

基于词嵌入辅助机制的情感分析

韩旭丽¹ 曾碧卿² 曾锋¹ 张敏¹ 商齐¹

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

摘要 文本情感分析是自然语言处理研究领域中的一个重要的研究方向,如何分析出长文本的情感极性是一个研究难点。目前,大部分研究工作倾向于将词嵌入应用在神经网络模型中进行情感分析,虽然这种方法的词特征表示能力较好,但是对于长文本来说有待优化,过长的文本会给模型带来沉重的负担,使模型在训练过程中耗费更多的时间和计算资源。针对此问题,文中提出了一种基于词嵌入辅助机制的注意力神经网络模型(Word Embedding Auxiliary Mechanism Based Attentional Neural Network Model, WEAN),并将其应用于长文本的情感分析任务。该模型使用词嵌入辅助机制解决了长文本在神经网络模型中的训练负担问题,利用双向循环神经网络获取序列中的上下文信息,同时应用注意力机制来捕获序列中不同重要程度的信息,提高了情感分类的性能。在 IMDB, Yelp 2013 和 Yelp 2014 数据集上的实验结果表明,与 NSC+LA 模型相比,所提模型的情感分析准确率分别提高了 1.1%, 2.0% 和 2.6%。

关键词 情感分析,词嵌入,注意力机制,神经网络,自然语言处理

中图分类号 TP391 文献标识码 A DOI 10.11896/jsjx.180901687

Sentiment Analysis Based on Word Embedding Auxiliary Mechanism

HAN Xu-li¹ ZENG Bi-qing² ZENG Feng¹ ZHANG Min¹ SHANG Qi¹

(School of Computer Science, South China Normal University, Guangzhou 510631, China)¹

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

Abstract Text sentiment analysis is an important research direction in natural language processing, and how to analyze the sentiment polarity of long text is a research hotspot. At present, majority of the researches tend to apply word embedding in neural network models for sentiment analysis. Although this method has a good representation ability of word features, it has some weaknesses for long text. Extremely long text will bring heavy burden to the model and make it consume more time and resources in training process. In light of this, this paper proposed an attention neural network model based on word embedding auxiliary mechanism, namely WEAN, and it is applied to sentiment analysis for long text. The model deals with some training burdens of long text in neural network model by using word embedding auxiliary mechanism and utilizes bidirectional recurrent neural network and mechanism to obtain context information. At the same time, it captures information with different importance degree in sequences, thus improving the sentiment classification performance. Experiment was conducted on IMDB, Yelp 2013 and Yelp 2014 datasets. The results show that the accuracy of the sentiment analysis of the proposed model increases by 1.1%, 2.0% and 2.6% respectively on three datasets compared with NSC+LA model.

Keywords Sentiment analysis, Word embedding, Attention mechanism, Neural network, Natural language processing

1 引言

文本的情感分析是自然语言处理中一个重要的任务。在当前的大数据时代,情感分析的任务受到社会广泛关注。现如今,利用情感分析来对大数据进行处理,并预测文本的情感倾向性成为了一个研究热点。文本情感分析已经应用到社会生活的各个领域,如推荐系统、心理学研究和舆情预测等。

广义上,文本情感分析是指计算机对一段文本进行研究

并分析其情感极性。情感分析的研究方法主要分为基于规则的、基于机器学习的和基于深度学习的情感分析方法。基于规则的情感分析主要依靠人工标记的情感词典进行文本情感的倾向性分析^[1];但随着时代的发展,大量的新词涌现,需要人工更频繁地更新情感词典,因此该研究方法浪费了大量的时间和财力,并且在性能方面也不够突出。基于机器学习的方法是将情感分析视为与文档分类和主题分类相似的任务^[2],该方法采用语言模型构建篇章特征,使用机器学习的算

收稿日期:2018-09-09 返修日期:2018-12-22 本文受国家自然科学基金项目(61772211,61503143)资助。

韩旭丽(1992-),女,硕士生,主要研究方向为自然语言处理、情感分析;曾碧卿(1969-),男,博士,教授,CCF 高级会员,主要研究方向为自然语言处理,E-mail:zengbiqing0528@163.com(通信作者);曾锋(1992-),男,硕士生,主要研究方向为自然语言处理;张敏(1991-),女,硕士生,主要研究方向为自然语言处理;商齐(1994-),男,主要研究方向为自然语言处理。

