

文章编号: 1003-0077(2018)09-0125-09

## 基于词注意力卷积神经网络模型的情感分析研究

王盛玉<sup>1</sup>, 曾碧卿<sup>1,2</sup>, 商 齐<sup>1</sup>, 韩旭丽<sup>1</sup>

(1. 华南师范大学 计算机学院, 广东 广州 510631;

2. 华南师范大学 软件学院, 广东 佛山 528225)

**摘 要:** 情感分类任务需要捕获文本中的情感特征, 利用重要的局部特征构建文本的特征表示。卷积神经网络(convolutional neural networks, CNN)已经被证明拥有出色的特征学习能力, 但是该模型无法判别输入文本中特征词与情感的相关性, 卷积层缺乏对单一词特征的提取。基于目前运用非常成功的注意力模型, 该文提出一种基于词注意力的卷积神经网络模型(word attention-based convolutional neural networks, WACNN)。相比于卷积神经网络, 该模型以篇章的文本信息作为输入, 首先在词嵌入层之后增加注意力机制层, 获取重要的局部特征词, 使模型有选择地进行特征提取; 然后在卷积层中增加大小为 1 的卷积核, 提取单一词的特征; 最后该方法对输入文本进行适当的文本填充, 保证每个词都存在上下文信息, 使模型有效提取到每个词的 n-grams 局部特征, 避免卷积处理过程中局部信息的丢失。该模型在 MR5K 和 CR 数据集上进行验证, 较普通卷积神经网络和传统机器学习方法, 在准确率上分别取得 0.5% 和 2% 的提升。

**关键词:** 卷积神经网络; 注意力模型; 情感分类

**中图分类号:** TP391 **文献标识码:** A

### Word Attention-based Convolutional Neural Networks for Sentiment Analysis

WANG Shengyu<sup>1</sup>, ZENG Biqing<sup>1,2</sup>, SHANG Qi<sup>1</sup>, HAN Xuli<sup>1</sup>

(1. School of Computer Science, South China Normal University, Guangzhou, Guangdong 510631, China;

2. School of Software, South China Normal University, Foshan, Guangdong 528225, China)

**Abstract:** Sentiment classification task needs to capture the sentiment features from document and then combines them to construct the document representation. In this paper, we propose the model of Word Attention-based Convolution Neural Networks(WACNN). Compared with CNN, our model takes the document information as input. In detail, we put a word attention layer after the word embedding layer and before the CNN layer. Attention layer enables our model to focus on certain part of the input document and learn weights of each word. We also add a convolve filter with size of 1 in the convolution layer to extract the features of single word. To ensure the existence of context for each word, we pad the input of the convolution layer. This method can be used to extract the n-grams local features of each word effectively, avoiding information loss caused by convolution processing. Compared with traditional CNN and machine learning methods, the accuracy is improved by 0.5% and 2% on MR5K and CR datasets, respectively.

**Key words:** convolutional neural networks; attention model; sentiment classification

## 0 引言

情感分类任务面向用户生成的文本, 分析文本

的情感信息, 最终得到文本的情感倾向: 正向、中性或负向。目前, 情感分析已经广泛应用于诸多领域, 如电子商务平台、电影推荐、舆情分析等。

当前, 情感分类主要有基于规则、机器学习和深

收稿日期: 2017-11-28 定稿日期: 2018-01-12

基金项目: 国家自然科学基金(61503143); 华南师范大学研究生创新计划(2016lkxm59)

度学习这三类方法。基于规则的方法主要借助于情感词或专家构建的特征,但特征构建的代价非常高。基于机器学习的方法将情感分类视为与文档分类或主题分类相似的任务,采用 unigram、bigram 或 one-hot 等语言模型构建篇章特征,使用机器学习算法作为分类器对篇章进行情感倾向性分析。该方法面临篇章特征向量稀疏、维度爆炸、特征提取困难等问题。基于深度学习的方法把神经网络模型参数训练与特征学习的过程并行化,篇章映射成矩阵作为模型的输入,模型学习到篇章的特征表示后,将其输入到分类器,完成情感分类任务。

近年来,随着深度学习在情感分析领域的研究,越来越多的研究者使用深度学习、神经网络模型对文本进行情感分析。卷积神经网络在情感分析任务中的有效性已被证明。篇章通过 word embedding 映射后作为 CNN 的输入,经过卷积层和池化层得到篇章的特征表示。Kim<sup>[1]</sup>利用提前训练的词向量,成功地将 CNN 应用于句子分类任务。Kalchbrenner<sup>[2]</sup>提出一种动态 k-max pooling 的 CNN 用于句子特征学习,取得较好效果。

注意力机制(attention mechanism)目前已经成功应用于自然语言处理与图像处理等多个领域。在情感分析领域,注意力机制的应用,有利于模型更有效地发现和构建特征,使模型在训练时有选择地进行特征提取。Xu K<sup>[3]</sup>将注意力机制应用于解决图片描述生成问题;在机器翻译领域<sup>[4-5]</sup>,注意力机制改进了原有的 encoder-decoder 翻译模型。Yin<sup>[6]</sup>等提出一种基于注意力机制的卷积神经网络,将其用于句子对建模任务中;Wang<sup>[7]</sup>等人利用基于多层注意力机制的卷积神经网络进行句子关系分类;这些方法的成功说明注意力机制与 CNN 结合的有效性。

