通道获取的局部特征, 形成一个特征向量, 并将该特征向量作为隐藏层的输入.

5) 隐藏层. 为了获取不同通道局部特征之间的联系, 本文采用一个隐藏层对局部特征向量进行特征提取, 并可以通过权重矩阵学习不同通道的相互联系.

6) 输出层. 本文使用函数  $\text{softmax}$  输出待分类句子的分类结果.

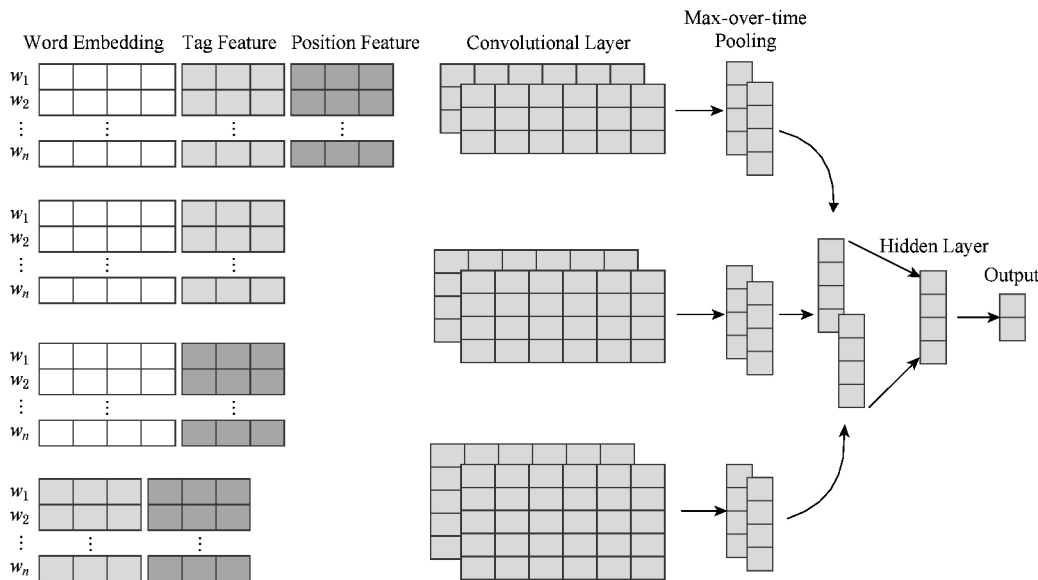


Fig. 3 Model architecture of MCCNN

图 3 多通道卷积神经网络模型结构

和普通的卷积神经网络相比, MCCNN 模型可以通过不同特征的组合变换形成新的通道作为网络的输入, 使模型可以根据多样化的输入, 在训练过程中获取到更多的语义信息. 因为不同特征结合除了形成新的特征之外, 也可以让特征之间有相互联系和影响. 同时, 不同通道的特征组合可以让网络在一次学习过程中就能完成对多个特征参数调整, 降低了网络模型的时间代价. 此外, MCCNN 模型中的每个通道是相互独立的, 在训练过程中模型可以对不同的通道使用不同的卷积核和不同的激活函数,

使模型学习到更加多样化的信息. MCCNN 模型除了使用 SWCNN 模型用到的词向量和词性特征之外, 还加入了词在文本中的位置特征. 对于长度为  $n$  的句子, 其位置特征:

$$\begin{aligned} & \text{position}_{1:n} = \text{position}_1 \\ & \text{position}_2 \quad \dots \quad \text{position}_n. \end{aligned} \quad (10)$$

对词特征、词性特征和位置特征采用不同的组合方式形成 4 个不同的通道作为网络的输入. 为了使网络模型简单化, 本文在实验中对特征的组合使用的是一种简单的拼接操作:

$$V_1 = w \quad tag \quad position, \quad (11)$$

$$V_2 = w \quad tag, \quad (12)$$

$$V_3 = w \quad position, \quad (13)$$

$$V_4 = tag \quad position, \quad (14)$$

和 SWCNN 模型一样,对于每个通道的特征信息,本文利用不同的卷积层对不同通道进行卷积操作来提取特征信息,对于长度为  $h$  的卷积核,卷积操作得到的特征向量图:

$$c_j = \text{relu}(W \cdot V_{i:i+h-1}^j + b), \quad (15)$$

然后利用池化层对特征信息进行过滤和提取,获取最重要的特征信息.实验中,本文对 4 个通道采用多窗口多卷积核的卷积操作,其中卷积核数量均为  $d$ ,通过池化操作,可以得到池化特征向量图:

$$\hat{C}_j = (\hat{c}_1, \hat{c}_2, \dots, \hat{c}_d), \quad (16)$$

式(15)(16)中,  $j = \{1, 2, 3, 4\}$  为通道下标.然后将不同通道的特征向量图通过合并层可以得到特征向量:

$$\hat{C} = \hat{C}_1 \quad \hat{C}_2 \quad \dots \quad \hat{C}_n. \quad (17)$$

此外,为了进一步获取更重要的特征信息和获取不同通道特征信息之间的联系,本文在 MCCNN 的池化层和全连接层之间加入一个隐藏层:

$$R = \text{relu}(W_h \hat{C}_h + b_h), \quad (18)$$

其中,  $R \in R^q$  为隐藏层输出,  $q$  为隐藏层输出维度.  $W_h \in R^{q \times d}$  为隐藏层权重矩阵,  $b_h \in R^q$  为偏置.通过该隐藏层可以更好地过滤影响分类性能的信息,获取到更重要的特征.然后将隐藏层提取到的特征向量作为全连接层的输入,和 SWCNN 模型一样,本文利用一个函数  $\text{softmax}$  输出待分类句子的分类结果.实验中,模型通过最小化交叉熵来调整模型参数,优化网络模型的性能:

$$\text{loss} = -\frac{1}{D} \sum_s y \ln \hat{y} + \lambda \|\theta\|^2, \quad (19)$$

其中,  $D$  为训练数据集大小,  $y$  为预测类别,  $\hat{y}$  为实际类别,  $\lambda \|\theta\|^2$  为正则项.

### 3 实 验

本文采用 2014 年中文观点倾向性分析评测 (COAE2014) 语料中的任务 4 数据集<sup>①</sup>和从网上爬取得到的微博文本形成不同的数据集进行对比实验,来对本文提出方法的性能进行评估.从 COAE2014 数据集中标注 6 000 条带有极性的数据,其中正面情

绪 2 864 条、负面情绪 3 136 条.为了丰富数据集的多样性,本文从不同领域爬取 5 000 条带有极性的微博文本,作为微博语料数据集 (micro-blog dataset, MBD),其中正面情绪和负面情绪各 2 500 条.此外,为了验证本文提出方法在混合数据集的情感分类有效性,本文从 COAE2014 和微博语料数据集各抽取 5 000 条数据形成混合数据集来完成对比试验,详细的数据统计如表 2 所示:

Table 2 Statistic of Datasets

表 2 实验使用数据统计

Dataset	Corpus	Positive	Negative
COAE-train	COAE2014	2 248	2 427
COAE-test	COAE2014	616	709
MBD-train	MBD	2 000	2 000
MBD-test	MBD	500	500
COAEMBD-train	COAE2014+MBD	4 000	4 000
COAEMBD-test	COAE2014+MBD	1 000	1 000

#### 3.1 数据预处理

本文使用 ICTCLAS 分词工具<sup>②</sup>对表 2 所示的实验数据进行分词和词性标注.词向量和词性特征由网上爬取的微博语料产生,采用 Google 的 word2vec 工具<sup>③</sup>的 skip-gram 模型在特征向量可训练实验中对词向量和词性特征向量进行训练,未登录词使用均匀分布  $U(-0.01, 0.01)$  来随机初始化.由于位置特征的取值较少,本文使用均匀分布  $U(-0.01, 0.01)$  对位置特征向量进行随机初始化.实验中,词向量维度为 100 维,词性特征为 50 维,位置特征为 10 维.本文剔除了出现次数少于 5 次的词条,其余参数使用 word2vec 的默认参数.

#### 3.2 超参数

本文在实验中使用了多种窗口卷积核对输入向量进行卷积操作,卷积核函数为 rectified linear units.训练过程采用 Zeiler<sup>[28]</sup>提出的 Adadelta 更新规则.模型的参数设置如表 3 所示:

Table 3 Hyper Parameters of Experiment

表 3 模型参数设置

Parameter	Parameters Description	Value
$h$	Window Size	3, 4, 5
$m$	No. of Features Map	100
$b$	Mini-batch Size	50
$s$	L2 Constrain	3
$p$	Dropout Rate	0.5

① <http://download.csdn.net/detail/hzssssshuo/8708735>

② <http://ictclas.nlpir.org/>

③ <http://word2vec.googlecode.com/svn/trunk/>

3.3 实验介绍

将本文提出的 2 种模型和文献[11-13]提出的方法在表 2 所示数据集上进行对比实验,验证本文提出方法的有效性,各实验介绍如下:

- 1) Rich-features. 文献[13]提出的多样化分类特征方法,实验中使用 SVM 分类器.
- 2) WFCNN-rand. 文献[11]提出的结合情感序列的卷积神经网络模型,但模型词向量随机初始化.
- 3) WFCNN. 文献[11]提出的 WFCNN 模型,并使用 word2vec 训练词向量.
- 4) CNN-rand. 文献[12]提出的卷积神经网络模型,实验中词向量随机初始化.
- 5) CNN. 文献[12]提出的模型,并利用 word2vec 训练词向量.
- 6) SWCNN-rand. 本文提出的将词性取值映射为多维向量加入情感信息的方法,但在实验中词性特征采用随机初始化.
- 7) SWCNN. 本文提出的将词性取值映射为多维向量加入情感信息的方法,实验中词性特征利用 word2vec 进行训练.
- 8) MCCNN-rand. 本文提出的多通道卷积神经网络模型,但在实验中随机初始化特征向量.
- 9) MCCNN. 本文提出的多通道卷积神经网络模型,并利用 word2vec 训练特征向量.

3.4 实验结果与分析

本文将在表 2 所示的 3 个不同数据集上完成 9 组对比实验来验证本文提出方法的有效性. 各数据集上的对比实验结果如表 4 所示:

Table 4 Accuracy on Sentiment Classification of Different Models

表 4 不同模型的情感分类性能对比

Model	Accuracy		
	COAE	MBD	COAEMBD
Rich-features	0.8204	0.8290	0.8185
WFCNN-rand	0.7887	0.8120	0.7840
WFCNN	0.8279	0.8400	0.8240
CNN-rand	0.7841	0.8040	0.7835
CNN	0.8242	0.8330	0.8220
SWCNN-rand	0.7970	0.8100	0.7905
SWCNN	0.8392	0.8460	0.8290
MCCNN-rand	0.8377	0.8520	0.8385
MCCNN	<b>0.8491</b>	<b>0.8580</b>	<b>0.8520</b>

从表 4 结果可以看出,本文提出的 MCCNN 模型在 3 个数据集上都取得了最好的情感分类效果,其中在最好的 MBD 数据集上取得了 85.80% 的分

类正确率,相比 WFCNN 模型的 84.00% 和 CNN 模型的 83.30% 分别提升了 1.80% 和 2.50%,验证了本文提出方法在情感分析任务中的有效性. 从表 4 结果也可以看出,利用 word2vec 训练词向量的 4 种深度学习模型在 3 个数据集上都取得了比使用传统方法的 Rich-features 模型更好的分类效果,说明深度网络模型在情感分析任务中相比传统方法有更好的效果. 对比文献[11]提出的 WFCNN 模型和文献[12]提出的 CNN 模型可以看出,加入情感信息的 WFCNN 模型在 3 个数据集上的分类正确率相比不使用情感特征信息的 CNN 模型都有不同程度的提升,说明在情感分析任务中加入情感信息能有效提高情感分析的正确率. 对比本文提出的将特征信息以词性向量的形式加入卷积神经网络的 SWCNN 模型和文献[11]提出的对情感信息进行二值取值的 WFCNN 模型可以看出,本文提出的 SWCNN 模型在使用 word2vec 训练特征向量和向量随机初始化实验中都取得了比 WFCNN 模型更好的分类效果. 其中,在提升幅度最高的 COAE 数据集上,本文提出的 SWCNN 相比 WFCNN 模型的分正确率分别提升了 1.13% 和 0.83%,说明本文提出的将情感信息以向量形式加入卷积神经网络的方法能使句子的情感特征信息在网络模型中得到更充分的利用,从而取得更好的情感分类效果. 此外,从表 4 中结果也可以看出,4 种深度网络模型在利用 word2vec 训练特征向量的实验中都取得了比向量随机初始化更好的分类效果,说明了特征向量的初始值会影响情感分析的分类效果. 为了进一步分析特征向量的初始值对分类性能的影响,本文从 COAE 数据集中随机抽取 5 000 条样本进行 10 倍交叉验证对比实验,对比结果如表 5 所示:

Table 5 Experimental Results of Cross Validation

表 5 交叉验证实验结果

Model	Accuracy
WFCNN-rand	0.9306
WFCNN	0.9462
CNN-rand	0.9302
CNN	0.9448
SWCNN-rand	0.9424
SWCNN	0.9572
MCCNN-rand	0.9583
MCCNN	<b>0.9643</b>

综合表 4 和表 5 实验结果,4 组对比实验在随机初始化特征向量和利用 word2vec 训练特征向量的实验对比结果如图 4 所示:



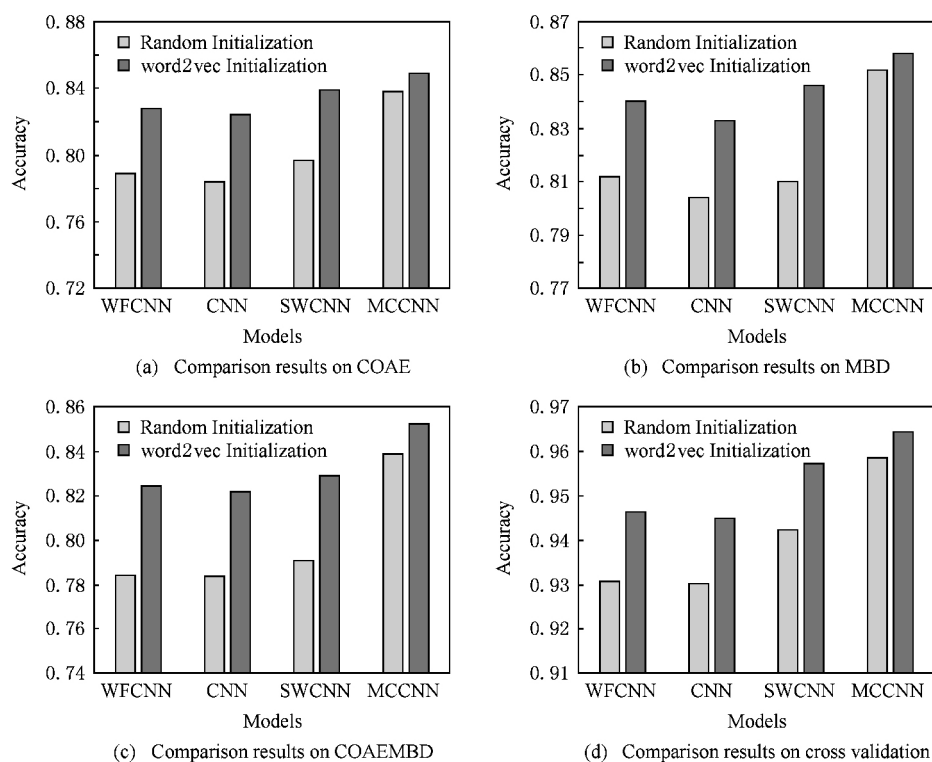


Fig. 4 Comparison results of random and word2vec initialization embedding in different datasets

图4 不同数据集上随机初始化特征向量和利用 word2vec 训练特征向量的实验对比

从图4结果可以看出,使用 word2vec 训练特征向量的模型在所有的实验中都取得比随机初始化特征向量更好的分类效果,说明利用 word2vec 训练特征向量能有效提高情感分类的正确率. 分析结果可知,相比随机初始化特征向量,利用 word2vec 训练词向量给词条赋予初始值可以使网络在训练过程中更好地学习和调整参数,使模型有更优的分类性能. 此外,从图4结果也可以看出,本文提出的 MCCNN 模型在随机初始化特征向量的实验中也能取得不错的分类效果. 同时,随机初始化特征向量相比用 word2vec 训练特征向量的分类正确率降幅较小,说明本文提出的 MCCNN 模型能有效降低模型对特征向量初始值的依赖. 分析实验结果可知,利用多通道输入来接收待分类句子多样化表示的 MCCNN 模型中每个通道的输入不仅包含了特征本身的信息,还包含了特征之间的联系,模型可以根据不同通道的特征信息能充分挖掘不同特征之间相互联系和更多的隐藏信息,有效弥补了使用单一通道因为特征向量初始值不合理而难以提取更多特征信息的不足,降低了模型对向量初始值的依赖.

此外,从表4结果可以看出, WFCNN, CNN, SWCNN 这3个模型在混合数据集 COAEMBD 上的分类结果都不理想,3个模型在 COAEMBD 数据

集上的分类正确率相比 COAE 和 MBD 数据集都有所下降. 说明当数据集使用混合领域的数据样本时,普通卷积神经网络难以结合不同领域数据中的输入句子来学习整个数据集的特征信息. 而对于本文提出的 MCCNN 模型,在 COAEMBD 数据集上的分类正确率为 85.20%,虽然比在 MBD 数据集上的分类正确率(85.80%)降低了 0.60%,但是比 COAE 数据集的分类正确率(84.91%)提升了 0.29%. 说明本文提出的基于多通道的卷积神经网络模型在混合数据集上也能取得不错的分类效果,能有效缓解因数据样本分布不均匀给模型调参带来的难度. 因为 MCCNN 模型每个通道中不同特征的结合可以使模型提取到输入句子更丰富的特征信息. 同时, MCCNN 模型加入了词语的位置特征信息,能有效表示出每个词语在句子中的重要程度,在某类句子样本数量有限时也可以充分挖掘句子的特征信息,从而取得比传统网络模型更好的分类效果.