法作为文本分类器来对文本进行情感倾向性分析。基于深度学习的方法将神经网络模型的参数训练和特征学习进行并行化处理,将文本映射成矩阵并作为模型的输入,再将模型学习到的特征表示在模型分类器中进行分类,从而完成情感分析的分类任务^[3]。

如何将文本表示成计算机容易识别的语言是研究重点。早期,最常见的方法是 one-hot 向量表示方法。但是 one-hot 表示方法有两个缺点:1)随着文本长度的增大,向量维度不断增大,从而导致数据稀疏问题;2)由于每个词都表示为一个向量,文本语义相关性在此表示方法中无法体现出来。词嵌入的方法由 Mikolov 等^[4]于 2013 年首次提出,该方法弥补了 one-hot 向量表示方法的不足。在神经网络研究方法中,通过词嵌入的方法将文本映射成向量并作为模型输入,其主要是利用词语的上下文关系和上下文与目标词之间的关系进行建模。词嵌入的方法将每个词转化为一个固定长度的实值向量,由此,文本转化为一个低维度的稠密型的向量空间。深度模型利用词嵌入表示文本的方法进行建模^[5]。词嵌入表示方法虽然很好,但是在长文本中,由于文本长度过长,会导致模型在训练过程中耗费过多的时间和资源。一些常见的模型试图通过改进模型来提高性能,常见的有融合多层感知机模型、卷积神经网络(Convolutional Neural Network,CNN)模型、循环神经网络(Recurrent Neural Network,RNN)模型等。多层感知机模型由神经网络中的多个神经元直接相连构成,许多研究者将其应用在情感分析上进行研究,研究者们发现随着神经网络层数的增加,模型将出现梯度消失等问题,这些问题使得模型训练变得比较复杂,并且容易出现过拟合。卷积神经网络建模方法将卷积层作为中间层^[6-7],利用卷积操作实现参数共享,解决了参数过多的问题。但是卷积神经网络方法在自然语言处理的并行化操作上没有考虑上下文信息关系。而循环神经网络具有处理时间序列信息的功能,其处理序列的方式是将每一个词的序列表示信息逐个输入并计算,每一个序列信息的输出结果都通过将前一个输出的结果与当前输入序列信息进行函数计算得到。但循环神经网络具有无法实现序列长期依赖的问题,并且随着序列的增长,可能导致梯度消失和梯度爆炸等问题。长短期记忆网络(Long Short-Term Memory,LSTM)可以解决梯度消失问题,实现序列长期依赖^[5,8],其原理在于通过门控单元实现选择记忆时间序列信息,从而解决梯度消失的问题。为了更好地利用上下文信息,双向的长短期记忆网络(Bi-LSTM)利用两个方向的 LSTM 实现了序列上下文信息依赖。但是在自然语言处理上,特别是对情感分析的研究中,不同的序列信息在文本中的重要程度亦不相同,循环神经网络无法解决此问题,但可以用注意力机制来解决这一缺陷问题,即通过加权计算每个词表示的权值来得到每个词语特征表示的重要程度值。

基于以上因素,本文提出了基于词嵌入辅助机制的注意力神经网络模型 WEAN。本模型不仅考虑了序列的上下文信息依赖,还解决了控制参数的指数爆炸等问题。该模型主要分为 3 个部分,第一部分为词嵌入,第二部分为词序信息,第三部分为情感分类。第一部分主要对数据进行特征表示;在词嵌入后加上词嵌入辅助机制,帮助处理复杂度计算问题

和序列顺序的问题。该部分主要针对计算复杂度问题进行池化操作,将 3 种池化方法分别作用于词嵌入后的词向量上,3 种池化方法分别是平均池化方法(Ave-Pooling)、最大池化方法(Max-Pooling),以及结合了平均池化和最大池化的方法。第二个部分以上下文信息和词序信息为切入点,将第一部分得到的词特征表示信息作为输入,使用双向的长短期记忆神经网络进行建模。由于每个词在不同上下文中的重要程度不同,其情感极性的作用也不同,因此在本层次模型上加上注意力机制来计算句子中不同词的重要程度值。第三个部分即分类层,根据得到的序列特征表示进行文本的情感分类。

