

doi: 10.12052/gdutxb.190042

基于双注意力卷积神经网络模型的情感分析研究

曾碧卿¹, 韩旭丽², 王盛玉², 徐如阳², 周 武²

(1. 华南师范大学 软件学院, 广东 佛山 528225; 2. 华南师范大学 计算机学院, 广东 广州 510631)

摘要: 卷积神经网络(Convolutional Neural Networks, CNN)无法判别输入文本中特征词与情感的相关性. 因此提出一种双注意力机制的卷积神经网络模型(Double Attention Convolutional Neural Networks, DACNN), 将词特征与词性特征有效融合后得到本文的特征表示, 确定情感倾向. 本文提出局部注意力的卷积神经网络模型, 改进卷积神经网络的特征提取能力, 采用双通道的局部注意力卷积神经网络提取文本的词特征和词性特征. 然后使用全局注意力为特征分配不同的权重, 有选择地进行特征融合, 最后得到文本的特征表示. 将该模型在MR和SST-1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0.7%和1%的提升.

关键词: 卷积神经网络; 注意力机制; 情感分类

中图分类号: TP391 文献标志码: A 文章编号: 1007-7162(2019)04-0010-08

Sentiment Classification Based on Double Attention Convolutional Neural Network Model

Zeng Bi-qing¹, Han Xu-li², Wang Sheng-yu², Xu Ru-yang², Zhou Wu²

(1. School of Software, South China Normal University, Foshan 528225, China;

2. School of Computer Science, South China Normal University, Guangzhou 510631, China)

Abstract: Convolutional Neural Networks (CNN) cannot discriminate the correlation between feature words and emotions in the input text. Therefore, this paper proposes a double attention mechanism convolutional neural network to incorporate word and POS(Part-of-speech Tagging) information into sentiment classification (Double Attention Convolutional Neural Networks, DACNN), which combines the word features with the part-of-speech features to obtain the feature representation of the text and determine the sentiment tendency. In this paper, we introduce an attention convolutional neural network to incorporate word and POS information into sentiment classification. The first local attention mechanism captures local feature from word and POS channels. And then we propose a novel architect for combining two features by a global attention mechanism, which can decide how much information to use from word or POS information. The model is validated on MR and SST-1 datasets, achieving 0.7% and 1% improvement in accuracy compared to conventional convolutional neural networks and traditional machine learning methods.

Key words: convolutional neural networks; attention mechanism; sentiment classification

情感分类是情感分析的一项基础任务. 情感分析的应用是针对互联网上生成的文本信息分析其情感极性, 当前极性分类有三分类、五分类和十分类, 例如三分类为积极、中性和消极. 目前, 情感分析已经解决了很多实际问题, 例如提取评论情感倾向用于优化电商平台、视频网站以及对选举活动进行舆情分析等.

研究人员通常进行人工构建特征, 以机器学习方法来训练文本分类器. Pang等^[1]根据传统的机器学习方法进行情感倾向性分析, 利用一元词、二元词和词性标注等特征进行实验, 论证了该方法在情感分类任务上是有效的. 随后, 研究学者们试图采用更有效的人工特征来构建文本分类模型, 例如采用情感词汇特征^[2-3]和n-grams词特征^[4-5]构建模型. 但是这些

情感分类方法过多依赖于人工特征,例如采用专家设计的特征或使用额外的情感词汇。

神经网络模型在获取特征表示时无需复杂的特征工程,该思想受到越来越多的研究者的关注^[6-7]。当前,神经网络模型已经在多个自然语言处理领域取得较好效果,例如神经网络机器翻译^[8]、句子对识别^[9]、自动问答^[10]、文本摘要^[11]和序列标注^[12]等。在自然语言处理领域,卷积神经网络在情感分析任务中的有效性已被证明。Kim等^[13]提出用词向量训练文本并用以表示文本信息,在卷积神经网络模型中成功地进行篇章分类。Kalchbrenner等^[14]提出用CNN模型进行句子分类,并利用动态最大池化方法将其融入句子分类模型中取得较好的实验效果。

虽然CNN具有很强的学习能力,但它不能自动确定文本的哪个部分在文本特征提取的过程中更为重要。在进行文本情感分析时,部分文本的情感信息对最终的情感倾向具有较大影响。理想的情感分析模型应该能够有选择地进行特征提取并构建生成句子特征表示。基于此,本文提出一种基于多注意力的卷积神经网络模型,以该模型进行情感分析研究。模型受启发于n-grams语言模型,同时在词和词性2个角度提取文本的特征表示。注意力机制可以使模型在训练的过程中对部分文本加强特征提取,并将得到的篇章特征构建的中间表示^[15]。

