



## 基于 AT-DPCNN 模型的情感分析研究

高伟军, 杨 杰, 张春霞, 师 阳

(兰州理工大学 计算机与通信学院, 兰州 730050)

**摘 要:** 情感分析是自然语言处理领域的一个重要分支, 卷积神经网络(CNN)在文本情感分析方面取得了较好的效果, 但其未充分提取文本信息中的关键情感信息。为此, 建立一种基于注意力机制的深度学习模型 AT-DPCNN。利用注意力矩阵重点关注文本序列中对情感走向影响较大的部分, 通过对提取到的注意力特征矩阵与原文本词向量进行运算得到注意力输入矩阵, 并利用 CNN 再次提取文本特征。同时为了更好地提取转折等复杂句式的特征, 在池化层进行分池操作。在多个不同类型数据集上的测试结果表明, 该模型具有较高的泛化性能, 处理转折等复杂句式时其分类准确率和 F1 值相对 WACNN、HAN 等模型均有明显提升。

**关键词:** 深度学习; 卷积神经网络; 情感分析; 注意力机制; 损失函数

开放科学(资源服务)标志码(OSID):



中文引用格式: 高伟军, 杨杰, 张春霞, 等. 基于 AT-DPCNN 模型的情感分析研究[J]. 计算机工程, 2020, 46(11): 53-60.

英文引用格式: GAO Weijun, YANG Jie, ZHANG Chunxia, et al. Sentiment analysis research based on AT-DPCNN model[J]. Computer Engineering, 2020, 46(11): 53-60.

## Sentiment Analysis Research Based on AT-DPCNN Model

GAO Weijun, YANG Jie, ZHANG Chunxia, SHI Yang

(School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China)

**[Abstract]** Sentiment analysis is a fundamental field of Natural Language Processing (NLP). Convolutional Neural Network(CNN) performs well when applied to this field, but fails to fully extract the key sentiment information in the text information. To address the problem, this paper proposes a deep learning model, AT-DPCNN, which is based on attention mechanism. The model uses the attention matrix to highly focus on the part of the text sequence that has significant influence on emotion tendency, and operates on the extracted attention feature matrix and the word vector of the original text to get the attention input matrix. Then the CNN is used to re-extract text features. Also, in order to better extract the features of complex sentence patterns such as transition, the divide pooling is performed at the pooling layer. The proposed model is tested on different types of datasets, and the experimental results show that the model has good generalization performance, and significantly improves the classification accuracy and F1 score compared with WACNN, HAN and other models when processing complex sentence patterns such as transition.

**[Key words]** deep learning; Convolutional Neural Network(CNN); sentiment analysis; attention mechanism; loss function

DOI: 10.19678/j.issn.1000-3428.0055950

### 0 概述

随着 5G 时代的到来以及社交媒体的快速发展, 大量用户通过博客、微博、Twitter 和淘宝等来评论热点事件、抒写影评观点以及描述产品体验等, 从而产生了大规模带有情感倾向的文本信息。文本信息爆炸式的增长使得通过文本数据发掘潜在信息成为自然语言处理领域的热点问题之一<sup>[1]</sup>, 其中, 对海量文本信息进行准

确的情感分析是文本数据发掘中的关键。

情感分析通常被分为 2 种类型, 即基于情感词典的分类方法和基于机器学习的情感分类方法<sup>[2]</sup>。基于情感词典的分类方法依赖于情感词典的构建, 情感词典包含大量的情感词, 利用对应的情感词对分类文本进行标注<sup>[3]</sup>。从中国古代“七情(喜、怒、哀、惧、爱、恶、欲)说”对情绪的划分, 到西方学者 PLUTCHIK 从强度、相似性和两极性三方面进行情

基金项目: 国家自然科学基金(61762059); 甘肃省引导创新发展项目(052004)。

作者简介: 高伟军(1973—), 男, 副教授, 主研方向为高性能计算、自然语言处理; 杨 杰、张春霞、师 阳, 硕士研究生。

收稿日期: 2019-09-08 修回日期: 2019-10-29 E-mail: 1043589204@qq.com

绪划分,情感词典一直在不断的发展完善。情感词典的构建目前主要有 4 种方法,分别为基于启发式规则的方法、基于词对齐的方法、基于图的方法和基于表示学习的方法<sup>[4]</sup>。然而,随着网络用语的快速增加,情感词典的构建和更新难度大幅提高<sup>[5]</sup>,此外,文本存在二义性、前褒后贬或者前贬后褒等特点,导致情感词典的泛化能力难以提高。因此,学者们开始探究效果更好的情感分类方法。文献[6]较早在电影评论的情感分析问题上应用最大熵、SVM 和朴素贝叶斯 3 种机器学习方法。大量的实验结果表明,机器学习在文本分类中能够取得较好的效果,因此,基于机器学习的方法是目文本分类的主流方法。

本文建立一种基于注意力机制的深度学习模型 AT-DPCNN,利用注意力矩阵重点关注文本序列中对情感走向影响较大的部分,将提取的注意力特征矩阵与原词向量相结合进行运算,形成注意力输入矩阵,再通过卷积神经网络二次提取特征,将提取的信息特征输入到全连接层得到最终的结果。

## 1 相关工作

文献[7]较早进行情感分析研究,随后众多研究人员的参与使得情感分类取得了较好的效果。词向量的研究为自然语言处理开辟了新路径。文献[8-9]提出 Word2Vec 模型,该模型包含 CBOW 和 Skip-gram 两种框架,使单词映射到低维度连续向量空间中,大幅提升了词向量模型的效率,随后 Word2Vec 被广泛应用于自然语言处理的多个领域。文献[10]采用 CNN 模型对文本序列进行分类,设置不同参数并对比不同的词向量构造方法。