本文模型在 IMDB, Yelp 2013 和 Yelp 2014 数据集上进行了实验验证,准确率分别达到了 49.8%,65.1%和 65.6%。相比于 NSC+LA 模型^[9],本文模型的准确率在这 3 个数据集上分别提升了 1.1%,2%和 2.6%。本文的贡献如下:

1)提出了词嵌入辅助机制,将池化方法应用在词嵌入表示的向量空间上,提取词表示信息,缩短了训练时间,并提升了模型性能。

2)考虑到每个词的重要程度不同,在模型中引入注意力机制,并计算每个词的重要程度信息。

3)在 IMDB, Yelp 2013 和 Yelp 2014 数据集上进行实验,利用本文模型得到的实验效果都有不同程度的提升。

2 相关工作

情感分析是自然语言处理中的一项基本工作。传统的情感分析方法主要是利用统计学习进行情感分析,构建情感词典,通过机器学习算法进行情感分类^[2]。Pang 等^[10]利用机器学习算法并借助词性标注等特征在电影评论数据集上进行情感分析。Turney 等^[11]利用无监督学习方法提取情感词和短语,来确定文本情感极性。Wang 等^[1]通过利用 Ngram 模型构建的特征选择器来进行情感分类。

近年来,深度学习在自然语言学习领域的应用越来越广泛^[12-13]。词嵌入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^[4,14-15]。大量研究方法利用词嵌入神经网络模型进行情感分析研究。Mikolov 等^[4]提出了词嵌入方法(即 Skip-Gram 和 CBOW 方法)来对词向量进行建模,这种方法的优点是在缩短训练时间的同时达到了较好的实验效果。一些研究者利用词嵌入和卷积神经网络进行建模,其在文档分类上的应用取得了较好的实验效果。诸如, Kalchbrenner 等^[16]提出了一个多层卷积神经网络模型,该模型随机初始化词向量并对输入层的词向量进行降维操作,利用卷积神经网络进行建模并在句子分类应用上取得了良好的实验效果。Kim 等^[6]提出利用卷积神经网络模型进行文本分类,并且通过微调参数达到了很好的分类效果。Johnson 等^[17]提出用词序列的卷积神经网络来进行文本分类,该方法考虑了词之间的先后顺序问题,提高了整体分类的精度。Zhang 等^[18]在卷积神经网络的分类模型上提出几种参数调整方案,得到参数调优的总结经验,并在文本分类中取得了良好的效果。循环神经网络具有捕获序列顺序信息的功能^[19]。Le 等^[20]提出用 Paragraph Vector 的方法对文本进行向量化表示,并在文本分类上取得良好性能,该方法在词向量表示上

优于词袋模型。Tang 等^[21]提出门控循环神经网络模型并在情感分类上进行实验,其优点在于模型提取特征值时考虑了句子间上下文的关系,使得模型在情感分类上取得了良好的分类效果。

注意力机制被广泛应用于深度学习的神经网络模型中^[9,22-23]。Mnih 等^[24]将注意力机制首次应用于图像分类,并在图像分类上取得了良好的分类效果。基于此,研究者将注意力机制应用于自然语言处理领域,Bahdanau 等^[25]首次将注意力机制用于机器翻译系统,减少了编码器的编码负担,提高了翻译的质量。Luong 等^[26]提出了全局注意力机制和局部注意力机制的概念,并将两者应用于机器翻译系统,取得了良好的翻译效果。Yang 等^[9]利用注意力机制与神经网络模型进行联合建模,在文本分类上取得了良好的实验效果。基于此,本文利用注意力机制和双向的长短期记忆网络进行联合建模,并对长文本进行情感分析研究。

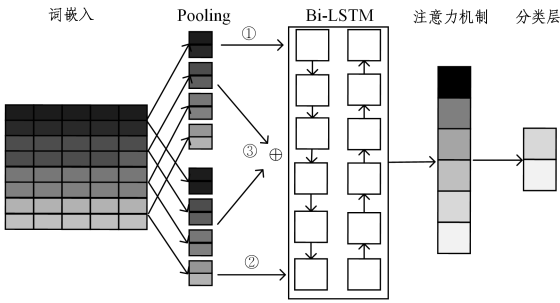


图1 基于词嵌入辅助机制的注意力神经网络模型

Fig.1 Attention neural network model based on word embedding auxiliary mechanism

3 WEAN 模型

本模型主要分为4个模块,第一个模块为词嵌入以及词嵌入辅助层 Pooling 机制;第二个模块为双向 LSTM,即 Bi-LSTM;第三个模块为注意力模块;第四个模块为分类层模块。其模型图如图1所示,其中,①代表用第一种词嵌入辅助 Ave-Pooling 方法,②表示用第二种词嵌入辅助 Max-Pooling 方法,③代表用第三种词嵌入辅助 Concat-Pooling 方法。