和传统的卷积神经网络相比,本文提出了一种将词性映射为多维词性向量的方法. 通过将词性向量和词语的词向量结合,使模型在训练过程中更好地学习输入句子的情感特征信息. 结合表4的实验结果,对比本文提出的 SWCNN 模型和传统方法的 WFCNN 和 CNN 模型在不同数据集上的情感分类

正确率来分析本文提出方法的有效性,对比结果如图 5 所示:

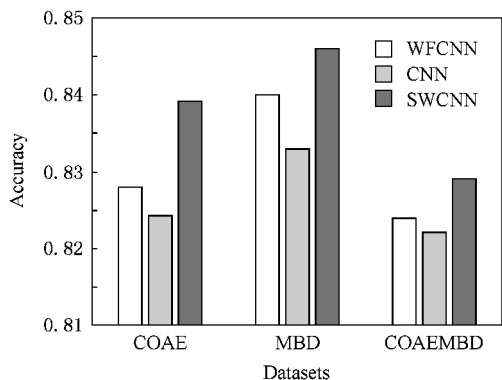


Fig. 5 Comparison result of different models on different datasets

图 5 模型在 3 种数据集上的对比结果

从图 5 结果可以看出,本文提出的将词性映射为多维向量的 SWCNN 模型在 3 个数据集上都取得了最好的分类效果.没有加入情感特征信息的 CNN 模型的情感分类效果并不理想,在 3 个数据集上的分类正确率都比不上加入情感特征信息的 WFCNN 和 SWCNN 模型,说明在情感分析任务中加入情感特征信息能有效提高情感分类正确率.此外,对比以二值特征形式加入情感信息的 WFCNN 模型,本文提出的 SWCNN 模型在情感分类任务中有更好的表现,在 3 个数据集上的情感分类正确率都超过了 WFCNN 模型.对比结果表明,本文提出的将词性映射为多维的向量特征可以使模型在训练过程中通过调整词性向量不同分量的取值来学习更深层次的情感信息,挖掘更多的隐藏特征,从而取得更好的情感分类效果.

为了进一步分析不同维度的词性特征对分类效果的影响,本文利用 SWCNN 和 MCCNN 模型在 COAE 数据集上使用不同维度的词性特征进行对比分析,实验结果如图 6 所示,其中词性向量维度为 0 表示不使用词性向量.

从图 6 可以看出,4 组实验在词性特征维度小于 50 时都呈现上升的趋势,其中随机初始化特征向量的 MCCNN-rand 模型上升最为明显;但是当词性特征维度超过 50 维之后,随机初始化特征向量的 SWCNN-rand 和 MCCNN-rand 模型随着维度的增加分类正确率呈现下降趋势,而利用 word2vec 训练特征向量初始值的 SWCNN 和 MCCNN 模型在词性特征维度超过 50 之后的分类正确率出现了波动.在随机初始化特征向量的 2 组实验中,当词性向量

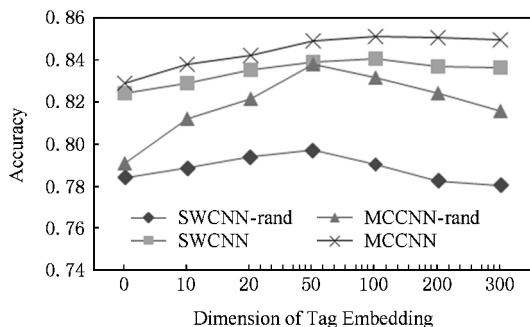


Fig. 6 Comparison of tag embedding in different dimensions

图 6 不同维度词性特征比较

的维度增加时,模型可以调整词性向量更多的分量参数来学习待分类句子的情感信息,所以在词性向量维度小于 50 时分类正确率呈现稳定上升的趋势.随着词性特征维度的增加,模型在一次迭代过程需要调整更多的权重和向量参数,随机初始化特征向量时,有可能会给词性赋予一个和真实值相差很大的特征向量,使得模型在训练过程中难以通过参数调整来逼近真实的特征向量.所以当特征维度超过某个阈值时,随机初始化特征向量的模型分类正确率会随着词性特征维度的增加而降低.对于使用 word2vec 训练特征向量初始值的 SWCNN 和 MCCNN 模型,当词性特征维度超过 50 时的分类正确率上升也不明显,而词性特征向量维度越大,模型的训练时间就越长.所以词性向量的维度并非越大越好.从图 6 也可以看出,在词性特征维度为 0,即不使用词性特征的实验中,加入位置特征的 MCCNN 模型相比 SWCNN 模型有更好的分类正确率,说明位置特征能提升模型分类正确率.