文献[11]提出神经概率语言模型,使用注意力机制建立词与词之间的关系,结合神经网络后应用于机器翻译中,其取得了较好的效果,验证了注意力机制在自然语言处理中的有效性。随后,注意力机制被广泛应用于基于深度学习的自然语言处理任务中。文献[12]针对文本序列的不同位置提出 global attention 和 local attention 两种注意力方式,其系统阐述了自然语言处理中的注意力机制。文献[13]提出了结合位置注意力机制与循环神经网络的方法,以解决特定主题下的情感分类问题。文献[14]在深度 LSTM 模型中加入注意力机制,在 SemEval-2017 情感分类任务中取得了较好的效果。文献[15]使用不同的注意力机制关注文本序列的不同部分,并应用于主题情感分析任务。文献[16]在 BiLSTM 的基础上加入 2 层注意力机制,分别针对词语级和句子级进行特征提取。文献[17]分析句子对建模任务,提出一种基于注意力机制的卷积神经网络。

相比传统文本情感分析模型,上述研究在情感分析方面取得了较好的效果,但受文本复杂性的影响,多数模型仍存在泛化能力不足、特征提取不全面等问题,还需进一步优化。本文将改进的注意力机制与卷积神经网络相结合,高度关注文本序列中对情感走向影响较大的部分,并针对转折等复杂句式进行分池操作,在此基础上,建立一种新的情感分析模型 AT-DPCNN。

## 2 网络模型

### 2.1 词向量表示

在自然语言处理领域,传统统计特征包含的信息量过少,限制了深度学习在自然语言处理中的应用。词向量由于包含了更丰富的信息,深度学习借助词向量可以提取大量的文本特征,从而更好地处理大部分 NLP 问题。经典的 one-hot 编码假设词典中不同词的数量为  $N$ ,每个词均可以和从 0 到  $N-1$  的连续整数相对应。假设一个词的对应整数为  $i$ ,为了得到该词的 one-hot 向量表示,创建一个元素全为 0、长为  $N$  的向量,并将其第  $i$  位设为 1。这种表示方法有 2 个缺点:一是由于深度网络模型的词表相对于传统的词表非常大,一般都在百万级别以上,因此利用 one-hot 处理文本很容易受维数灾难的困扰;二是其默认 2 个词之间是孤立的,任意 2 个词向量内积为 0,简单的如“老师”和“学生”这样的词之间没有语义关系。

为解决 one-hot 编码存在的“语义鸿沟”问题,本文选择 Google 推出的 NLP 词向量工具 Word2Vec。Word2Vec 模型由 MIKOLOV 等人提出,基于当代语言学中的重要定理——“距离相似性定理”,该定理假设词语之间的相似性取决于其相邻词汇是否相识。Word2Vec 通过高维向量表示词语,相近词语放在相近位置,映射出序列局部间数据的关联关系,弥补了 one-hot 编码的不足。通过 Word2Vec 训练得到的词向量具有高维性,解决了词语多方向发散的问题,保证了模型的稳定性,因此,Word2Vec 适合处理序列数据。Word2Vec 作为一种层次较浅的神经网络,其包括如下 2 种模型:

1) 连续词袋(CBOW)模型,如图 1 所示,其通过上下文预测当前词。由背景词生成任一中心词的概率如下:

$$\prod_{i=1}^T P(\mathbf{w}^{(i)} | \mathbf{w}^{(i-m)}, \mathbf{w}^{(i-m+1)}, \dots, \mathbf{w}^{(i-1)}, \mathbf{w}^{(i+1)}, \dots, \mathbf{w}^{(i+m)}) \quad (1)$$

其中,  $T$  为文本序列的长度,  $\mathbf{w}^{(i)}$  为预测的中心词,即时间步  $i$  的对应词,  $\mathbf{w}^{(i-m)}$ 、 $\mathbf{w}^{(i+m)}$  分别为中心词的上下文背景词。

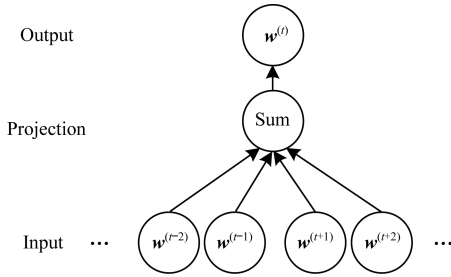


图 1 连续词袋模型结构

Fig. 1 Structure of continuous bag of words model

2) 跳字 (Skip-gram) 模型, 其通过当前词预测上下文。给定中心词生成背景词的概率如下:

$$\prod_{t=1}^T \prod_{-m \leq j \leq m, j \neq 0} P(w^{(t+j)} | w^{(t)}) \quad (2)$$

文本数据无法通过 CNN 直接处理, 必须通过词向量将每一个句子转化为矩阵的形式。长度为  $n$  个单词、 $k$  个维度的文本序列表示为:

$$X_{0:n-1} = X_0 \oplus X_1 \oplus \cdots \oplus X_i \oplus \cdots \oplus X_{n-1} \quad (3)$$

其中,  $X_i \in \mathbf{R}^d$  ( $i = \{0, 1, \dots, n-1\}$ ) 是  $K$  个维度词向量的文本序列中第  $i$  个单词的向量表示。

## 2.2 卷积神经网络

卷积神经网络是一种改进的前馈神经网络, 其避免了输入数据的复杂预处理过程, 可以直接对输入数据进行处理, 实现端对端的表示。卷积神经网络改变了输入层和隐藏层之间采用全连接结构的方式, 其采用稀疏连接, 每一层的神经元响应一部分范围内的周边单元, 即每一个卷积层节点只与一个固定大小的区域相连接, 连接的权重矩阵称为卷积核。这种方式类似于生物神经网络, 以局部连接和共享权值的方式调整参数, 既没有遗漏过多的维度特征,