3.1 词嵌入

假设文本用 X 表示,一个文本由 L 个词组成,则文本可以表示为 $\{W_1, W_2, \dots, W_i, \dots, W_L\}$,其中 W_i 表示文本中的第 i 个词,用 GloVe 进行词向量映射,得到词嵌入特征表示 $\{v_1, v_2, \dots, v_i, \dots, v_L\}$,其中 $v_i \in R^K$, K 表示每个词的维度。本文在词嵌入的基础上引入了词嵌入的辅助机制,即在词嵌入后加上一层池化层,得到新的词嵌入表示。下面主要介绍3种词嵌入的辅助机制。

(1)平均池化操作,即 Ave-Pooling。设池化窗口大小为2,即在 GloVe 词嵌入的基础上进行取平均池化操作,得到池化后的词嵌入特征表示为 $\{Z_1, Z_2, \dots, Z_i, \dots, Z_{\frac{L}{2}}\}$,其中 $Z_i \in R^K$,如式(1)所示, Z_i 表示池化后得到的特征值。由于池化窗口大小取2,因此池化后特征表示个数将变化为 $\frac{L}{2}$ 。

$$Z_i = \text{Ave-Pooling}(v_{2i-1}, v_{2i}) \quad (1)$$

(2)最大池化操作,即 Max-Pooling。设窗口大小为2,与平均池化操作类似,得到的词表示为 $\{Z_1, Z_2, \dots, Z_i, \dots, Z_{\frac{L}{2}}\}$,其中 $Z_i \in R^K$ 。如式(2)所示, Z_i 表示池化后得到的特征值。由于池化窗口大小取2,因此池化后特征表示个数将变化为 $\frac{L}{2}$ 。

$$Z_i = \text{Max-Pooling}(v_{2i-1}, v_{2i}) \quad (2)$$

(3)前两种方法的并行化连接操作,即 Concat-Pooling。用并行的方式使两种池化同时进行,将得到的池化后的表示连接在一起作为新的表示 $\{Z_1, Z_2, \dots, Z_i, \dots, Z_L\} \{Z_1, Z_2, \dots, Z_i, \dots, Z_L\}$ 。 Z_i 表示连接后得到的特征值,如式3所示:

$$Z_i = \text{Concat-Pooling}(\text{Ave-Pooling}(v_{2i-1}, v_{2i}), \text{Max-Pooling}(v_{2i-1}, v_{2i})) \quad (3)$$

3.2 Bi-LSTM

在词嵌入辅助层后加上神经网络用于提取文本特征。目前在深度学习的应用方面,比较常见的神经网络模型有卷积神经网络和循环神经网络,用这两种神经网络进行建模得到的模型在自然语言处理领域达到的实验效果也比较理想。循环神经网络(RNN)能够处理序列顺序信息的问题,并且可以处理不同长度的序列。其利用激活函数序列输入特征表示 Z_i 和前一时刻的隐层状态输出值 h_{t-1} ,并将其转化为当前的隐层状态输出值 h_t :

$$h_t = f(h_{t-1}, Z_t) \quad (4)$$

但是循环神经网络不能解决长期依赖的问题,并且随着序列长度的增加,将出现梯度消失和梯度爆炸等情况。Schmidhuber 等^[5]利用长短期记忆网络(LSTM)解决了这个问题。LSTM 主要有3个门,即输入门、遗忘门和输出门,其转化函数的公式如下:

$$i_t = \sigma(W_i \cdot [h_{t-1}, w_t] + b_i) \quad (5)$$

$$f_t = \sigma(W_f \cdot [h_{t-1}, w_t] + b_f) \quad (6)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, w_t] + b_o) \quad (7)$$

$$g_t = \tanh \cdot (W_r \cdot [h_{t-1}, w_t] + b_r) \quad (8)$$

$$C_t = i_t \odot g_t + f_t \odot C_{t-1} \quad (9)$$

$$h_t = o_t \odot \tanh(C_t) \quad (10)$$

其中, i_t 代表输入门, f_t 代表遗忘门, g_t 代表候选记忆单元, C_t 代表当前时刻记忆单元, o_t 代表输出门, h_t 代表最终输出。 $W_i, W_f, W_o, W_r, b_i, b_f, b_o, b_r$ 分别是输入门、遗忘门和输出门的参数。 σ 是 sigmoid 函数。