目前,大部分的神经网络模型将句子或篇章作为模型输入,经过模型得到情感特征表示。多特征的融合更具特征表示能力,可以得到更优的情感分类效果。但是,当前使用串联的方式无法判别融合多个特征。因此,本文提出一种注意力方式来替代串联方式进行特征融合。在进行多特征融合时,本文使用全局注意力方式来判别词特征和词性特征各自的重要性,对2个特征有选择地构建最终的篇章表示。

本文的主要贡献如下:(1) 本文结合注意力机制和CNN来解决情感分类问题。(2) 本模型中运用了双注意力机制,局部注意力机制可以发现文本中重要的词和词性。随后全局注意力将两类特征有效融合。(3) 相较于对比实验,本文提出的模型在情感分析任务上取得较优的效果,同时,实验证明多特征的融合较单一特征对情感分析更有效。

1 研究基础

随着深度学习的发展,深度学习模型已在多个情感分析数据集评比上均取得较好结果。由于人工构建的特征无法有效表征文本的语法和语义信

息,因此词嵌入表示方法被提出。词嵌入(Word Embedding)^[16-18]的功能是将文本中词映射成一个连续的低维的实质向量矩阵,作用是为了更好地表示文本的语法和语义特征。大多数情感分析研究者利用词嵌入构建深度学习模型的方法进行情感分析研究^[19],Kim等^[13]提出利用卷积神经网络构建多通道的情感分类模型。Zhang等^[20]提出一种卷积神经网络的方法,利用高维度的one-hot向量作为模型的输入。Socher等^[21]利用递归神经网络进行建模,在情感分类任务上取得了不错效果。循环神经网络在序列化文本中具有较强的特征学习能力,在情感分析领域得到很好的应用。Tai等^[22]提出一种树型结构的LSTM模型用于情感分类。使用循环神经网络构建层次化的模型,该模型通过RNN能够更好地提取出篇章的特征,而篇章的特征通过词和句子2个维度刻画,从而达到良好的实验效果^[23-24]。

注意力机制(Attention Mechanism)成功应用于NLP各个领域,在情感分析方向上,注意力机制作用在于发现更有利于模型训练的特征。Xu等^[25]在处理计算机视觉问题时使用注意力机制改善计算机图像描述不准确的问题。注意力机制在机器翻译领域应用比较广泛^[26-27],Bahdanau等^[26]利用注意力机制在原始的encoder-decoder翻译模型上进行改进。Yang等^[28]在层次化模型中融合注意力机制,从词和句子2个角度关注特征提取并取得较好效果。Yin等^[9]利用卷积神经网络模型构建一种情感分类任务模型,并利用注意力机制将模型进行融合,在构建句子分类任务中实验效果良好。基于CNN的层次化模型在融合注意力机制后用于句子关系分类问题,取得较好效果^[29-30]。

2 词注意力卷积神经网络模型

本文提出一种双注意力的卷积神经网络模型(Double Attention Convolutional Neural Networks, DACNN),如图1所示。整个模型包括2个通道:词特征通道和词性特征通道。在每个通道获取到特征表示后,在全局注意力层对两种特征进行特征融合。最后,将生成的篇章特征表示作为全连接层的输入,得到最终的情感倾向。