CNN 虽然拥有强大的学习能力,但在文本特征提取时,数据集未进行预处理,导致输入文本中存在大量非关键词;其次,CNN 无法自动判别输入文本中哪些局部特征词较为重要,而这对情感分析至关重要;再者,CNN 在情感分析中往往采用不同大小的卷积核,但忽略对单一词特征的提取;最后,CNN 在情感分析领域的应用缺乏理论解释,同时卷积操作容易忽略篇章首尾词的上下文信息。因此,本文结合卷积神经网络和注意力机制,以词特征提取为目标,提出词注意力卷积神经网络模型(word attention-based CNN, WACNN)。该模型针对情感分析任务,做出如下三方面工作:①词注意力机制。模

型首先在词嵌入层后构造词注意力层,注意力机制作用在词嵌入层输出的文本特征表示上,以局部窗口的方式,获取窗口中心词的上下文信息,得到每个词的情感权重。②卷积核。卷积层的卷积操作可以视为获取文本的 n-grams 特征,在该层加入大小为 1 的卷积核,获取单一词的特征。③输入填充。借鉴 n-grams 思想,对卷积层的输入进行适当的 0 填充(Padding),使卷积操作可以根据每个词的上下文提取特征。最后在 MR5K 和 CR 数据集上进行实验验证,取得了比普通卷积神经网络和传统机器学习方法更好的效果。

本文的主要贡献如下:

(1) 本文首次结合注意力机制和 CNN 来解决情感分类问题。

(2) 基于词注意力的卷积神经网络模型既能在词嵌入层自动地获取关键特征词,还能在卷积层分别提取词特征和整体的语义特征。

(3) 本文给出卷积在情感分析上的 n-grams 解释,并利用 Padding 方法模拟该特征提取,实验证明了 Padding 方法在文本特征提取上的有效性。

## 1 相关工作

### 1.1 情感分析

情感分析任务针对用户生成的文本,判断用户的观点、情感或态度<sup>[8]</sup>。情感分类作为情感分析的一项基础任务,通过对文本的情感分析获取其情感倾向性。传统的情感分析方法主要借助于机器学习算法和人工特征构建。Pang B<sup>[9]</sup>利用机器学习方法,借助一元词、二元词、词性标注等特征进行电影评论情感分析。Wang 和 Manning<sup>[10]</sup>将支持向量机与朴素贝叶斯方法相结合,在多个数据集上取得较好效果。

在语言模型上,传统方法多使用词袋模型,但该模型无法有效地获取文本的深层次语义特征。词嵌入(word embedding)<sup>[11]</sup>将词或短语映射到一个低维向量空间中,用以表征它的语法和语义特征。大量基于 word embedding 的神经网络模型应用于情感分类任务。Kim<sup>[1]</sup>将 CNN 应用于文本分类;Johnson 等<sup>[12]</sup>提出一种类似于 Kim 的方法,但将高维度的 one-hot 向量作为模型的输入;Socher<sup>[13]</sup>提出递归神经网络对电影评论进行分析;Tang<sup>[14]</sup>使用 CNN 与带有门机制的 RNN 构建层次化的情感分

类模型。

注意力机制方面, Bahdanau<sup>[4]</sup> 提出注意力机制, 并首次应用于翻译领域, Luong<sup>[5]</sup> 则提出了两种注意力方式 global attention 和 local attention。两者将注意力机制在翻译领域的成功应用, 使其迅速成为研究的热点。Zhou<sup>[15]</sup> 结合 LSTM 和注意力机制, 用于特定目标的情感分类, 将特定目标作为模型的关注点, 训练中获得与特定目标关联性强的信息, 分配高权重, 从而有效识别目标的情感倾向。

## 1.2 卷积神经网络模型

本文以 Kim 的卷积神经网络为基础, 卷积神经网络主要包括输入层、词嵌入层、卷积层、池化层和全连接层。假设输入的情感篇章为  $S = \{w_1, w_2, \dots, w_{n-1}, w_n\}$ , 篇章长度为  $n$ , 通过词嵌入层的词嵌入矩阵可得到篇章的 Word2Vec 特征表示  $X$ , 如式(1)所示。