考虑到上下文信息在情感分析中的重要性,本文模型用双向的长短期记忆网络进行建模,双向的长短期记忆网络由一个正向的 LSTM 和一个反向的 LSTM 组成,计算公式如下:

$$\vec{h}_t = f_{\text{lstm}}(h_{t-1}, Z_t) \quad (11)$$

$$\overleftarrow{h}_t = f_{\text{lstm}}(h_{t+1}, Z_t) \quad (12)$$

$$h_t = [\vec{h}_t, \overleftarrow{h}_t] \quad (13)$$

其中, \vec{h}_t 表示正向的 LSTM 得到的特征表示, \overleftarrow{h}_t 表示反向的 LSTM 得到的特征表示。

3.3 注意力机制

在神经网络层后加入注意力机制层。由于随着序列

的增长,无法区分 LSTM 产生的定长的序列信息的重要程度,因此为了突出与当前输出更相关的信息,需要对某些词进行重点关注。本文利用注意力机制解决此类问题。每个词的特征由得分函数产生的得分值 α 与每个时刻的输出特征表示进行加权求和得到新的特征表示 C_i 。计算公式如下:

$$C_i = \sum_{j=1}^L \alpha_j h_j \quad (14)$$

$$\alpha_i = \frac{\exp(e_i^T A)}{\sum_{k=1}^L \exp(e_k^T A)} \quad (15)$$

$$e_i = \tanh(W_i h_i + b) \quad (16)$$

其中,注意力机制利用得分函数突出重点部分,得分值的取值范围为 $[0, 1]$,每个词的特征得分值表示为 α ,其中 α_i 表示句子的第 i 个词的特征得分。 α 与每个时刻的状态信息的上下文有关,由此引入上下文向量 \mathbf{A} ,其中 \mathbf{A} 是在训练过程中进行统一初始化的。 W_i 和 b 是权重参数和偏置参数,通过模型的不训练得到。

3.4 情感分类

情感分类作为模型的最后一层,主要用于构造一个分类器。用线性函数将文本得到的特征表示转化为实值向量,并将该向量映射到一个 C 类的向量空间上(类别数是 C),最后加上 softmax 层来进行文本的情感分类,输出的分类概率为 p_C :

$$p_C = \frac{\exp(W_C C_i + b_C)}{\sum_{k=1}^C \exp(W_k C_k + b_k)} \quad (17)$$

在训练数据时,用监督学习方法对本文模型进行评估。使用交叉熵对真实值与预测值之间的误差进行评估:

$$\text{loss} = - \sum_{d \in T} \sum_{c=1}^C p_C^e(d) \log(p_C(d)) \quad (18)$$

其中, p_C^e 是真实分类, T 代表训练的文本。

4 实验

本节主要进行对比实验,即将情感分析领域的一些模型的方法与本文模型 WEAN 的方法进行对比,分析其差别以及优缺点。本节首先对实验设置进行描述;然后进行对比实验,得出实验结果并分析各个方法的优缺点;接下来在本实验的基础上进行实验分析,比较本模型一些模块的优缺点;最后将本文模型的注意力机制层进行可视化展示。

4.1 实验设置

首先对本文使用的数据集进行描述,然后进行参数设置,描述评估方法,最后对对比方法进行描述。

4.1.1 数据集

本文实验使用 3 个大型数据集,详细信息如表 1 所列。IMDB 数据集是电影网的评论数据集,包含了 84 919 条文本评论数据,被分为了 10 类,它的评分范围是 1—10 分,其中 1 分为最低分,10 分为最高分。分数越高,表达的文本情感积极性越大;同理,分数越低,表示的情感消极性越大;分数为 5 时,表示的文本情感为中性。Yelp 数据集来自最大点评网的数据集,其中, Yelp 2013 数据集是最大点评网站数据集中 2013 年的餐厅评论数据集,包含了 78 977 条文本评论数据,被分为 5 类。Yelp 数据集的评分范围按照类别分为 1—5