2.1 局部注意力卷积神经网络模型

本文从词和词性两种特征中提取篇章特征,双通道的机制相同,都采用局部注意力卷积神经网络模型,故本小节仅从词角度介绍模型结构。

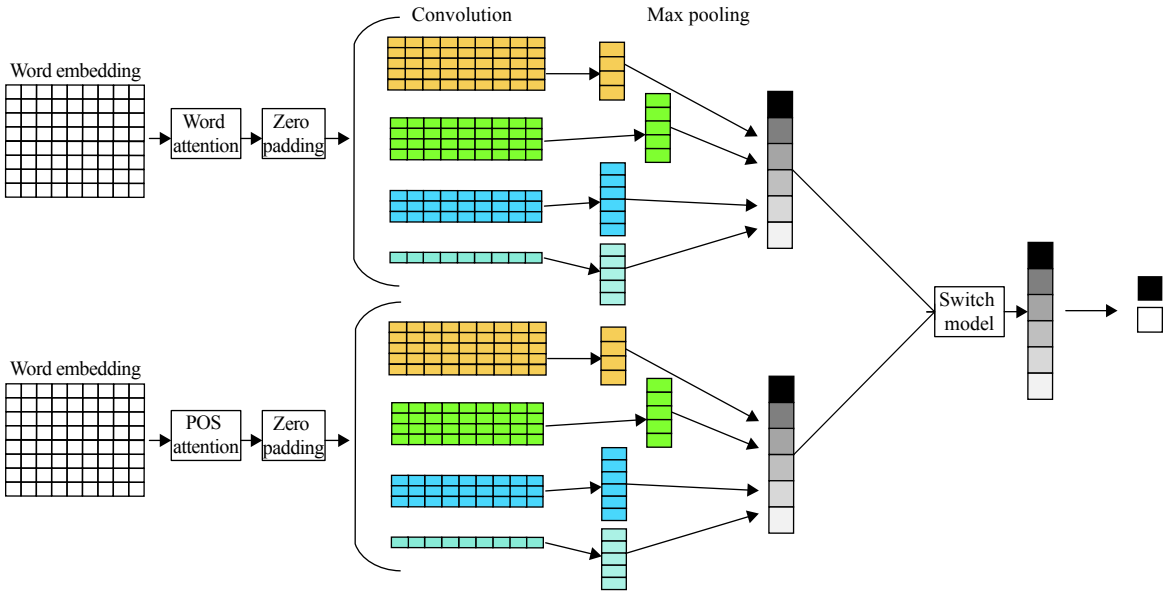


图1 双注意力卷积神经网络模型

Fig.1 The model of double attention convolutional neural networks

如图1所示,本文提出的局部注意力卷积神经网络模型,其主要分为5个层次:(1)词嵌入层:文本进行序列化操作后进行词嵌入表征文本信息,得到特征表示作为模型的输入。(2)词注意力层:在词嵌入层后增加注意力层,本层的作用是让模型可以在文本的训练过程中对文本进行重点关注词特征。(3)填充层:在词注意力层后添加填充层,其作用为文本中每个词特征表示进行上下文填充处理,并根据卷积核大小对卷积层进行填充,以确保文本中所有的词都具备上下文。(4)卷积层:本文使用的卷积层中使用了多个大小不一的卷积核,本文所取卷积核大小分别是1、3、4和5,作用是使本文可以提取不同方面的特征。(5)池化层:对输入特征进行筛选,得到最重要特征。

通过以上的层次,在情感分析中,使模型在训练过程中具有文本词特征的鉴别能力。本文模型训练过程中,卷积层的主要作用在于不仅可以获得文本每个词的特征表示,并且根据文本特征表示获得文本的局部特征。以下对5个层次进行详细介绍。

词嵌入层:文本篇章通过模型进行词特征映射得到特征表示 $S = \{w_1, w_2, \dots, w_{n-1}, w_n\}$, 定义中 w_i 表示文本中词语或者短语在整个文本中的词汇序列中第 i 个词。采用 one-hot 模型进行篇章的特征表示。经过词嵌入层将词语或者短语映射成一个 d 维向量。

$$\mathbf{x}_i = w_i \mathbf{e}, \quad \mathbf{x}_i \in \mathbb{R}^d. \quad (1)$$

式(1)中 $\mathbf{e}$ 表示为词向量矩阵,词嵌入后的矩阵为

$\mathbb{R}^{d \times |v|}$, $|v|$ 代表数据集的整个词汇表的大小。

词注意力层:在词嵌入层后,添加词注意力层,主要作用是对文本特征表示信息进行区分,将重点词特征进行突显,使模型可以有选择地被关注。本文利用了 n -grams 的语言模型,将文本中词语或短语的上下文的词作为该词的一个特征表达。本文用 D 表示一个中心词的上下文范围,词窗口大小为 $L = [p_i - D, p_i + D] = 1 + 2D$ 的词特征作为此词新的表示。 $\mathbf{W}_{\text{att}} \in \mathbb{R}^{L \times d}$ 作为滑动窗口的矩阵参数,在此步骤中需计算每个词 $\mathbf{x}_i$ 的特征值权重 α_i ,用于表示对文本中的词的重要程度信息值。

$$\begin{cases} \mathbf{X}_i^{\text{att}} = \{\mathbf{x}_{i-D}, \dots, \mathbf{x}_i, \dots, \mathbf{x}_{i+D}\}; \\ \mathbf{X}^{\text{att}} = \{\mathbf{X}_1^{\text{att}}, \dots, \mathbf{X}_n^{\text{att}}\}. \end{cases} \quad (2)$$

$$\alpha_i = h(\mathbf{X}_i^{\text{att}} \otimes \mathbf{W}_{\text{att}} + \mathbf{b}_{\text{att}}). \quad (3)$$

其中 $\mathbf{X}_i^{\text{att}}$ 表示以 $\mathbf{x}_i$ 作为中心词,设置中心词上下文大小为 L , $h(\cdot)$ 为 sigmoid 激活函数, $\mathbf{b}_{\text{att}}$ 代表偏置量, $\otimes$ 操作代表两矩阵对应元素相乘后求和,即:

$$\mathbf{X}_i^{\text{att}} \otimes \mathbf{W}_{\text{att}} = \sum \mathbf{X}_i^{\text{att}} \times \mathbf{W}_{\text{att}}. \quad (4)$$