此外,由于内容层面的词特征是卷积神经网络最主要的特征,所以当词性特征的维度过大时,模型将无法主要针对内容层面的词向量进行学习和调参,从而会影响模型分类性能.为了分析不同维度的词向量对模型分类性能的影响,本文对 SWCNN 和 MCCNN 模型取不同维度的词向量在 COAE 数据集上进行对比实验,实验结果如图 7 所示.

从图 7 结果可以看出,在词向量维度小于 100 时,4 组实验的正确率都有明显的上升趋势.在词向量维度大于 100 维的时候,利用 word2vec 训练词向量的 2 组实验的分类正确率呈现很平缓的上升,说明利用 word2vec 训练词向量实验的分类正确率能随着词向量维度的增加而提升.但是对于随机初始

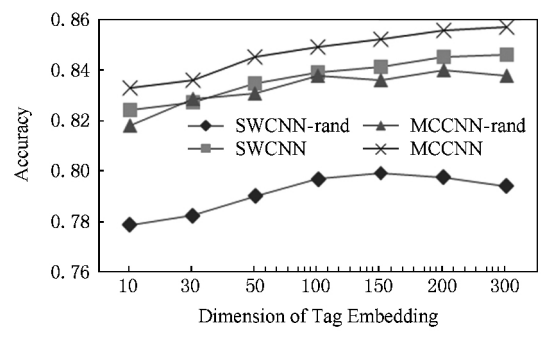


Fig. 7 Comparison of word embedding in different dimensions

图 7 不同维度词向量比较

化词向量的 2 组实验,在词向量维度超过 100 的时候分类正确率都出现了波动. 由于对向量初始值的依赖,随着词向量维度的增加,利用随机初始化赋予初始值的模型并不能很好地学习向量的特征信息,所以分类正确率出现了波动. 并且,随着向量维度的增加,模型的训练时间也会增加. 所以本文在实验中的词向量维度设为 100 维,词性向量维度为 50 维,而作为辅助网络模型训练的位置特征向量则取 10 维.

3.5 典型例子分析

为了进一步分析本文提出方法对比传统的深度学习模型的优点,本文从 COAE 数据集中抽取一些典型句子的分类结果进行对比分析.

如表 6 所示,对于句子 1 和句子 2 这类情感极性明显、结构简单的句子,普通的 CNN, WFCNN 模型和本文提出的 MCCNN 模型都能正确识别这类句子的情感极性,得到正确的分类结果. 句子 3 属于语句较为复杂的句型. 此类句型句子长度较长,且无明显的情感信息,所以普通的卷积神经网络模型很难通过有限的训练样本学习到这类句型的情感极性,所以 CNN 模型得到一个错误的分类结果. 而对

于加入特征信息的 WFCNN 模型和本文提出的 MCCNN 模型,可以通过情感特征来学习句子的隐藏特征信息,从而能有效识别这类句子的情感极性,取得正确的分类结果. 对于句子 4 这类字面上没有明确的情感信息且带有否定词的句型,普通 CNN 模型和加入情感信息的 WFCNN 都无法根据有限的特征信息来学习这类句子的情感极性,从而得到一个错误的分类结果. 而对于 MCCNN 模型,因为多通道卷积神经网络学习的不仅是单个特征,还有特征之间的联系,所以 MCCNN 模型可以根据不同特征之间的联系来学习到更多的语义信息,比如某个词的内容、该词的词性以及该词在句子中不同的位置,根据这些信息,模型就可以自动学习到这类情感极性语句的结构,从而可以得到和人工标注一致的结果. 对于句子 5 和句子 6,是一类关于反问句的样例,这类句子字面上往往有着和正确极性相反的情感,所以加入情感信息的 WFCNN 模型仅从单纯的字面信息来判断句子的情感极性,得到一个错误的分类结果. 同样的,普通的 CNN 模型也很难根据句子的信息来判断情感极性,所以句子 6 判断错误. 对于 MCCNN 模型,这类字面上有“?”、“什么”等词的句型,因为本文提出的方法对这类词语进行了词性特征的学习,所以 MCCNN 模型可以根据这些词在句子中的位置和词性特征来学习这类句型的情感信息,根据不用通道的输入信息学习不同特征之间的联系,让模型学习到和训练集标注一致的分类结果. 对于句子 7 和句子 8,这一类属于更复杂的带讽刺情感的句型, CNN 和 WFCNN 模型都很难判断这类数据的极性,如句子 7 和句子 8 的结果都和人工标注相反. 本文提出的 MCCNN 模型通过学习更多隐藏的语义信息,对这类数据也有比较好的极性判别效果,如句子 8 结果和人工标注一致.