$$X = \{x_1, x_2, \dots, x_n\} \quad (1)$$

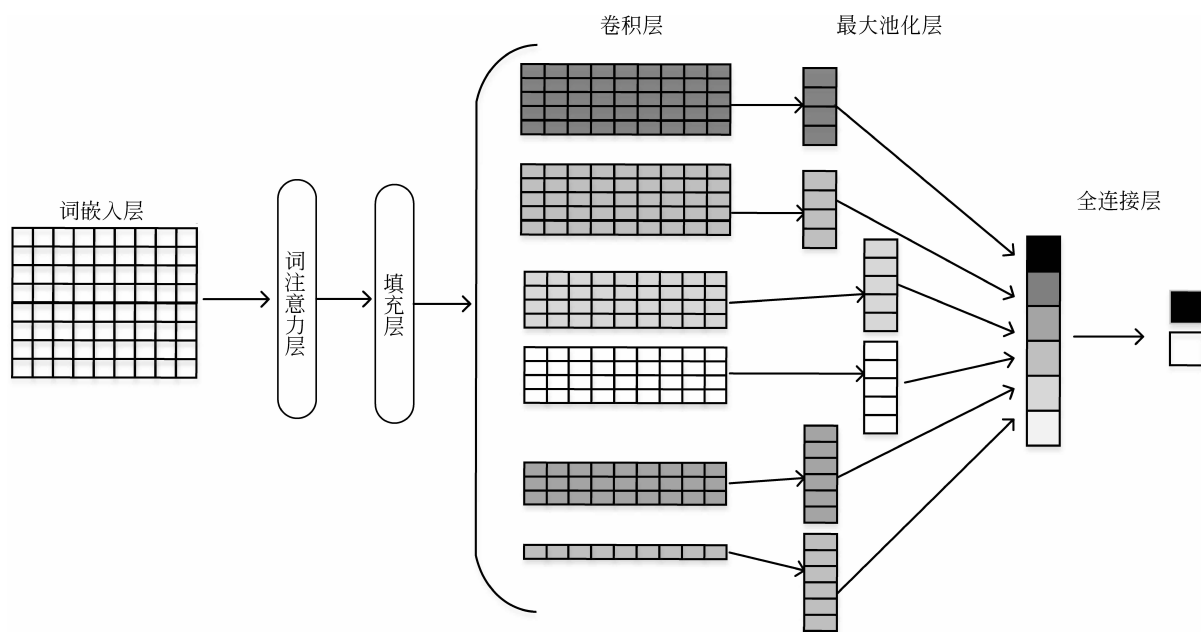


图1 词注意力机制卷积神经网络

(2) 词注意力层: 在词嵌入层后添加注意力层, 使模型在训练的过程中有选择地进行词特征提取, 降低噪声对情感分析的影响。

(3) 填充层: 本文以每个词的上下文含义作为词特征表示, 根据卷积核大小对卷积层的输入进行填充, 保证每个词都存在上下文。

(4) 卷积层: 卷积层除了使用多个大小为  $[3, 4, 5]$  的卷积核, 还增加大小为 1 的卷积核。卷积层

$d$  代表词向量的维度, 其中  $x_i \in R^d$ 。卷积层使用大小不同、数量不等的卷积核对输入矩阵进行卷积操作, 获取每个局部的特征, 生成输入篇章的特征图, 如式(2)所示。

$$C = f(\omega \cdot X + b) \quad (2)$$

其中  $\omega \in R^{h \times d}$  代表卷积核权重矩阵,  $h$  代表卷积核大小,  $b$  代表偏置量,  $f(\cdot)$  代表卷积核函数。得到特征图后, 池化层进行下采样操作, 提取重要的特征。最后, 将经过池化后的卷积层输出串联作为全连接层的输入, 得到最终的分类结果。

## 2 词注意力卷积神经网络模型

如图1所示, 本文提出的词注意力卷积神经网络模型 WACNN(word attention-based CNN)主要分为五个步骤:

(1) 词嵌入层: 将序列化后的输入文本进行词嵌入, 得到文本的特征表示。

对填充后的注意力层输出进行特征提取, 获取局部特征。

(5) 后两部分与 CNN 相似, 包括池化层和全连接层。

通过步骤(2)~(4), 使卷积神经网络在情感分析时, 具备鉴别特征词的能力; 卷积层既能获取每个词的特征, 也能基于每个词的上下文得到局部特征。

## 2.1 词注意力机制

**词嵌入层：**篇章经过序列化处理后得到  $S = \{\omega_1, \omega_2, \dots, \omega_{n-1}, \omega_n\}$ ,  $\omega_i$  代表词在整个数据集词汇表中的下标, 篇章的特征表示采用 one-hot 模型。在词嵌入层采用词映射的方式将每个词映射为  $d$  维的向量, 如式(3)所示。

$$\begin{aligned} x_i &= \omega_i e \\ x_i &\in R^d \end{aligned} \quad (3)$$

$e$  代表词向量矩阵  $e \in R^{|\nu| \times d}$ ,  $|\nu|$  表示整个数据集中词汇的数量。

**词注意力层：**注意力层的作用是对输入文本进行筛选, 决定哪些词汇较为重要, 使模型在训练过程中重点关注。如图 2 所示, 我们采取 n-grams 语言模型, 取该词和其上下文的含义作为该词的语义表达。设置词的上下文范围为  $D$ ,  $p_i$  表示词在文本中的位置, 则词  $\omega_i$  的语义可表示为: 大小为  $L = [p_i - D, p_i + D] = 1 + 2D$  的窗口内上下文的含义。设置滑动窗口矩阵参数  $W_{att} \in R^{L \times d}$ , 计算每个词  $x_i$  的特征值权重  $\alpha_i$ , 用以表示其对情感分析的重要程度。