将词注意力层采用平移滑动的方式,进行更深层次地表达,由词嵌入输出 $\mathbf{X}$ 的特征值,每一个输出的局部特征值 $\mathbf{X}$ 得到一个词权重值 α ,如图2所示。

$$\alpha = \{\alpha_1, \dots, \alpha_n\}, \quad \alpha \in \mathbb{R}^n. \quad (5)$$

其中 n 为篇章中句子的长度值。 α 用于表示句子中词或短语的重要程度值,将 α 与每个词特征表示 $\mathbf{x}_i$ 相乘输出新的特征表示 $\mathbf{X}_{\text{att}}$ 。

$$\begin{aligned} X_i^{\text{att}} &= \alpha_i x_i, X_i^{\text{att}} \in \mathbb{R}^d. \\ X_{\text{att}} &= \{X_1^{\text{att}}, \dots, X_i^{\text{att}}, \dots, X_n^{\text{att}}\}. \end{aligned} \quad (6)$$

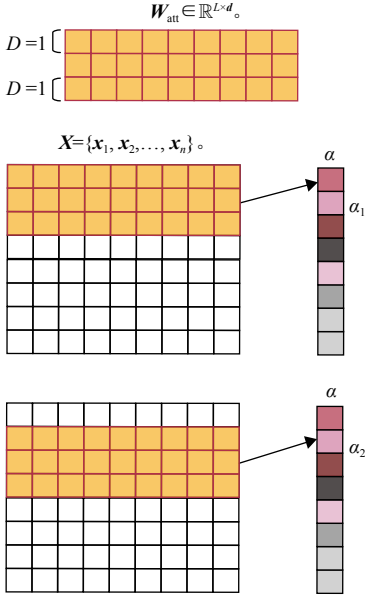


图2 词注意力层

Fig.2 Layer with attention of words

填充层:传统的情感分析处理自然语言的方法一般采用一元语言模型、二元语言模型和三元语言模型提取情感分析特征表示. Pang等^[1]曾使用该方法进行情感分析,并取得良好的实验效果. 卷积神经网络模型利用卷积操作提取局部特征值,它的工作原理在于每进行一次卷积操作,卷积核参数 $\omega \in \mathbb{R}^{h \times d}$ 都将 x_i 为中心,进行上下文窗口大小为 D 的卷积提取特征,视为提取每个词的 n -grams特征. 在本文中,以卷积核大小为3举例,每次取文本每一个中心词的上下文的范围为1,遵循马尔科夫原则,当卷积核大小取5时,上下文范围则扩大为2. 在模型训练中存在文本信息提取不足的问题,在进行卷积操作时,首部词无法取到上文和尾部词无法取到下文情况. 即首部和尾部有 $\lfloor h/2 \rfloor$ 个词无法提取到 n -grams特征的问题. 解决此问题,本文采取首尾填充操作. 主要采用2种填充方法,第一种采用首尾部填充的方法,第二种采用尾部填充的方法,其具体操作如图3所示,其填充方式如下.

(1) **Padding-1:** 在首尾部分进行填充,即对卷积层的输入的首部和尾部分别进行0向量补全,由于首尾部各存在一半窗口的信息缺失,所以补全大小取 $\lfloor h/2 \rfloor$,目的为保证其首部词存在上文,尾部词存在下文.

$$X = \underbrace{X_0 \oplus X_0 \oplus \dots \oplus X_{\text{att}}}_{\lfloor h/2 \rfloor} \oplus \dots \oplus \underbrace{X_0 \oplus X_0}_{\lfloor h/2 \rfloor}. \quad (7)$$

(2) **Padding-2:** 对卷积层的输入末尾部分进行大小为 $h-1$ 的0向量补全.

$$X = X_{\text{att}} \oplus \dots \oplus \underbrace{X_0 \oplus X_0}_{h-1} \quad (8)$$

其中 $\oplus$ 表示串联操作.

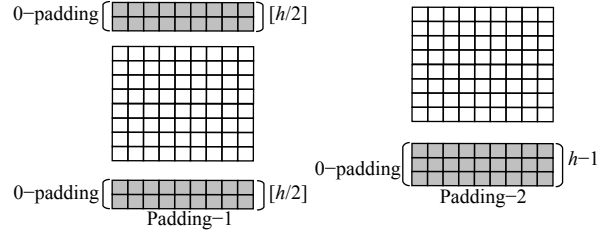


图3 两种padding方式

Fig.3 Methods of the padding

卷积层和池化层:在模型的词注意力层之后添加卷积层和池化层,卷积层的作用主要是将前一步骤提取到的特征进行卷积操作,取多个大小不一的卷积核得到多种特征表示. 如图1所示,卷积层负责提取单一词特征和词的上下文信息. 使用多个大小不同的卷积核得到多种特征图. 例如在本层中加入大小为1的卷积核,用于对次特征进行提取,用 X_{att} 进行表示. 其中取大小为 h 的卷积核 $\omega \in \mathbb{R}^{h \times d}$ 对输入的特征表示信息进行提取局部特征,随后使用最大池化得到特征图的最大特征 c .