又减少了权重规模并降低了网络的复杂度, 使整个网络不断地进行并行学习。在文本数据中, 情感十分强烈的词语往往都依赖于同一个主语或者相关关注点, 反映在网络层中即为“共享权重值”, 这种共享机制使得卷积层具有变换等价性, 即输入发生变化时, 输出也随之变化。从函数角度来看, 卷积通过 2 个函数  $f$  和  $g$  生成第 3 个函数的一种算子, 其中, 共享机制中的卷积函数  $f$  和  $g$  等价, 卷积公式如下:

$$h(x) = f(x) * g(x) = \int_{-\infty}^{+\infty} f(t) g(x-t) dt \quad (4)$$

卷积神经网络作为神经网络中的一种, 是多层的监督学习神经网络, 每一层有多个特征图, 每个特征图通过一种卷积滤波器提取输入的一种特征, 每个特征图有多个神经元。该网络模型采用梯度下降法最小化损失函数, 对网络中的权重参数逐层反向调节, 通过频繁的迭代训练提高网络的精度。卷积神经网络的低隐层是由卷积层和池化层交替组成, 高层是全连接层, 对应传统多层感知器的隐含层和逻辑回归分类器。第一个全连接层的输入是由卷积层和子采样层进行特征提取得到的特征图像, 最后一层输出层是一个分类器, 输入层为长度为  $n, k$  维的文本序列矩阵  $X \in \mathbf{R}^{n \times d}$ 。

## 2.3 AT-DPCNN 模型

AT-DPCNN 模型将卷积神经网络共享权值、多特征图等特点与注意力机制关注局部内容的特点相融合。利用 Word2Vec 对文本信息进行处理得到词向量, 经过注意力矩阵计算得到注意力词向量, 再将两者相结合输入到卷积神经网络, 经过注意力加权、分池操作, 最终通过全连接层后利用分类器进行分类。AT-DPCNN 模型结构如图 2 所示。

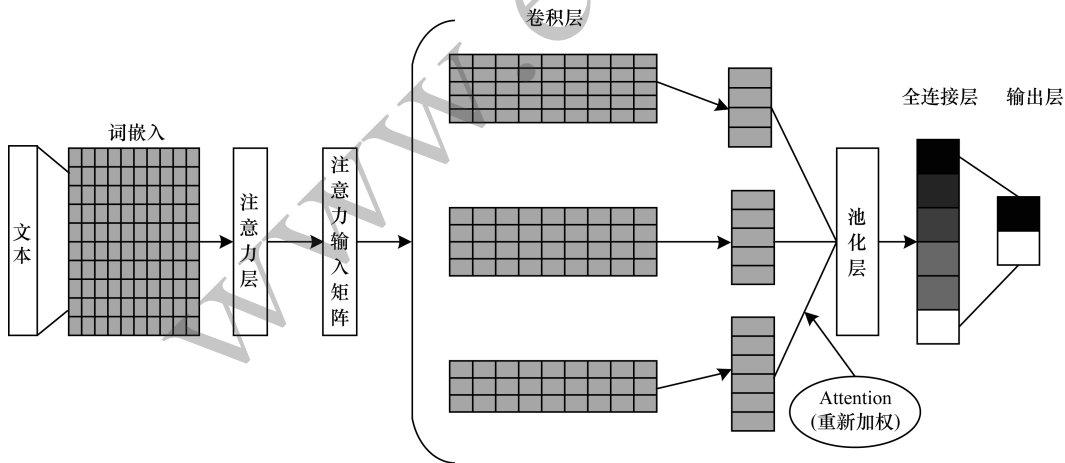


图 2 AT-DPCNN 模型框架

Fig. 2 AT-DPCNN model framework

### 2.3.1 注意力层

传统神经网络在处理数据时往往会忽略原有数据之间的联系, 特别是处理文本信息时, 当输入句子

较长时, 中间向量难以存储足够的信息, 从而限制了模型的性能。在自然语言处理过程中, 注意力机制允许模型根据输入文本以及其到目前为止已经生成的

隐藏状态来学习要注意的内容,不同于标准神经网络将全部原文本编码成固定长度的不含语义的向量,注意力机制将研究聚焦到特定区域,很好地解决了中间向量存储信息不足的问题。该机制灵活地为神经网络隐层单元分配不同的概率权重,使得关键节点信息得到有效关注,降低了冗余信息对情感走向的影响。利用注意力机制可以使模型更好地理解语义信息,全方位地关注更重要的内容以及整体的情感走向。

本文定义  $e(\mathbf{x}_i, \mathbf{y}_j)$  是文本序列中各个单词与当前预测词的相似度函数,其计算如下:

$$e(\mathbf{x}_i, \mathbf{y}_j) = u \tan \mathbf{x}_i (v \mathbf{x}_i + w \mathbf{y}_j) + \theta \quad (5)$$

其中,  $u, v, w$  为模型参数。 $e(\mathbf{x}_i, \mathbf{y}_j)$  构成了一个全连接的前向神经网络,仅包含一个隐藏层。为了防止部分词向量相关度过高,过多削弱其他间隔较远词向量之间的相关度或影响整体的情感极性,本文引入相关度调节参数  $\theta$  来平衡  $e(\mathbf{x}_i, \mathbf{y}_j)$ 。除此之外,也可以直接按照如下公式进行计算:

$$e(\mathbf{x}_i, \mathbf{y}_j) = \mathbf{x}_i^T w \mathbf{y}_j \quad (6)$$

或者使用 2 个状态之间的点乘:

$$e(\mathbf{x}_i, \mathbf{y}_j) = \mathbf{x}_i^T \mathbf{y}_j \quad (7)$$

$$\mathbf{a}'_{i,j} = \frac{\exp(e(\mathbf{x}_i, \mathbf{y}_j))}{\sum_{i=1}^l \exp(e(\mathbf{x}_i, \mathbf{y}_j))} \boldsymbol{\lambda} \quad (8)$$