$$\begin{aligned} X_i^{att} &= \{x_{i-D}, \dots, x_i, \dots, x_{i+D}\} \\ X^{att} &= \{X_1^{att}, \dots, X_n^{att}\} \end{aligned} \quad (4)$$

$$\alpha_i = h(X_i^{att} \otimes W_{att} + b_{att}) \quad (5)$$

$X_i^{att}$  表示以  $x_i$  为中心, 大小为  $L$  的上下文,  $h(\cdot)$  为 sigmoid 激活函数,  $b_{att}$  代表偏置量,  $\otimes$  操作代表滑动窗口覆盖区域内, 参数矩阵  $W_{att}$  与对应位置  $X_i^{att}$  相乘, 并将相乘所得进行求和, 即:

$$X_i^{att} \otimes W_{att} = \sum X_i^{att} * W_{att} \quad (6)$$

$\{X_1^{att}, X_n^{att}\}$  存在缺失上文或下文的问题, 采用 2.2 节的填充方法进行处理。

对于词嵌入层的输出  $X$ , 注意力层采用滑动的方式, 对每个局部提取词的权重  $\alpha_i$ , 得到权重向量  $\alpha$ , 如图 2 所示。

$$\begin{aligned} \alpha &= \{\alpha_1, \dots, \alpha_n\} \\ \alpha &\in R^n \end{aligned} \quad (7)$$

$n$  代表输入句子的长度。

通过  $\alpha$  可以标识每个词的重要程度, 将其与对应位置的词向量相乘得到新的文本特征表示  $X_{att}$ :

$$\begin{aligned} x_i^{att} &= \alpha_i x_i \\ x_i^{att} &\in R^d \\ X_{att} &= \{x_1^{att}, \dots, x_i^{att}, \dots, x_n^{att}\} \end{aligned} \quad (8)$$

## 2.2 填充(Padding)方法

在情感分析领域, 传统自然语言处理常采用

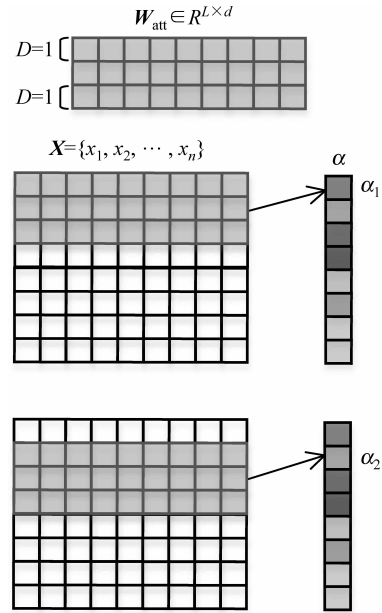


图 2 词注意力层

unigram、bigrams 或 trigrams 作为语言特征。

Pang B<sup>[9]</sup> 曾使用该方法进行情感分析, 此方法基于马尔科夫链, 考虑每个词的上下文信息, 研究结果表明该方法非常有效。卷积层提取局部特征, 可以将卷积操作近似地视为提取文本的 n-grams 特征。每次卷积操作, 卷积核参数  $\omega \in R^{h \times d}$  以  $x_i$  为中心, 考虑词的上下文, 依次提取每个词的 n-grams 特征。以大小为 3 的卷积核为例, 每次考虑中心词的上下文范围为 1, 即遵循隐马尔可夫原则, 而大小为 5 时, 上下文范围就扩大到 2。

但在卷积层中, 处于篇章首尾的词无法被卷积操作覆盖到, 且无上下文信息, 即首尾各存在  $\lfloor h/2 \rfloor$  个词无法提取到 n-grams 特征, 产生信息丢失。如图 3 所示, 我们使用两种方式对卷积层的输入进行适当的填充, 以使卷积操作可以提取到每个词的上下文信息。

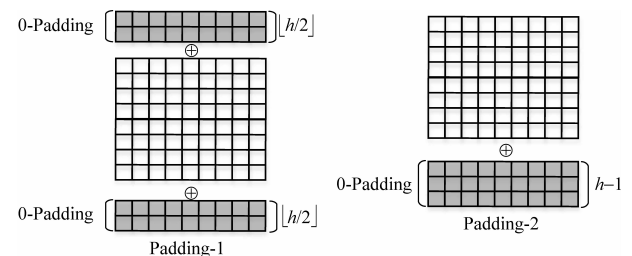


图 3 两种 Padding 方式

(1) Padding-1: 对卷积层的输入在首尾位置进行 0 向量补全, 补全大小为  $\lfloor h/2 \rfloor$ , 以保证开始词存在上文, 结尾词存在下文, 如式(9)所示。

$$X = \underbrace{x_0 \oplus x_0 \oplus \cdots \oplus x_0}_{\lfloor h/2 \rfloor} \oplus X_{\text{att}} \oplus \underbrace{\cdots \oplus x_0 \oplus x_0}_{\lfloor h/2 \rfloor} \quad (9)$$