$$c = \max(f(\omega \cdot X_{\text{att}} + b)). \quad (9)$$

使用 m 个卷积核对同一位置进行特征提取,得到特征表示 v .

$$v = \{c_1, \dots, c_m\}. \quad (10)$$

$$v \in \mathbb{R}^m. \quad (11)$$

2.2 全局注意力融合层

假设当前在双通道中分别得到篇章的词特征表示 V_W 和词性特征表示 V_{POS} . 一方面,词特征表示和词性特征表示对情感特征表示的贡献度不同,在特征融合时给予的权重也不同. 另一方面,采用传统的串联方式进行特征融合,融合后篇章特征表示维度将倍增. 因此提出一种全局的注意力机制对不同特征进行融合,给重要的特征分配较高权重.

$$u = \tanh(W_{\text{sm}}[v_w, v_{\text{pos}}] + b_{\text{sm}}), u \in [u_w, u_{\text{pos}}]. \quad (12)$$

$$\alpha_t = \frac{\exp(u_t)}{\sum_t \exp(u_t)}, \alpha_t \in [\alpha_w, \alpha_{pos}].$$

(13)

$$s = \alpha_w v_w + \alpha_{pos} v_{pos}.$$

(14)

将局部注意力得到的特征表示 $[v_w, v_{pos}]$ 作为多层感知机(Multi-Layer Perceptron, MLP)的输入,得到输入 u ,其中 W_{sm} 是MLP的权重.对输入 u 经过softmax函数归一化后得到每个特征的权重 α_t .最后,特征表示 $[v_w, v_{pos}]$ 与对应的权重相乘后相加,得到篇章的特征表示 s .

2.3 全连接层

将全局注意力层的输出作为全连接层的输入,全连接层负责获取最终的情感倾向.

$$P = \text{softmax}(W_{fc} \cdot s + b_{fc})$$

(15)

其中softmax为激活函数, W_{fc} 代表全连接层的权重, b_{fc} 代表偏置.

3 实验数据与实验设置

3.1 数据集

实验在以下2个开放数据集上进行:

SST-1: 斯坦福情感树数据集,已划分为训练集/验证集/测试集.

MR: 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.

如表1所示,展示了2个数据的详细信息,对数据集的标签进行了人为处理,所以本文实验可以忽略标签和实验数据集中不相符合的情况.

表 1 实验数据集
Tab.1 The Statistical Data of Datasets

数据集	类别	$ V $	N	average	max
MR	2	18 000	10 662	20	56
SST-1	5	17 000	11 855	18	56

其中 $|V|$ 为词汇总量;average为篇章总量;max为所有包含的最多词的篇章中所含词的数量.

3.2 实验参数与数据处理

下列内容将详细地介绍实验参数设置和数据预处理过程.数据集被划分为训练集、验证集和测试集,所占比例分别为8:1:1.使用Keras提供的tokenizer接口实现自动分词,所以本文将其作为数据预处理工具.词向量采用Mikolov等^[16]提前训练好的公开结果集,向量维度映射为300维(<https://code.google>.

[com/archive/p/word2vec/](https://code.google.com/archive/p/word2vec/)).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,激活函数为sigmoid函数.全局注意力多层感知机权重 $W_{sm} \in \mathbb{R}^{400 \times 1}$.其他参数设置同Kim等^[13]构建情感分析的单层多通道的卷积神经网络模型相类似,卷积核大小分别取 $[1,2,3,4,5]$ 用于卷积层实验.其卷积核数量设置为100,激活函数使用的是Rectified Linear Units(RELU),池化层采用的是Max-pooling进行池化操作.模型训练过程中对超参数进行微调,优化器参数采用Adadelata算法作为模型优化操作.Dropout值设为0.5,batch_size为64.全连接层隐藏单元200个.用十折交叉取平均值作为实验结果.本文代码使用Keras实现并优化,采用Nvidia Tesla K40 GPU加速训练过程.

3.3 对比方法

将前人研究在情感分析上的一些较好的模型方法作为本文的基线模型,将本文的方法与基线模型方法进行对比,对比方法如下.