其中,  $\mathbf{a}'_{i,j}$  为每个词向量的注意力系数矩阵,  $l$  为原句的词向量长度,  $\boldsymbol{\lambda}$  为可调参数。

$$\mathbf{c}_j = \sum_i \mathbf{a}'_{i,j} \mathbf{x}_i \quad (9)$$

其中,  $\mathbf{c}_j$  为最终通过原文本词向量和注意力特征矩阵运算得到的注意力输入矩阵。

### 2.3.2 卷积层

卷积层是卷积神经网络的核心网络结构,卷积神经网络中每层卷积层由若干卷积单元组成,每个卷积单元的参数都是通过反向传播算法优化而得。卷积运算的目的是提取文本输入的不同特征,每个卷积核的“层数”需要与输入的“层数”一致,即局部感知野的大小对应于卷积核的大小。卷积层的主要任务是从不同的层面来选择前一层特征图的各角度特征使其具有位移不变性,卷积的本质就是对前一层的特征图进行处理,得到这一层的特征图,表达式如下:

$$f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + \mathbf{b} \quad (10)$$

其中,  $\mathbf{x}$  为卷积核窗口词向量矩阵,  $\mathbf{w}$  为权重矩阵,  $\mathbf{b}$  为偏置,  $f$  为激活函数。

长度为  $l$  的文本序列作用于卷积核矩阵  $\mathbf{V} = \mathbf{v}_0 \oplus \mathbf{v}_1 \oplus \dots \oplus \mathbf{v}_{l-1}$  得到一个卷积后的特征表示  $\mathbf{c}_i$ , 则对于每个长度为  $l$  的序列,有:

$$\mathbf{C}_i = f(\mathbf{w} \cdot \mathbf{x}_{n:n+l-1}) + \mathbf{b} =$$

$$f\left(\|\mathbf{w}\| \times \|\mathbf{x}_{n:n+l-1}\| \times \frac{\mathbf{w} \cdot \mathbf{x}_{n:n+l-1}}{\|\mathbf{w}\| \times \|\mathbf{x}_{n:n+l-1}\|} + \mathbf{b}\right) \quad (11)$$

其中,子序列  $\mathbf{x}_{i:i+l-1} = \mathbf{x}_i \oplus \mathbf{x}_{i+1} \oplus \dots \oplus \mathbf{x}_{i+l-1}$  是词向量的串联组合,  $\|\cdot\|$  是二阶  $l_2$  范数,  $\mathbf{b}$  为偏置矩阵,  $l$  为滑动窗口大小。此处的  $\mathbf{w}$  和  $\mathbf{x}_{n:n+l-1}$  基于切比雪夫定律,  $f$  表示非线性激活函数,使用该函数增强网络的拟合能力。结合实际调参过程,利用多个卷积核进行卷积得到:

$$\mathbf{c} = (\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_{n-l+1}) \quad (12)$$

激活函数可以单独构成卷积神经网络的一个“空”网络层——激活层,作为空层,其没有实际可训练的参数,仅对输入起到激活作用。激活函数加入了非线性因素,其中,线性整流函数(ReLU)又称修正线性单元,是一种人工神经网络中常用的激活函数,通常指代以斜坡函数及其变种为代表的非线性函数。ReLU 函数拥有更加有效的梯度下降以及反向传播性能,避免了梯度爆炸和梯度消失问题,其构建稀疏矩阵,去除数据中的冗余信息,最大可能地保留数据的特征。因为神经网络不断地反复计算,实际上是在不断尝试用一个大多数元素为 0 的矩阵来表达数据特征,因此,利用 ReLU 函数的稀疏性可以使得梯度下降效果更好,又不会丢弃过多的文本特征。相比 Sigmoid 和 tanh 函数,ReLU 在使用梯度下降时收敛速度更快,其只需一个门限值,即可在得到激活值的同时加快计算速度。因此,本文的卷积神经网络中采用 ReLU 函数。在经过卷积层后,得到的向量维度为  $d' = l - h + 1$ 。

### 2.3.3 池化层

池化层又称下采样层。通常在卷积层之后会得到维度很大的特征,将特征分成几个区域,取其最大值或平均值,得到新的或维度较小的特征,从而更加聚焦于文本中最主要的信息。池化层抽取的信息更具有细粒度,有效防止了模型过拟合问题,在一定程度上提高了算法的泛化性能。池化对每个输入的深度切片独立操作,规模一般为  $2 \times 2$ ,池化层一般选择最大池化操作,即取 4 个点的最大值。

传统的卷积神经网络模型并未充分利用卷积神经网络的特性去理解句子的语义结构,一般选择最大池化提取整个句子的最优特征。对于大量的情感转折句式,传统的池化操作很难有效提取其特征。一方面,情感极性突变的转折句往往蕴含了大量情感信息,这很符合人类语言的特点;另一方面,网络文本信息存在大量语法结构模糊、用词不完整的现

象,传统的池化操作对于这种类型的文本很难准确分析其情感变化。为了充分考虑转折句前后的情感极性变化,本文对传统的池化进行“分池”操作,以“情感转折词”为节点分解为若干“小池”,各个小池分别抓取文本序列特征中最有效的部分,再利用传统池化操作保留情感信息特征。具体如下:

$$\mathbf{c}_i = \mathbf{c}_{i1} + \mathbf{c}_{i2} + \cdots + \mathbf{c}_{im} \quad (13)$$