(2) Padding-2: 在卷积层的输入末尾进行大小为  $h-1$  的 0 向量补全, 如式(10)所示。

$$X = X_{\text{att}} \oplus \underbrace{\cdots \oplus x_0 \oplus x_0}_{h-1} \quad (10)$$

式(10)中,  $\oplus$  表示串联操作。

### 2.3 卷积层和池化层

卷积层负责提取单一词特征和词的上下文信息。经过填充后的文本特征表示  $X_{\text{att}}$ , 将其作为卷积层的输入。卷积层使用大小为  $h$  的卷积核  $\omega \in R^{h \times d}$  对输入进行局部特征提取, 如式(11)所示。

$$c_i = f(\omega \cdot X_{\text{att}, (i: i+h-1)} + b) \quad (11)$$

其中  $X_{\text{att}, (i: i+h-1)}$  代表文本特征矩阵从第  $i$  行到第  $i+h-1$  行的局部特征,  $f(\cdot)$  代表卷积核函数,  $b$  代表偏置量。卷积层对特征矩阵  $X$  的不同位置进行卷积操作, 得到特征图(feature map)  $c$ , 如式(12)所示。

$$c = \{c_1, c_2, \dots, c_{n-h}, c_{n-h+1}\} \quad (12)$$

其中  $c \in R^{n-h+1}$ , 采用最大池化进行下采样, 提取最重要特征  $\hat{c} = \max(c)$ , 卷积核数量为  $m$  个, 得到特征表示  $v$ , 如式(13)所示。

$$v = \{\hat{c}_1, \dots, \hat{c}_m\} \quad (13)$$

### 2.4 全连接层

上述为单一大小的卷积核, 本文使用多个大小的卷积核, 数量为  $z$ , 将池化层的输出串联作为全连接层的输入, 全连接层负责情感分类, 得到结果, 如式(14)所示。

$$O = v_1 \oplus \cdots \oplus v_z$$

$$O \in R^{mz}$$

$$P(y | O, W_{fc}, b_{fc}) = \text{softmax}(W_{fc} \cdot O + b_{fc}) \quad (14)$$

式(14)中,  $\oplus$  表示串联操作,  $O$  代表模型在多个不同大小的卷积核下生成的结果串联,  $z$  表示选取了几种大小的卷积核,  $m$  表示每个卷积核的数量,  $W_{fc}$  为全连接层的权重,  $b_{fc}$  为偏置量,  $y \in \{+1, -1\}$ , 代表情感类别, 此处为二分类。

## 3 实验数据与实验设置

### 3.1 数据集

实验部分在以下两个公开数据集上进行:

(1) CR: 用户在不同商品上的评论数据集, 用于对评论进行情感倾向性分析<sup>[16]</sup>。

(2) MR5K: 用户对电影的评论数据。每一条评论数据包含正向和负向的情感倾向<sup>[17]</sup>。

上述的这些数据集已经人工对每条评论设置了情感标签, 实验中暂不考虑标签与文本是否相符的情况。数据集详细信息如表 1 所示。

表 1 实验数据集

数据集	类别	V	N	average	max
MR5K	2	18 000	10 662	20	56
CR	2	5 340	3 775	19	105

表 1 中  $|V|$  代表数据集词汇数量;  $N$  代表数据集篇章数量; average 代表篇章词数量的平均值; max 代表篇章词数量的最大值。

### 3.2 实验参数与数据处理

本实验按照 8 : 1 : 1 将数据集划分为训练集、验证集和测试集。数据集的预处理采用 keras 自带的 tokenizer 接口进行分词。文本使用预训练的词向量进行映射, 词向量采用 Mikolov<sup>[11]</sup> 提前训练好的公开结果集, 向量维度 300 维<sup>①</sup>。为了降低数据分配对实验结果的影响, 每组实验结果都取十折交叉验证的平均值。

实验中卷积神经网络的参数同 Kim<sup>[1]</sup> 构建的单层卷积神经网络参数相似, 激活函数使用 rectified linear units(RELU), 卷积核大小分别为 [1, 3, 4, 5], 每个卷积核数量为 100 个, Dropout 层接收比例为 0.5, batch\_size 为 64, 池化层采用最大池化, 训练时对词向量进行微调, 采用 Adadelta<sup>[18]</sup> 算法优化模型参数, 词注意力层取上下文范围大小  $D=2$ , 即  $L=5$ , 滑动窗口数量为 1。全连接层隐藏单元 200 个。

### 3.3 对比方法

本文提出的方法与多种不同的方法进行对比, 包括支持向量机(support vector machine, SVM)、卷积神经网络等。对比实验数据来自 Kim<sup>[1]</sup>。

对比方法如下:

**NBSVM, MNB:** 使用朴素贝叶斯和支持向量机进行情感和主题分类<sup>[10]</sup>。

① <https://code.google.com/archive/p/word2vec/>

**CNN-K:** Kim<sup>[1]</sup> 将卷积神经网络应用于篇章分类。

**CNN-Z:** Zhang Y<sup>[19]</sup> 实验观察了卷积神经网络对各参数的敏感性。

**CNN:** 在本实验环境下,使用 CNN 在各数据集上的实验结果。

**WACNN1:** 词注意力卷积神经网络模型,填充层使用 Padding-1 方法。

**WACNN2:** 词注意力卷积神经网络模型,填充层使用 Padding-2 方法。

**WACNN-N:** 模型在填充层不使用 Padding 方法。

### 3.4 实验结果与分析

本文在 MR5K 和 CR 数据集上进行多组实验,得到目标的情感倾向。表 2 给出了实验在不同数据集上得到的分类结果。

表 2 不同模型在特定目标的情感分类结果

模型	MR5K	CR
WACNN1	81.96	<b>86.26</b>
WACNN2	<b>82.02</b>	86.24
WACNN-N	81.81	85.86
CNN	81.49	85.40
CNN-K	81.50	84.30
CNN-Z	81.24	83.92
NBSVM	79.40	81.80
MNB	79.00	80.00

从表 2 可以看出,本文提出的方法在两个领域的数据集上都取得更好的效果。WACNN1/WACNN2/WACNN-N 都取得了比原始 CNN 和传统机器学习方法更好的效果。CNN 模型将所有词都同等对待,提取每一处的局部特征,没有识别关键词的能力,基于注意力机制的 WACNN 模型相比 CNN 情感分类正确率有明显提升。在 CR 数据集上提升了 2%,这说明注意力机制能使 CNN 在训练过程中有选择地关注特定目标词汇,降低噪声影响,从而更好地识别目标情感倾向性。本组实验验证了注意力机制同卷积神经网络结合有助于特定目标的情感分类。

从表 2 还可以看出,在多个数据集上,WACNN-N 模型的表现要比 WACNN1 和 WACNN2 略差。从图 4 可以更显著地看出,WACNN-N 效果优于原始 CNN,但较 WACNN1 和 WACNN2 要差一些。卷积层在获取文本开始和结尾处的局部特征时,缺乏

词的上下文信息,造成信息丢失,填充层的加入可有效提升模型效果,保证卷积层可提取到每个词的 n-grams 特征。实验验证了本文提出的两种填充方法在 WACNN 模型上的有效性。

从图 4 可以看出,两种填充方法的优劣并无绝对差别,在不同的目标数据中表现不一。两种方法哪种更优需要根据特定目标而定。

### 3.5 填充(PADDING)方法

卷积神经网络在卷积层的卷积操作,我们将其理解为依据中心词提取词的 n-grams 特征,并给出了两种填充方法保证卷积操作可以提取到每个词的上下文信息。

表 2 和图 4 验证了填充方法在 WACNN 模型上的有效性,本节在 MR5K 数据集上设计两组实验来验证理论猜想的正确性、填充方法在 CNN 上的有效性。分别考虑多卷积核大小和单一卷积核的 CNN,形成表 3 和表 4 的实验结果。CNN-1 代表使用 Padding-1 方法,CNN-2 代表使用 Padding-2 方法。

从表 3 的实验结果可以看出,伴随着卷积核的变大,卷积层在文本开始和结尾处丢失的信息更多,取得的效果也越差。填充层的加入使输入根据卷积核的大小进行填充,卷积层提取到每一个词的上下文信息,能更好地表达文本的情感信息。CNN-1 和 CNN-2 取得了比原始 CNN 更好的效果。实验验证了填充方法在多卷积核卷积神经网络上的有效性。

表 3 Padding 对多卷积核模型的影响

Region Size	CNN	CNN-1	CNN-2
[3,4,5]	81.49	81.49	<b>81.53</b>
[4,5,6]	81.28	<b>81.73</b>	81.71
[5,6,7]	81.19	81.55	<b>81.73</b>
[7,8,9]	81.13	81.37	<b>81.44</b>
[10,11,12]	80.43	<b>81.16</b>	81.05

表 4 的实验结果表明,在不同大小的卷积核下,填充方法可取得 0.3%~1% 的模型提升效果。填充方法在单卷积核的情况下同样有效,但是两种方法的优劣并没有绝对差别,效果取决于目标数据。