Table 6 Analysis of Typical Sentences

表 6 经典句型例子分析

Sentences	Content	WFCNN	CNN	MCCNN	Label
1	粉色的奥迪,好喜欢啊	Positive	Positive	Positive	Positive
2	三星手机就是垃圾	Negative	Negative	Negative	Negative
3	车船使用税怎么这么高了,快赶上保险钱了,不带这么抢钱的啊.	Negative	Positive	Negative	Negative
4	我那年带的三星手机,坚持不了一会.	Positive	Positive	Negative	Negative
5	咱人民银行还买什么赔什么呢,怎能相信一个卖保险的?	Positive	Negative	Negative	Negative
6	拿什么信任你? 我的蒙牛	Positive	Positive	Negative	Negative
7	高峰期停在站台的奥迪小轿车,所有公交车靠站都有躲它,真牛!	Positive	Positive	Positive	Negative
8	哎真厉害! 为了 6 块钱值得么? 修车得多少钱啊	Positive	Positive	Negative	Negative

## 4 总 结

本文提出一种基于多通道卷积神经网络的微博情感分析模型,该模型利用多通道卷积神经网络提取更多的语义信息和学习更多的隐藏信息.实验结果表明,本文提出的多通道卷积神经网络在不同的数据集均取得了比对比方法更好的分类性能,在混合数据集 COAEMBD 数据集上也取得不错的分类效果,验证了多通道卷积神经网络的鲁棒性.此外,多通道卷积神经网络模型降低了对特征向量初始值的依赖性,这一结果表明多通道卷积神经网络可以结合更多的特征来学习和优化模型.但是,通过分析经典例子也可以看出,对于含有讽刺情感的句子,本文提出的多通道卷积神经网络模型仍然不能很好地识别这类句子的情感极性.

在下一步工作中,可以对不同的通道采用不同的特征组合方式,以及在不同的通道采用不同的激活函数,让模型学习到更多的特征信息,并且针对带有讽刺情感的句型来改进本文提出的多通道卷积神经网络模型.

## 参 考 文 献

- [1] Pang B, Lee L. Opinion mining and sentiment analysis [J]. Foundations and Trends in Information Retrieval, 2008, 2(1-2): 1-135
- [2] Joshi A, Balamurali A R, Bhattacharyya P, et al. C-Feel-It: A sentiment analyzer for micro-blogs [C] //Proc of the ACL-HLT 2011 System Demonstrations. Stroudsburg, PA: ACL, 2011: 127-132
- [3] Chesley P, Vincent B, Xu Li, et al. Using verbs and adjectives to automatically classify blog sentiment [J]. Training, 2006, 580(263): 233-235
- [4] Boiy E, Moens M F. A machine learning approach to sentiment analysis in multilingual Web texts [J]. Information Retrieval, 2009, 12(5): 526-558
- [5] Ye Qiang, Zhang Ziqiong, Law R. Sentiment classification of online reviews to travel destinations by supervised machine learning approaches [J]. Expert Systems with Applications, 2009, 36(3): 6527-6535
- [6] Yu Kai, Jia Lei, Chen Yuqiang, et al. Deep learning: Yesterday, today, and tomorrow [J]. Journal of Computer Research and Development, 2013, 50(9): 1799-1804 (in Chinese)  
(余凯, 贾磊, 陈雨强, 等. 深度学习的昨天、今天和明天 [J]. 计算机研究与发展, 2013, 50(9): 1799-1804)
- [7] Kim Y. Convolutional neural networks for sentence classification [C] //Proc of the 2014 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2014: 1746-1751
- [8] Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences [C] //Proc of the 52nd Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2014: 655-665
- [9] Wang Xin, Liu Yuanchao, Sun Chengjie, et al. Predicting polarities of tweets by composing word embeddings with long short-term memory [C] //Proc of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Int Joint Conf on Natural Language Processing. Stroudsburg, PA: ACL, 2015: 1343-1353
- [10] Vo D T, Zhang Yue. Target-dependent Twitter sentiment classification with rich automatic features [C] //Proc of the 24th Int Joint Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2015: 1347-1353
- [11] Chen Zhao, XuRuifeng, Gui Lin, et al. Combining convolutional neural networks and word sentiment sequence features for Chinese text sentiment analysis [J]. Journal of Chinese Information Processing, 2015, 29(6): 172-178 (in Chinese)  
(陈钊, 徐睿峰, 桂林, 等. 结合卷积神经网络和词语情感序列特征的中文情感分析 [J]. 中文信息学报, 2015, 29(6): 172-178)
- [12] Liu Longfei, Yang Liang, Zhang Shaowu, et al. Convolutional neural networks for Chinese micro-blog sentiment analysis [J]. Journal of Chinese Information Processing, 2015, 29(6): 159-165 (in Chinese)  
(刘龙飞, 杨亮, 张绍武, 等. 基于卷积神经网络的微博情感倾向性分析 [J]. 中文信息学报, 2015, 29(6): 159-165)
- [13] Zhang Zhilin, Zong Chengqing. Sentiment analysis of Chinese micro blog based on rich-features [J]. Journal of Chinese Information Processing, 2015, 29(4): 134-143 (in Chinese)  
(张志琳, 宗成庆. 基于多样化特征的中文微博情感分类方法研究 [J]. 中文信息学报, 2015, 29(4): 134-143)
- [14] Pang B, Lee L, Vaithyanathan S. Thumbs up?: Sentiment classification using machine learning techniques [C] //Proc of the ACL-02 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2002: 79-86
- [15] Bengio Y, Ducharme R, Vincent P, et al. A neural probabilistic language model [J]. Journal of Machine Learning Research, 2003, 3(Feb): 1137-1155
- [16] Mikolov T, Sutskever I, Chen Kai, et al. Distributed representations of words and phrases and their compositionality [C] //Proc of the 27th Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 2013: 3111-3119
- [17] Go A, Bhayani R, Huang Lei. Twitter sentiment classification using distant supervision [R/OL]. Palo Alto: Stanford Library Technologies Project, 2009 [2017-01-01]. <https://cs.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.pdf>