(1) NBSVM, MNB: 朴素贝叶斯和SVM用于对情绪和主题进行分类^[30]. (2) CNN-K: Kim等^[13]构建的情感分类的多通道单层卷积神经网络模型. (3) CNN-Z: Zhang等^[20]通过对深度学习进行构建模型并总结了卷积神经网络对各种参数的分类效果. (4) CNN-A: 词嵌入和注意机制联合应用卷积神经网络的情感分析方法^[31]. (5) CNN: 在本文的实验环境下,使用CNN在各数据集上的实验. (6) Paragraph Vector: Le and Mikolov等^[32]提出的一种文档分布式情感分析方法. (7) Sent-Parser: 利用统计分析架构方法进行情感分析^[33]. (8) DACNN1: 其他设置保持不变,在填充层上用padding-1方法在本文的双注意力卷积神经网络模型进行实验. (9) DACNN2: 其他设置保持不变,在填充层上用padding-2在本文的双注意力卷积神经网络模型进行实验. (10) DACNN-N: 本文的双注意力卷积神经网络模型中对填充层不使用任何padding方法.

3.4 实验结果与分析

本文在两组数据集MR和SST-1上进行多组不同的实验,得到目标的情感倾向准确率值.表2展示了通过对不同数据集的实验获得的分类结果.其中加粗数字为最好的结果.

根据表2中的结果可以看出,本文提出的方法在两个领域的数据集上达到的实验准确率效果最优.较传统的机器学习方法和原始CNN相比,DACNN1、DACNN2、DACNN-N在情感分析任务中取得较好的

实验效果. CNN模型对所有单词一视同仁,从每一处提取局部特征,无法识别关键词. 与一般的神经网络模型相比,DACNN可以并行提取原始文本的特征和词性信息. 与CNN模型相比,基于注意机制的DACNN模型在情感分类准确率上有显著提高,与CNN-K和CNN-A相比,DACNN在MR5K和SST-1数据集上分别提高了近0.7%和1%. 表明将文本的词特征与词性特征的有效融合,可以充分表达原文的情感信息,更有利于分析文本的情感倾向性. 并且局部注意力有效地捕捉到词语和词性信息中有关情感的重要信息,而全局注意力则有效地融合了这两种特征,从而已达到了预期的效果.

表 2 情感分析模型的实验结果

Tab. 2 The results in the models of sentiment analysis %

模型	MR	SST-1
DACNN1	82.07	49.77
DACNN2	82.17	49.41
DACNN-N	81.96	49.14
CNN-K	81.5	48.0
CNN-Z	81.24	48.08
CNN-A	—	48.8
DCNN	—	48.5
Paragraph Vector	—	48.7
Sent-Parser	79.4	—
NBSVM	79.4	—
MNB	79.0	—

从表2可以看出,DACNN-N模型在多个数据集上的性能略低于DACNN1和DACNN2. 根据图4显示,DACNN-N性能较原始CNN得到改进,但与DACNN1、DACNN2两个模型进行比较,性能较差. 模型在训练过程中,在卷积层对文本进行特征进一步提取,添加填充层防止信息丢失的问题,提高实验的效果.

图4显示不同填充方法的优劣是不确定的,其效果在不同的目标数据中表现不同. 具体哪一种更好取决于具体的目标.

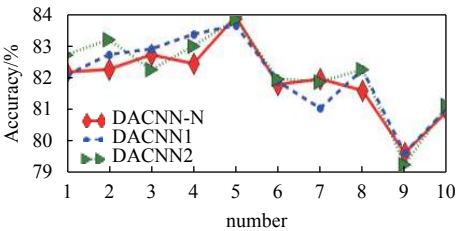


图 4 MR数据集上模型上十折交叉验证的实验结果

Fig.4 10-fold cross validation experiment on MR dataset

3.5 填充(PADDING)方法

了解卷积层中卷积神经网络的卷积运算,根据中心词提取单词的n-grams特征,本文利用2种不同的填充方式,确保模型在卷积运算过程中能够提取到篇章中所有的上下文信息.

如表2和图4所示,在MR和SST-1数据集上模型证明了该方法在改进DACNN模型的效果. 在本节中,通过在MR数据集上的两组实验来验证提出的理论,推测正确性和CNN上填充方法的有效性.

进行更深层次实验验证DACNN模型的有效性. 如表3和表4所示,验证了模型的理论预测正确性和CNN上填充方法有效性. 其中CNN-1代表使用padding-1的方法,CNN-2代表使用padding-2的方法. 将单个卷积核和多个卷积核分别进行实验. 从表3可以看出,随着卷积核的增加,效果越差,原因是卷积层在文本的开头和结尾丢失的信息与卷积核的数量有关,卷积核数量越多,丢失信息越多. 本文通过添加填充层,可以更好地表达文本的情感信息. 根据实验结果,CNN-1和CNN-2均取得了比原CNN更好的效果,该填充方法在多卷积核的卷积神经网络中是同样有效的. 从表4中的实验结果可知,使用不同大小的卷积核,填充方法可以将模型效果提高0.3%~1%. 由此可得,该方法不仅适用于DACNN模型,而且适用于传统的卷积神经网络模型.