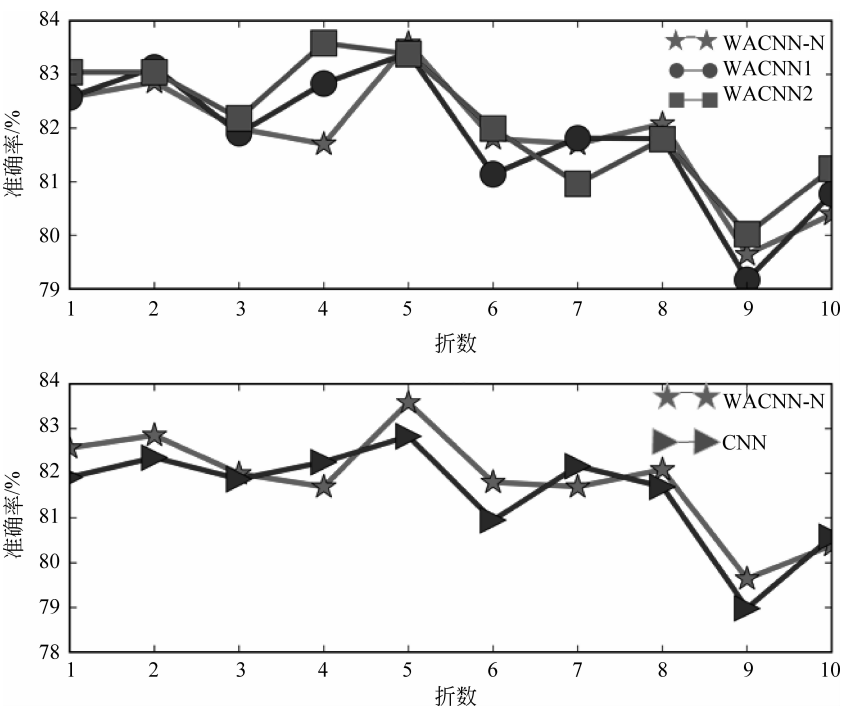


图4 MR数据集上模型10折交叉验证的实验结果

表4 Padding对单卷积核模型的影响

Region Size	CNN	CNN-1	CNN-2
3	80.81	80.89	80.89
4	81.02	81.36	<b>81.38</b>
5	81.23	81.44	<b>81.62</b>
6	80.96	<b>81.42</b>	81.24
7	80.88	<b>81.52</b>	81.09
10	80.44	81.41	81.41
15	80.11	<b>81.14</b>	81.10

3.4节和3.5节的实验表明：①理论的正确性。情感分析任务中，卷积层以每个词为中心，依据词的上下文提取词的n-grams特征。②填充方法的有效性。两种填充方法有效地保证了每个词具有上下文信息，使卷积层可提取到每个词的特征。

3.6 词注意力可视化

情感分析任务中，卷积神经网络取得了不错的效果，但是模型在特征提取、特征选择等环节都缺乏可解释性和可视性。本文在词嵌入层之后卷积层之前添加词注意力层，使得模型在训练过程中有选择地进行特征提取。因此，我们可以对模型的特征提取和选择进行可视化，并给予解释。

本组实验从MR5K和CR数据集中随机各选

出两篇文本，模型的词注意力层得到每个词的权重，根据权重大小加深词的颜色，深色阴影代表最高权重级别，浅色阴影代表较高权重级别。表5和表6中只标识出权重较高的词，其余不做标注。

表5 CR数据集文本注意力机制可视化

数据集：CR	类别：正向
it's <b>very sleek</b> <b>looking with</b> a <b>very good front panel</b> button layout, and it has a <b>great feature set</b> .	
i did not want to have high expectations for this <b>apex</b> player because of the price but it is <b>definitely working out much better</b> than what i would expect <b>from an expensive high-end</b> player.	

表6 MR5K数据集文本注意力机制可视化

数据集：MR5K	类别：负向
the <b>movie's something-borrowed construction</b> feels <b>less</b> the product of <b>loving</b> , well integrated homage and <b>more like</b> a mere <b>excuse for</b> the wan, <b>thinly sketched</b> story. <b>kill-</b> ing time, that's all that's going on here.	
i'm <b>not</b> saying that ice age <b>doesn't have</b> some fairly pretty pictures, <b>but</b> there's not enough substance in the story to <b>actually</b> give them life.	

从表5和表6中可以看出，模型训练时会将形容词语（形容词+名词）和形容词视为较重要的词。例如在CR数据中“very sleek”“good front panel”“thinly sketched”，MR5K数据中“thinly

sketched”“integrated”等。除此之外,否定词转折词、名词短语和比较短语也有很高的权重值。如“much better”“more like”“i’m not”“but”等。

本组实验证明:词注意力层确实可以使模型在训练过程中发现关键词并有选择地提取数据特征。在情感分类任务中,情感词、形容词短语、比较词、否定词和转折词都对情感倾向性判断具有较大影响。

#### 4 结论

本文提出一种基于词注意力的卷积神经网络模型,并将其用于情感分析领域,实验取得了较好效果。模型借鉴传统自然语言处理中的 n-grams 特征,在词嵌入层之后加入注意力层,提升了模型训练时的特征选择能力,并使模型具备特征选择可视化 and 可解释性;填充层的填充方法保证每个词都具备上下文信息;卷积层以词为中心,根据词的上下文提取特征。最后,模型在 MR5K 和 CR 数据集上验证了效果。