其中,  $m$  为分解的小池数量。考虑到文本序列的逻辑性,分解的小池一般不超过 3 个,即  $m \leq 3$ ,原因是句子中很难出现 3 个以上的情感极性转变。

取情感极性最高的情感特征进行矩阵相加,得到最终的情感属性特征向量,如下:

$$\mathbf{c}_{ii}^* = \sum_{i=1}^{n-l+1} \max(\mathbf{c}_{ii}) \quad (14)$$

$$\mathbf{c}_i^* = \mathbf{c}_{i1}^* \oplus \mathbf{c}_{i2}^* \oplus \cdots \oplus \mathbf{c}_{im}^* \quad (15)$$

### 2.3.4 输出层

在全连接层之前,如果神经元数目过大,学习能力过强,可能出现过拟合问题。本文引入 dropout、正则化等操作随机删除神经网络中的部分神经元,用以解决过拟合问题,通过局部归一化(LRN)、数据增强和交叉验证等来增强模型的鲁棒性。全连接层拼接所有特征,得到一个高质量的特征图并输入到输出层。输出层通过 Softmax 分类器输出预测分类节点,每一个节点代表一种分类,节点之和为 1。

$$\sigma_i(\mathbf{z}) = \frac{e^{z_i}}{\sum_{j=1}^m e^{z_j}} \quad (16)$$

$$\sum_{i=1}^j \sigma_i(\mathbf{z}) = 1 \quad (17)$$

其中,  $i$  为节点次序,  $\mathbf{z}_i = \mathbf{w}_i \mathbf{x} + \mathbf{b}_i$ 。

Softmax 使用交叉熵作为损失函数来衡量预测的(0,1)概率分布与实际的(0,1)值是否匹配。交叉熵越小,说明匹配越准确,模型精度越高。损失函数如下:

$$\text{loss} = -\frac{1}{n} \sum_{i=1}^n y_i (\ln a) + (1 - y_i) \ln(1 - a) \quad (18)$$

其中,  $P(y=1|\mathbf{x}) = a, a = \frac{1}{1 + e^{-x}}$ 。

最后选择 Adam Optimizer 优化整体网络参数,模型在使用梯度下降时,可以合理地动态选择学习速率,分配梯度下降幅度。当训练中损失函数接近最小值时减小每步的梯度下降幅度;如果损失函数曲线相对陡峭,则适当增大下降幅度。

## 3 实验结果与分析

### 3.1 实验数据

为了更好地测试模型的泛化能力,本文选择国内外不同类型的公开数据集,包含酒店评论、电影评

论和商品评论等,具体数据集信息如下:

1) MR<sup>[1]</sup>, 二元情绪分类数据集,包括 10 662 个电影评论(5 331 条正面评论和 5 331 条负面评论),每条评论仅包含一个句子。

2) CR<sup>[18]</sup>, 由用户在不同商品上的正、反评论组成的数据集,主要用于对评论进行情感倾向性分析。

3) MPQA (Multiple-Perspective QA) 库<sup>[19]</sup>, 由 JANYCE W 等人开发的深度标注语料库,包含 535 篇不同视角的新闻评论。

4) SemEval2014<sup>[20]</sup>, SemEval 发布的关于 Twitter 情感分析任务的语料库,分为酒店评论集和手提电脑评论集,每个集合由训练数据、验证数据和测试数据三部分组成。

5) IMDB, 互联网电影评论数据集,一共收集了来自 IMDB 的 25 000 条电影评论,以正面、反面情绪进行标记,全部评论经过预处理并编码为词索引的整数序列表示。

6) CHNSENTICORP(简称为 CSC), 国内学者谭松波收集整理的较大规模的酒店评论语料库,共 4 个数据集,包含 10 000 篇评论。

7) SST-2<sup>[21]</sup>, 斯坦福大学情绪数据集 SST-1 (SST-1 是 MR 数据集的扩展)的简化版本,仅包含二进制标签(正面和负面),共 9 613 条数据。

8) Subj<sup>[22]</sup>, 由具有二进制类别的 10 000 个用户生成的评论所组成的主观性数据集。

### 3.2 参数设置与训练

实验数据按照 8:1:1 分为训练数据、验证数据、测试数据。中文数据选择 jieba 分词工具,英文数据选择 NLTK 分词工具。模型初始参数设置如表 1 所示,评价指标为准确率、F1 值。

表 1 模型参数设置  
Table 1 Model parameters setting

参数	数值
卷积核窗口大小	3、4、5
词向量维度	300
每种卷积核数量	100
隐藏单元	200
Dropout	0.3

在其他参数保持一致的情况下,观察不同卷积核窗口大小对模型准确率的影响,结果如图 3 所示。从图 3 可以看出,训练过程中当窗口小于 3 时,模型的准确率随着窗口的增大而增大;当窗口大于 4 时,模型的准确率随着窗口的增大而减小;当窗口大小为 4 时,准确率曲线达到峰值。由此可以看出,窗口大小直接影响模型效果,窗口大小为 3、4、5 时,2 条曲线的准确率均相对较高,变化幅度也相对较小,模型处于最佳状态。

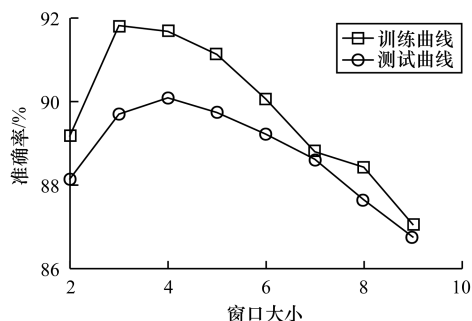


图3 窗口大小对模型性能的影响

Fig.3 Impact of window size on model performance

### 3.3 对比分析

#### 3.3.1 词向量测试

将文本信息转化为模型可以理解的语言,对自然语言处理至关重要<sup>[23]</sup>。一方面,词向量的训练可以学习到一些基本的情感信息,另一方面,不同的词向量对模型的影响程度不同。通过对比不同的词向量来选择适合模型的文本转换方式,从而为后续的模型改进提供依据。