表 3 Padding对多卷积核模型的影响

Tab. 3 Effect of filter region size with several region sizes %

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

表 4 Padding对单卷积核模型的影响

Tab. 4 Effect of single filter region size %

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

3.6 全局注意力融合层

从表2可以看出整体模型的效果要优于单一特

征的模型,为进一步验证全局注意力层的特征融合效果,设计本组实验,如图5所示. 为排除局部注意力的影响,实验在简单CNN上进行.

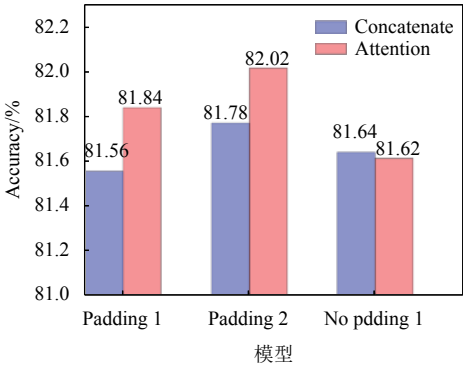


图 5 全局注意力层特征融合效果对比
Fig.5 Effect of the Global Attention for Integrating

3.7 词注意力可视化

为了帮助更好地理解词注意力对模型的影响,此部分对特征提取进行词注意力可视化. 如图6和图7所示. 可视化实验数据由本文的2个实验数据集中随机抽取的2条文本. 其中图6表示取正向积极情感的文本数据,图7取负向消极情感的数据文本. 在词注意力层获取词特征表示的权重值. 用不同颜色深度给文本进行标注,其中实验过程中颜色深的得到权重比较大. 浅颜色表示权重比较小. 为了便于视觉查看,其余比较低的权重词特征表示未进行标注. 如图6为SST-1中随机抽取的数据文本,可以看出词注意力确实具有将关键词重点关注的作用.

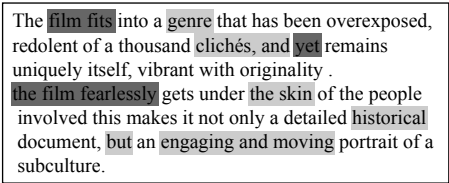


图 6 SST-1数据集文本注意力机制可视化(正向)
Fig.6 Visualization of review text with highlights on the SST-1 datasets

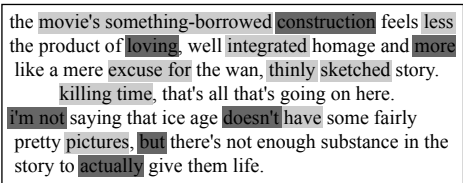


图 7 MR数据集文本注意力机制可视化(负向)
Fig.7 Visualization of review text with highlights on the MR datasets

4 结论

通过对情感分析研究,提出一种基于双注意力的卷积神经网络模型,实验取得了较好效果. 首先,模型从词和词性个通道提取特征. 其次,采用局部注意力机制的方式使模型在训练过程中进行特征词提取,提高了模型的词特征选择能力. 最后对注意力机制层进行可视化分析,充分证明了本文在模型中添加的词注意力机制对情感分析实验是有帮助的. 最后,填充层的填充方法保证每个词都具备上下文信息,篇章的特征表示由全局注意力层进行特征融合. 模型在MR和SST-1数据集上得到验证.

参考文献:

[1] 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. Stroudsburg: ACM. 2002: 79-86.

[2] PEREZ-ROSAZ V, BANEJA C, MIHALCEA R. Learning sentiment lexicons in spanish[C]//LREC. Istanbul: ELRA. 2012, 12: 73.

[3] KIRITCHENKO S, ZHU X, MOHAMMAD S M. Sentiment analysis of short informal text [J]. *Journal of Artificial Intelligence Research*, 2014, 50: 723-762.

[4] WANG S, MANNING C D. Baselines and bigrams: simple, good sentiment and topic classification[C]// Meeting of the Association for Computational Linguistics: Short Papers. Jeju Island: ACL, 2012: 90-94.

[5] ZHANG L, WANG S, LIU B. Deep learning for sentiment analysis: a survey [J]. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2018, 8(4): e1253.

[6] 张玥杰, 徐智婷, 薛向阳. 融合多特征的最大熵汉语命名实体识别模型[J]. *计算机研究与发展*, 2008, 45(6): 1004-1010.

ZHANG Y J, XU Z T, XUE X Y. Fusion of multiple features for chinese named entity recognition based on maximum entropy model [J]. *Journal of Computer Research and Development*, 2008, 45(6): 1004-1010.

[7] CHEN P, SUN Z, BING L, et al. Recurrent attention network on memory for aspect sentiment analysis[C]//Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen: ACL, 2017: 452-461.