然而,本文将注意力机制应用在词特征的选择上,并未在语义层面进行特征筛选。所以接下来将对注意力机制进行改进,使其对语义的特征选择更加有效。

#### 参考文献

- [1] Kim Y. Convolutional neural networks for sentence classification[J]. arXiv preprint arXiv: 1408.5882, 2014.
- [2] Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences[J]. arXiv preprint arXiv: 1404.2188, 2014.
- [3] Xu K, Ba J, Kiros R, et al. Show, attend and tell: Neural image caption generation with visual attention [C]//Proceedings of International Conference on Machine Learning, 2015: 2048-2057.
- [4] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint arXiv: 1409.0473, 2014.
- [5] Luong M T, Pham H, Manning C D. Effective approaches to attention-based neural machine translation[J]. arXiv preprint arXiv: 1508.04025, 2015.
- [6] Yin W, Schütze H, Xiang B, et al. ABCNN: Attention-based convolutional neural network for modeling sentence pairs[J]. arXiv: 1512.05193v4, 2015.
- [7] Wang L, Cao Z, Melo G D, et al. Relation classification via multi-level attention CNNs [C]//Proceedings of Meeting of the Association for Computational Linguistics, 2016: 1298-1307.
- [8] Liu B. Sentiment analysis and opinion mining[J]. Synthesis Lectures on Human Language Technologies, 2012, 5(1): 1-167.
- [9] Pang B, Lee L, Vaithyanathan S. Thumbs up?: Sentiment classification using machine learning techniques [C]//Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing-Volume 10. Association for Computational Linguistics, 2002: 79-86.
- [10] Wang S, Manning C D. Baselines and bigrams: Simple, good sentiment and topic classification [C]//Proceedings of Meeting of the Association for Computational Linguistics: Short Papers. Association for Computational Linguistics, 2012: 90-94.
- [11] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space[J]. arXiv preprint arXiv: 1301.3781, 2013.
- [12] Johnson R, Zhang T. Effective use of word order for text categorization with convolutional neural networks [J]. arXiv preprint arXiv: 1412.1058, 2014.
- [13] Socher R, Perelygin A, Wu J Y, et al. Recursive deep models for semantic compositionality over a sentiment treebank [C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2013, 1631: 1642.
- [14] Tang D, Qin B, Liu T. Document modeling with gated recurrent neural network for sentiment classification [C]//Proceedings of Conference on Empirical Methods in Natural Language Processing, 2015: 1422-1432.
- [15] Zhou X, Wan X, Xiao J. Attention-based LSTM network for cross-lingual sentiment classification [C]//Proceedings of Conference on Empirical Methods in Natural Language Processing, 2016: 247-256.
- [16] Hu M, Liu B. Mining and summarizing customer reviews [C]//Proceedings of Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, Washington, Usa, August. DBLP, 2004: 168-177.
- [17] 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.
- [18] Zeiler M D. ADADELTA: An adaptive learning rate method[J]. arXiv preprint arXiv: 1212.5701, 2012.
- [19] Zhang Y, Wallace B. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification[J]. arXiv preprint arXiv: 1510.03820, 2015.





王盛玉(1992—), 硕士研究生, 主要研究领域为自然语言处理、情感分析。

E-mail: wangshengyu\_1992@163.com



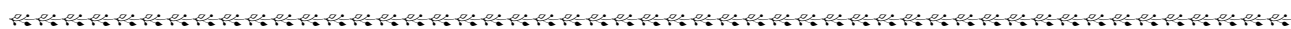
曾碧卿(1969—), 通信作者, 教授, 研究生导师, 主要研究领域为人工智能、机器学习与自然语言处理。

E-mail: zengbiqing0528@163.com



商齐(1992—), 硕士研究生, 主要研究领域为自然语言处理、推荐系统。

E-mail: 1162276945@qq.com



(上接第 92 页)

- [15] Le Q, Mikolov T. Distributed representations of sentences and documents[C]//Proceedings of the 31st International Conference on Machine Learning, 2014: 1188-1196, 4.
- [16] 李武军, 周志华. 大数据哈希学习: 现状与趋势[J]. 科学通报, 2015(60): 485-490, 5.
- [17] Datar M, Immorlica N, Indyk P, et al. Locality-sensitive hashing scheme based on p-stable distributions [C]//Proceedings of the Twentieth Annual Symposium on Computational Geometry, 2004: 253-262, 5.
- [18] Charikar M S. Similarity estimation techniques from rounding algorithms[C]//Proceedings of the Thirty-Fourth Annual ACM Symposium on Theory of Computing, 2002: 380-388, 5.
- [19] Jiang Q Y, Li W J. Scalable graph Hashing with feature transformation[C]//Proceedings of the twenty-fourth international joint conference on artificial intelligence, 2015: 2248-2254, 5.
- [20] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[C]//Proceedings of Advances in neural information processing systems. 2013: 3111-3119, 6.
- [21] Dahl G E, Sainath T N, Hinton G E. Improving deep neural networks for LVCSR using rectified linear units and dropout[C]//Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2013: 8609-8613, 7.
- [22] Duchi J, Hazan E, Singer Y. Adaptive subgradient methods for online learning and stochastic optimization[J]. Journal of Machine Learning Research, 2011, 12: 2121-2159, 7.
- [23] Rehurek R, Sojka P. Software framework for topic modelling with large corpora[C]//Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, 2010: 45-50, 8.



张帆进(1994—), 博士研究生, 主要研究领域为机器学习、数据挖掘。

E-mail: zjf17@mails.tsinghua.edu.cn



顾晓韬(1994—), 博士研究生, 主要研究领域为机器学习、数据挖掘。

E-mail: guxt1994@gmail.com



姚沛然(1997—), 本科生, 主要研究领域为机器学习、数据挖掘。

E-mail: ypr15@mails.tsinghua.edu.cn