分,其中 5 分为最高分,1 分为最低分。5 分表达的情感积极性最大,1 分表示的情感消极性最大,3 分表示中性情感。Yelp 2014 数据集是最大点评网站数据集 2014 年的餐厅评论数据集,包含了 231 163 条文本评论数据,被分为了 5 类。本文按照 8:1:1 的比例分别将每个数据集分割为训练集、验证集和测试集。

表 1 3 个数据集的统计数据

Table 1 Statistical data of three datasets

数据集	类别	文本个数
IMDB	10	84 919
Yelp 2013	5	78 977
Yelp 2014	5	231 163

4.1.2 参数设置

数据集的预处理采用 Keras 自带的 Tokenizer 接口进行分词处理,使用 GloVe 词嵌入作为文本中词语的初始化操作,并将每个词映射成 300 维的特征向量。对于未登录词,按照均匀分布范围 $[-0.01, 0.01]$ 进行初始化。本文模型用到的优化器是 Adadelta 优化器,学习率为 1.0。为了防止过拟合,模型用了两层 Dropout,分别放在词嵌入辅助层之后和词注意力层之后,其值设为 0.5。训练批次大小设置为 256。LSTM 的单元个数设置为 256。

4.1.3 评估方法

本文评估模型的方法是准确率,通过准确率(见式(19))的高低来判断模型的性能是否优于基线模型的性能。

$$\text{acc} = \frac{T}{N} \quad (19)$$

其中, acc 代表模型的实验准确率, T 是通过模型预测的评分为准确评分的数量, N 是文档的总数。

4.1.4 对比实验

本节首先对一些基线模型进行描述。为比较本文模型在情感分析上的差异,将 9 个模型作为对比模型。其次对本文模型进行对比实验,以验证其优缺点。

9 个基线模型如下。

1) Majority: 将训练集中占多数的情感标签视为测试集的情感标签。

2) Ngram: 将一元词、二元词和三元词作为特征训练 SVM 分类器^[27]。

3) TextFeature: Kiritchenko 等^[28]提出一种有监督的文本分类的方法,该方法用于提取文本的词特征、字符特征和词性特征,并利用 SVM 作为分类器进行模型的训练。

4) AvgWordvec: 将文档的词向量平均化,得到文档表示,以此训练 SVM 分类器^[29]。

5) SSWE: 使用特别训练的情感词向量生成特征, SVM 为分类器^[30]。

6) NSC: 用双层的 LSTM 分别获取词特征和句子特征,构建篇章表示。

7) NSC+LA: 通过结合双层 LSTM 和全局注意力机制来提取篇章的特征表示。

8) RNTN+RNN: 使用递归神经网络提取句子的特征表示,并用循环神经网络得到篇章的特征表示^[31]。

9) Paragraph Vector: 将 PVDM 模型用于情感分类^[32]。

本文模型如下。

1) WEAN-MAX: 用 Max-Pooling 的方法进行词嵌入辅助机制的实验。

2) WEAN-AVE: 用 Ave-Pooling 的方法进行词嵌入辅助机制的实验。

3) WEAN-CON: Concat-Pooling, 结合以上两种方法进行词嵌入辅助机制的实验。

4) WEAN-NP: 保持其他设置不变, 去掉模型中的词嵌入辅助机制层, 即去掉 Pooling 层, 验证本模型中 Pooling 机制的性能好坏。

5) WEAN-NA: 保持其他设置不变, 去掉模型中的词注意力层, 验证本模型中注意力的性能好坏。

6) WEAN-CNN: 保持其他设置不变, 将模型中的 Bi-LSTM 模块替换成卷积神经网络进行实验, 卷积层窗口大小保持默认设置, 卷积核大小设为 3, 4 和 5。用以验证上下文信息的重要性。