本文选择 CNN 和 BiLSTM 两种模型,测试词向量 one-hot、Word2Vec 对模型的影响。为了对比词向量效果并加快实验效率,词向量测试模型参数取最简值,2 种测试词向量的对比结果如图 4 所示。

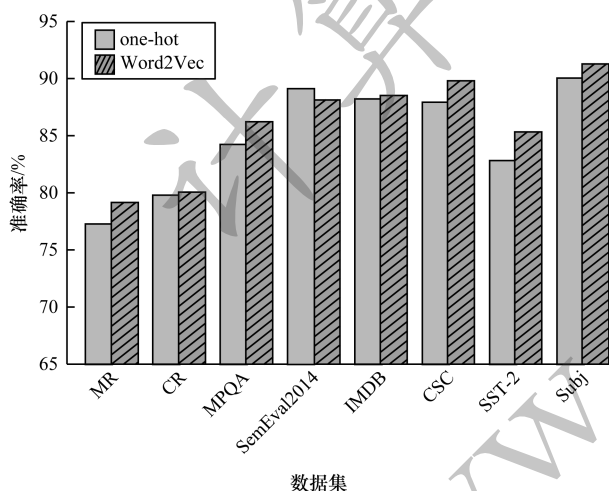


图4 2种词向量的性能对比结果

Fig.4 Performance comparison results of two word vectors

从图 4 可以看出,对于不同的数据集,2 种模型在使用 Word2Vec 时准确率均有明显提升,主要原因是 Word2Vec 关注了词向量空间位置中相同语境下相近位置出现的词语语义情感。因此,后续实验均在 Word2Vec 下进行。

#### 3.3.2 基线实验对比

本节实验对比模型如下:

1) SVM<sup>[24]</sup>:改进的基于支持向量机的早期机器学习算法模型。

2) LSTM<sup>[25]</sup>:使用长短期记忆网络,并将网络的最后一个隐藏状态反馈给一个带有 Softmax 函数的

线性分类器。

3) CNN<sup>[10]</sup>:较早使用卷积神经网络进行文本分类。

4) TREE-LSTM<sup>[26]</sup>:将长短期记忆网络推广到树形拓扑结构,解决了语义关联情感分类问题。

5) MNB<sup>[27]</sup>:改进的朴素贝叶斯模型。

6) DPCNN:本文的基准模型,没有添加注意力机制。

从图 5 可以看出,深度学习方法明显优于传统机器学习方法,尤其是当文本数据量较大时,深度学习可以很好地学习文本特点。其中,在数据量较小的 MR 数据集上,本文模型的准确率相较于传统 CNN 模型平均提升了 0.05%,同样在数据量较小且数据内容随机的 MPQA 数据集上,本文模型的准确率相较于传统深度学习模型平均提升了 2.50%。

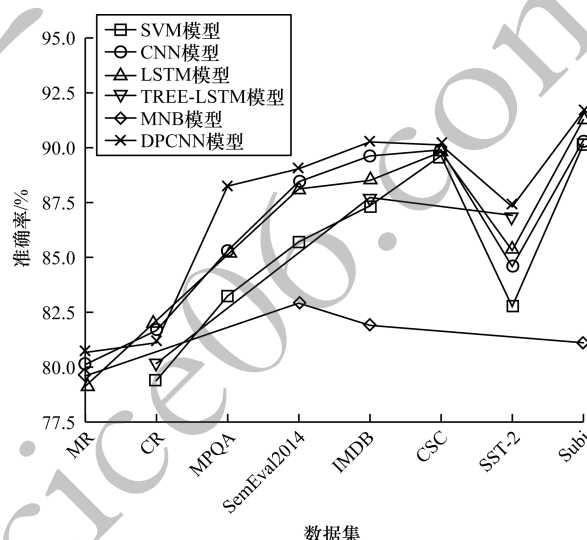


图5 6种传统模型的准确率对比结果

Fig.5 Accuracy comparison results of six traditional models

#### 3.3.3 注意力机制模型实验对比

本节实验对比模型具体如下:

1) WACNN<sup>[28]</sup>:在卷积层分别提取词语特征和整体的语义特征,并利用 Padding 方法模拟特征提取。

2) HAN<sup>[29]</sup>:将文本结构化,利用注意力机制在词、句 2 个不同的层级分别关注文本内容,更好地区分文本内容的重要程度。

3) AT-LSTM<sup>[15]</sup>:在长短期记忆网络的基础上引入注意力机制,在编码、解码过程中得到了更好的信息特征,为进行对比实验,本文不考虑主题层级的情感分析。

4) SELF-AT<sup>[30]</sup>:自注意增强递归神经网络,主要利用多头自注意力机制来改进特征选择,使信息在较长时间内保持依赖关系。

5) AT-DPCNN:本文的基准模型结合改进的注意力机制。



从图6可以看出,采用注意力机制的深度学习模型在自然语言处理方面优势明显,准确率相对传统模型平均提升了2%。在平均句长较长、文本数据较为整洁的 Subj 数据集上,采用注意力机制的模型均达到了最佳效果。其中,基于长短期记忆网络的注意力模型 AT-LSTM 和本文模型效果最好。主要原因在于长短期记忆网络善于捕捉句子中不同位置之间词的关系,随着句长的增加,AT-LSTM 的性能提升效果较为明显,因此,基于长短期记忆网络的文本算法模型更适合处理长文本序列以及文档级分类工作。本文模型高度关注了长文本中对整个句子起关键作用的局部情感信息,结合卷积神经网络提取到关键节点内容,模型在数据量较小的 MR 数据集上依旧得到了性能提升。对于文本信息嘈杂、个人主观性较强的商品评论文本数据集 CR,本文模型在准确率上也提升明显,从另一方面证明了本文模型较好的泛化性能。