[8] WU Y, SCHUSTER M, CHEN Z, et al. Google's neural machine translation system: Bridging the gap between human and machine translation[J]. *arXiv preprint arXiv: 1609.08144*, 2016.

[9] YIN W, SCHÜTZE H, XIANG B, et al. Abcnn: attention-based convolutional neural network for modeling sentence pairs [J]. *Transactions of the Association for Computational*

Linguistics, 2016, 4: 259-272.

- [10] GOLUB D, HE X. Character-level question answering with attention[J]. arXiv preprint arXiv: 1604.00727, 2016
- [11] RUSH A M, CHOPRA S, WESTON J. A neural attention model for abstractive sentence summarization[J]. arXiv preprint arXiv: 1509.00685, 2015.
- [12] REI M, CRICHTON G K O, PYYSALO S. Attending to characters in neural sequence labeling models[J]. arXiv preprint arXiv: 1611.04361, 2016.
- [13] KIM Y. Convolutional neural networks for sentence classification[J]. arXiv preprint arXiv: 1408.5882, 2014
- [14] KALCHBRENNER N, GREFFENSTETTE E, BLUNSON P. A convolutional neural network for modelling sentences[J]. arXiv preprint arXiv: 1404.2188, 2014.
- [15] VASWANI A, SHAZEER N, PARMAR N, *et al.* Attention is All you Need [J]. Neural information processing systems, 2017: 5998-6008.
- [16] MIKOLOV T, SUTSKEVER I, CHEN K, *et al.* Distributed representations of words and phrases and their compositionality[C]//Advances in Neural Information Processing Systems. Lake tahoe Nevada: NIPS, 2013: 3111-3119.
- [17] BENGIO Y, DUCHARME R, VINCENT P, *et al.* A neural probabilistic language model [J]. Journal of machine learning research, 2003, 3(6): 1137-1155.
- [18] MIKOLOV T, CHEN K, CORRADO G, *et al.* Efficient estimation of word representations in vector space[J]. arXiv preprint arXiv: 1301.3781, 2013.
- [19] 梁斌, 刘全, 徐进, 等. 基于多注意力卷积神经网络的特定目标情感分析[J]. 计算机研究与发展, 2017, 54(8): 1724-1735.
LIANG B, LIU Q, XU J, *et al.* Aspect-based sentiment analysis based on multi-attention CNN [J]. Journal of Computer Research and Development, 2017, 54(8): 1724-1735.
- [20] 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.
- [21] SOCHER R, PERELYGIN A, WU J, *et al.* Recursive deep models for semantic compositionality over a sentiment treebank[C]//Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. Seattle, Washington: ACL, 2013: 1631-1642.
- [22] TAI K S, SOCHER R, MANNING C D. Improved semantic representations from tree-structured long short-term memory networks[J]. arXiv preprint arXiv: 1503.00075, 2015.
- [23] TANG D, QIN B, LIU T. Document modeling with gated recurrent neural network for sentiment classification[C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal: ACL, 2015: 1422-1432.
- [24] BHATIA P, JI Y, EISENSTEIN J. Better document-level sentiment analysis from rst discourse parsing[J]. arXiv preprint arXiv: 1509.01599, 2015.
- [25] XU K, BA J, KIROS R, *et al.* Show, attend and tell: Neural image caption generation with visual attention[C]//International Conference on Machine Learning. Lille, France: ACM, 2015: 2048-2057.
- [26] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint arXiv: 1409.0473, 2014.
- [27] LUONG M T, PHAM H, MANNING C D. Effective approaches to attention-based neural machine translation[J]. arXiv preprint arXiv: 1508.04025, 2015.
- [28] YANG Z, YANG D, DYER C, *et al.* Hierarchical attention networks for document classification[C]//Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Diego, California: ACL. 2016: 1480-1489.
- [29] WANG L, CAO Z, MELO G D, *et al.* Relation classification via multi-level attention CNNs[C]// Meeting of the Association for Computational Linguistics. Berlin: ACL, 2016: 1298-1307.
- [30] WANG S, MANNING C D. Baselines and bigrams: simple, good sentiment and topic classification[C]// Meeting of the Association for Computational Linguistics: Short Papers. Association for Computational Linguistics, Jeju Island: ACL, 2012: 90-94.
- [31] SHIN B, LEE T, CHOI J D, *et al.* Lexicon integrated CNN models with attention for sentiment analysis [J]. Empirical Methods in Natural Language Processing, 2017: 149-158.
- [32] LE Q V, MIKOLOV T. Distributed representations of sentences and documents [J]. International Conference on Machine Learning, 2014: 1188-1196.
- [33] DONG L, WEI F, LIU S, *et al.* A statistical parsing framework for sentiment classification [J]. Computational Linguistics, 2014, 41(2): 293-336.