4.2 实验结果

本节首先对基线模型进行对比实验以验证其性能, 并分析对比模型的差异性以及一些模型具有优越性的原因。其次对本文模型中的一些模块进行分析, 探索各模块对性能的影响。

在 3 个数据集上分别得到模型的准确率, 本文将模型的对比实验分为 3 组: 第一组以传统特征分类器进行情感实验, 与其对比的主要模型是 Majority, Ngram 和 Text Feature; 第二组基于循环神经网络, 结合全局注意力机制, 进行情感分析实验, 与其对比的主要模型是 AvgWordvec, SSWE, ParagraphVecto, RNTN+RNN, NSC 和 NSC+LA; 第三组是基于注意力机制的词嵌入神经网络模型的实验, 利用该模型提取文本特征并分析得到文本情感倾向性。

表 2 实验结果对比

Table 2 Comparison of experimental results

(单位: %)			
模型	IMDB	Yelp 2013	Yelp 2014
Majority	19.6	41.1	39.2
Ngram	39.9	56.9	57.7
TextFeature	40.2	55.6	57.2
AvgWordvec	30.4	52.6	53.0
SSWE	31.2	54.9	55.7
Paragraph Vector	34.1	55.4	56.4
RNTN+RNN	40.0	57.4	58.2
NSC	44.3	62.7	63.7
NSC+LA	48.7	63.1	63.0
WEAN-MAX	49.8	65.1	65.6
WEAN-AVE	49.4	64.4	65.1
WEAN-CON	49.3	64.5	65.2

分析表 2 可知, 传统方法和简单基于词向量的方法较其余两组的情感分类效果差距较大。循环神经网络模型方法使用 LSTM 构建了层次化的模型, 该模型在各个数据集上都比传统方法和基于词向量的方法有较大幅度的提升。NSC 利用双层 LSTM 构建模型, NSC 模型的实验性能较 Paragraph Vector 提升了 8% 左右。表 2 证明注意力机制有助于使模型

关注有效信息, 提升了模型的效果, NSC+LA 使用注意力改进了原模型, 取得了更优的效果。

将本文模型 WEAN 在 IMDB, Yelp 2013 和 Yelp 2014 数据集上进行了实验验证, 相比于基线模型, 其性能得到了明显的提升。在 IMDB 数据集上, 相比最好的模型的准确率, 本文模型提升了 1.1%; 在 Yelp 2013 数据集上, 相比最好的模型的准确率, 其提升了 2%; 在 Yelp 2014 数据集上, 相比比较好模型的准确率, 其提升了 2.6%。相比第二组神经网络模型, 本文模型较为简单, 参数量较少, 计算复杂度较低。并且本文将池化层作为模型的词嵌入辅助层, 使得模型在训练过程中提取的词特征有助于提升模型的性能。因此本文模型的性能效果比基线模型的性能效果更为突出。

4.3 模型分析

4.3.1 Pooling 机制对模型性能的影响

对模型中的词嵌入的辅助项 Pooling 机制进行实验验证。以本文模型中效果最好的模型 WEAN-MAX 做基础实验, 保持其他参数不变, 去掉模型中的 Pooling 机制层进行验证, 分析 Pooling 机制层是否对模型实验效果有提升作用。其中, WEAN-NP 表示在词嵌入中未加上辅助机制 Pooling 机制层, WEAN-MAX 表示加上了 Pooling 机制层。实验结果如表 3 所列, 从中可以看出没有词嵌入辅助机制比加上词嵌入辅助机制的性能效果要差, 在同种参数配置下, 在 IMDB 数据集上进行实验验证, WEAN-NP 的实验准确率为 49.0%, WEAN-MAX 的实验准确率为 49.8%, 准确率提高了 0.8%。在 Yelp 2013 数据集上进行实验, WEAN-NP 的实验准确率为 64.2%, WEAN-MAX 的实验准确率为 65.1%, 准确率提高了 0.9%。在 Yelp 2014 数据集上进行实验, WEAN-NP 的实验准确率为 64.0%, WEAN-MAX 的实验准确率为 65.6%, 准确率提高了 1.6%。由此证明词嵌入 Pooling 机制在本模型中是可以提高实验性能的。

表 3 Pooling 机制对模型性能的影响

Table 3 Influence of Pooling mechanism on model performance

(单位: %)			
模型	IMDB	Yelp 2013	Yelp 2014
WEAN-NP	49.0	64.2	64.0
WEAN-MAX	49.8	65.1	65.6