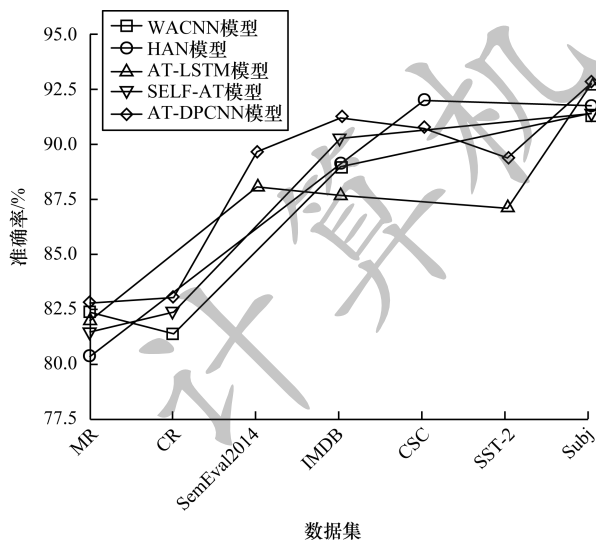


图6 5种注意力模型的准确率对比结果

Fig.6 Accuracy comparison results of five attention models

为了更好地测试转折句与非转折句对于文本分类结果的影响,从CSC数据集中选取数据进行实验,如图7所示。

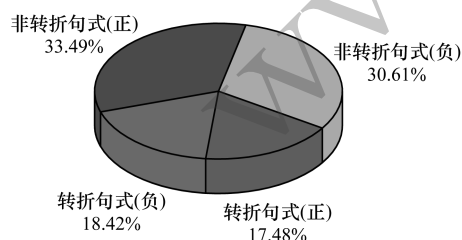


图7 分池实验数据

Fig.7 Experimental data of cell division

从图8可以看出,测试模型在非转折句式下的F1值均高于转折句式,表明测试模型在转折句式下的文本处理效果较差。相比WACNN和HAN模型,

本文模型在转折句式下F1值平均提升了3%,与非转折句式下的F1值差距缩小到1.5%,说明该模型在转折句式的处理中取得了很大的性能提升。

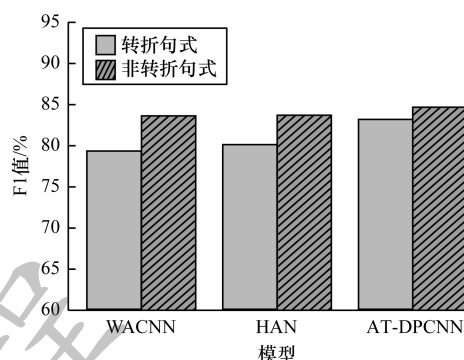


图8 2种句式的实验结果对比

Fig.8 Comparison of experimental results of two sentence patterns

#### 4 结束语

本文将卷积神经网络共享权值、具有多特征图等特点与注意力机制关注局部内容的特点相结合,建立一种新的情感分析模型AT-DPCNN。改进的卷积神经网络可以提取文本信息的潜在有效特征,注意力矩阵能够获得关键局部特征,从而全方位提取文本的有效特征,解决传统卷积神经网络在提取特征过程中存在的信息丢失问题,提高模型的准确率。此外,该模型通过分池操作有效提高了转折句式的情感判别效果。在多个不同类型的数据集上进行实验,结果表明,该模型具有较高的泛化性能,适用于实际中的文本信息处理任务。

大数据环境下的数据类型更加多样,表情包等其他情感元素在文本信息中的普及使得信息的复杂性大幅提升。因此,如何更好地处理多元素复杂文本信息并利用深度学习技术在多模态领域进行情感分析,将是下一步的研究重点。

#### 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. New York, USA: ACM Press, 2005: 115-124.
- [2] HONG Wei, LI Min. A review: text sentiment analysis methods[J]. Computer Engineering and Science, 2019, 41(4): 750-757. (in Chinese)  
洪巍,李敏. 文本情感分析方法研究综述[J]. 计算机工程与科学, 2019, 41(4): 750-757.
- [3] YAN Xiaodong, HUANG Tao. Tibetan sentence sentiment classification based on emotion dictionary[J]. Journal of Chinese Information Processing, 2018, 32(2): 75-80. (in Chinese)  
闫晓东,黄涛. 基于情感词典的藏语文本句子情感分类[J]. 中文信息学报, 2018, 32(2): 75-80.