- [18] Vo D T, Zhang Yue. Don't count, predict! An automatic approach to learning sentiment lexicons for short text [C] // Proc of the 54th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2016: 219-224
- [19] Tang Duyu, Qin Bing, Zhou Lanjun, et al. Domain-specific sentiment word extraction by seed expansion and pattern generation [J/OL]. [2017-01-04]. <https://arxiv.org/pdf/1309.6722.pdf>
- [20] 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
- [21] Hubel D H, Wiesel T N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex [J]. The Journal of Physiology, 1962, 160(1): 106-154
- [22] Hu Baotian, Lu Zhengdong, Li Hang, et al. Convolutional neural network architectures for matching natural language sentences [C] //Proc of the 28th Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 2014: 2042-2050
- [23] Wang Mingxuan, Lu Zhengdong, Li Hang, et al. genCNN: A convolutional architecture for word sequence prediction [C] //Proc of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Int Joint Conf on Natural Language Processing. Stroudsburg, PA: ACL, 2015: 1567-1576
- [24] Wang Jin, Yu L C, Lai K R, et al. Dimensional sentiment analysis using a regional CNN-LSTM Model [C] //Proc of the 54th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2016: 225-230
- [25] Chen Yubo, Xu Liheng, Liu Kang, et al. Event extraction via dynamic multi-pooling convolutional neural networks [C] //Proc of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Int Joint Conf on Natural Language Processing. Stroudsburg, PA: ACL, 2015: 167-176
- [26] Tang Duyu, Wei Furu, Qin Bing, et al. Building large-scale Twitter-specific sentiment lexicon: A representation learning approach [C] //Proc of the 25th Int Conf on Computational

Linguistics: Technical Papers. Stroudsburg, PA: ACL, 2014: 172-182

- [27] Collobert R, Weston J, Bottou L, et al. Natural language processing (almost) from scratch [J]. Journal of Machine Learning Research, 2011, 12(Aug): 2493-2537

- [28] Zeiler M D. ADADELTA: An adaptive learning rate method [J/OL]. [2017-01-03]. <https://arxiv.org/pdf/1212.5701.pdf>



**Chen Ke**, born in 1964. Professor. His main research interests include data mining and machine learning.



**Liang Bin**, born in 1993. Master candidate. His main research interests include sentiment analysis, machine learning, and deep learning.



**Ke Wende**, born in 1976. PhD, professor. His main research interests include artificial intelligence and machine learning.



**Xu Bo**, born in 1982. PhD, associate professor. His main research interests include cloud computing, computational intelligence, and multi-agent system.



**Zeng Guochao**, born in 1978. Experimentalist. His main research interests include data mining, data visualization and machine learning.