- [ 4 ] MEI Lili, HUANG Heyan, ZHOU Xinyu, et al. A survey on sentiment lexicon construction [ J ]. Journal of Chinese Information Processing, 2016, 30(5) : 19-27. (in Chinese)  
梅莉莉, 黄河燕, 周新宇, 等. 情感词典构建综述 [ J ]. 中文信息学报, 2016, 30(5) : 19-27.
- [ 5 ] YANG Xiaoping, ZHANG Zhongxia, WANG Liang, et al. Automatic construction and optimization of sentiment lexicon based on Word2Vec [ J ]. Computer Science, 2017, 44(1) : 42-47, 74. (in Chinese)  
杨小平, 张中夏, 王良, 等. 基于 Word2Vec 的情感词典自动构建与优化 [ J ]. 计算机科学, 2017, 44(1) : 42-47, 74.
- [ 6 ] PANG B, LEE L. Opinion mining and sentiment analysis [ J ]. Foundations and Trends<sup>®</sup> in Information Retrieval, 2008, 2(1/2) : 1-135.
- [ 7 ] PANG B, LEE L, VAITHYANATHAN S. Thumbs up? [ C ] // Proceedings of ACL Conference on Empirical Methods in Natural Language Processing. New York, USA: ACM Press, 2002 : 79-86.
- [ 8 ] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality [ EB/OL ]. [ 2019-08-20 ]. <https://arxiv.org/pdf/1310.4546.pdf>.
- [ 9 ] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space [ C ] // Proceedings of International Conference on Intelligent Text Processing and Computational Linguistics. Berlin, Germany: Springer, 2013 : 430-443.
- [ 10 ] KIM Y. Convolutional neural networks for sentence classification [ C ] // Proceedings of 2014 Conference on Empirical Methods in Natural Language Processing. New York, USA: ACM Press, 2014 : 1746-1751.
- [ 11 ] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate [ EB/OL ]. [ 2019-08-20 ]. <https://arxiv.org/abs/1409.0473>.
- [ 12 ] LUONG T, PHAM H, MANNING C D. Effective approaches to attention-based neural machine translation [ C ] // Proceedings of 2015 Conference on Empirical Methods in Natural Language Processing. New York, USA: ACM Press, 2015 : 1412-1421.
- [ 13 ] CHEN P, SUN Z Q, BING L D, et al. Recurrent attention network on memory for aspect sentiment analysis [ C ] // Proceedings of 2017 Conference on Empirical Methods in Natural Language Processing. New York, USA: ACM Press, 2017 : 463-472.
- [ 14 ] BAZIOTIS C, PELEKIS N, DOULKERIDIS C. Data stories at SemEval-2017 task 4: deep LSTM with attention for message-level and topic-based sentiment analysis [ C ] // Proceedings of the 11th International Workshop on Semantic. New York, USA: ACM Press, 2017 : 747-754.
- [ 15 ] WANG Yeguan, HUANG Minlie, ZHU Xiaoyan, et al. Attention-based LSTM for aspect-level sentiment classification [ C ] // Proceedings of 2016 Conference on Empirical Methods in Natural Language Processing. New York, USA: ACM Press, 2016 : 606-615.
- [ 16 ] KUMAR A, KAWAHARA D, KUROHASHI S, et al. Knowledge-enriched two-layered attention for sentiment analysis [ C ] // Proceedings of the 9th American Chapter of the Association for Computational Linguistics. [ S. l. ] : ACL, 2018 : 253-258.
- [ 17 ] YIN W P, 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.
- [ 18 ] HU Mingqing, LIU Bing. Mining and summarizing customer reviews [ C ] // Proceedings of 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA: ACM Press, 2004 : 168-177.
- [ 19 ] WIEBE J, WILSON T, CARDIE C. Annotating expressions of opinions and emotions in language [ J ]. Language Resources and Evaluation, 2005, 39(2/3) : 165-210.
- [ 20 ] TANG D Y, QIN B, FENG X C, et al. Effective LSTMs for target-dependent sentiment classification [ EB/OL ]. [ 2019-08-20 ]. <https://arxiv.org/abs/1512.01100>.
- [ 21 ] SOCHER R, PERELYGIN A, WU J Y. Recursive deep models for semantic compositionality over sentiment treebank [ EB/OL ]. [ 2019-08-20 ]. [http://www.cs.toronto.edu/~fidler/teaching/2015/slides/CSC2523/sentiment\\_zhcong.pdf](http://www.cs.toronto.edu/~fidler/teaching/2015/slides/CSC2523/sentiment_zhcong.pdf).
- [ 22 ] PANG B, LEE L. A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts [ EB/OL ]. [ 2019-08-20 ]. <https://arxiv.org/abs/cs/0409058>.
- [ 23 ] WU Chen, ZHANG Quan. Research on rules for detecting Chinese sentence groups in nature language processing [ J ]. Computer Engineering, 2007, 33(4) : 157-159. (in Chinese)  
吴晨, 张全. 自然语言处理中句群划分及其判定规则研究 [ J ]. 计算机工程, 2007, 33(4) : 157-159.
- [ 24 ] SILVA J, COHEUR L, MENDES A C, et al. From symbolic to sub-symbolic information in question classification [ J ]. Artificial Intelligence Review, 2011, 35(2) : 137-154.
- [ 25 ] GRAVES A. Generating sequences with recurrent neural networks [ EB/OL ]. [ 2019-08-20 ]. <https://arxiv.org/pdf/1308.0850v5.pdf>.
- [ 26 ] TAI K S, SOCHER R, MANNING C D. Improved semantic representations from tree-structured long short-term memory networks [ J ]. Computer Science, 2015, 5(1) : 36-40.
- [ 27 ] WANG S, MANNING C D. Baselines and bigrams: simple, good sentiment and topic classification [ C ] // Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics. [ S. l. ] : ACL, 2012 : 90-94.
- [ 28 ] WANG Shengyu, ZENG Biqing, SHANG Qi, et al. Word attention-based convolutional neural networks for sentiment analysis [ J ]. Journal of Chinese Information Processing, 2018, 32(9) : 123-131. (in Chinese)  
王盛玉, 曾碧卿, 商齐, 等. 基于词注意力卷积神经网络模型的情感分析研究 [ J ]. 中文信息学报, 2018, 32(9) : 123-131.
- [ 29 ] YANG Z C, YANG D Y, DYER C, et al. Hierarchical attention networks for document classification [ C ] // Proceedings of the 9th American Chapter of the Association for Computational Linguistics. [ S. l. ] : ACL, 2016 : 1480-1489.
- [ 30 ] LIN Z H, FENG M W, SANTOS C, et al. A structured self-attentive sentence embedding [ EB/OL ]. [ 2019-08-20 ]. <https://arxiv.org/abs/1703.03130>.