4.3.2 注意力机制对模型性能的影响

通过对模型中的注意力机制进行实验验证, 基于本文模型 WEAN-MAX 做基础实验, 保持其他参数不变, 去掉注意力机制层进行实验验证, 分析模型中注意力机制是否对模型性能有影响。其中, WEAN-NA 表示模型中未加入注意力机制层。由表 4 可以看出, 注意力机制在本实验中对模型性能的影响较明显, 在每个数据集上均提高了 5% 左右的准确率。

表 4 注意力机制对模型性能的影响

Table 4 Influence of attention mechanism on model performance

(单位: %)			
模型	IMDB	Yelp 2013	Yelp 2014
WEAN-NA	43.3	59.6	60.5
WEAN-MAX	49.8	65.1	65.6

4.3.3 上下文信息对模型性能的影响

考虑到文本的上下文信息对特征表示的影响,本文利用双向的长短期记忆网络(Bi-LSTM)对模型进行建模,由此验证上下文信息对本文文本表示的作用。保持其他参数设置不变,将模型中的循环神经网络层替换为卷积神经网络层,这是因为卷积神经网络是并行化处理数据的网络,没有考虑到上下文序列信息的作用。对比实验结果如表 5 所列,其中,WEAN-CNN 是将模型中的循环神经网络层替换为卷积神经网络模型。可以看出,双向的长短期记忆网络的模型性能比卷积神经网络的模型性能更优异,其在 3 个数据集上分别提高了 2%~3% 的准确率,由此可知,上下文信息对情感分析比较重要。

表 5 上下文信息的影响

Table 5 Influence of context information

(单位:%)

模型	IMDB	Yelp 2013	Yelp 2014
WEAN-CNN	47.5	62.5	62.8
WEAN-MAX	49.8	65.1	65.6

4.3.4 模型复杂度和训练时间的比较

本文验证本文模型的复杂性和其在实验中的模型训练时间是否有优势,以 Yelp 2013 数据集为例进行比较,WEAN-MAX 是加入了词嵌入辅助机制 Max-Pooling 层,WEAN-N 是未加入词嵌入辅助机制和注意力机制。由表 6 可以看出,加上词嵌入辅助机制层和注意力机制层的模型的总参数量和训练时间明显要少,而没有此机制的模型的训练时间相当于本文模型的双倍,由此证明了词嵌入辅助机制对本文模型的训练时间是有帮助的。

表 6 模型复杂度和训练时间的比较

Table 6 Comparison of model complexity and training time

模型	参数/M	时间消耗/s
WEAN-N	34	4816
WEAN-MAX	24	2308

4.4 案例可视化

为了证明本文模型在情感分析中的意义,本节主要对注意力层进行可视化操作。在 IMDB 数据集的测试集中选择了一个文本的部分句子进行注意力权重可视化分析,根据权重大小对句子中的词颜色进行划分,其中深颜色表示最高权重级别的词,浅颜色代表较高权重级别的词。如图 2 所示,可以看出,注意力机制在模型中确实具有发现重点词并关注重点词的作用。

this is a **stunningly beautiful** movie. the music by phillip glass is just a work of pure genius. i can watch this movie again and again.

图 2 注意力可视化

Fig. 2 Attention visualization

结束语 本文提出了一种词嵌入辅助机制的注意力神经网络模型,在原始的词嵌入的基础上加上词嵌入辅助层来对词向量进一步提取特征表示,减少了模型训练的复杂性,同时缩短了训练时间。该模型将训练时间降为原来的一半,同时其中的注意力机制的作用也比较明显。在情感分析中,由于

句子成分不同,利用注意力机制分析句子时将重点关注重要的词语。通过在多个数据集上和多个基线模型上进行的对比实验可知,本文模型的性能比基线模型的性能要好。未来工作中,我们将结合自然语言处理的其他方向继续研究神经网络模型中的词嵌入辅助功能。

参考文献

- [1] 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, Korea: ACL 2012; 90-94.
- [2] BENGIO Y. Learning Deep Architectures for AI [J]. Foundations & Trends[®] in Machine Learning, 2009, 2(1): 1-127.
- [3] SOCHER R, LIN C, MANNING C D, et al. Parsing Natural Scenes and Natural Language with Recursive Neural Networks [C]// International Conference on Machine Learning. Bellevue, Washington: ICML, 2011; 129-136.
- [4] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space [J]. arXiv: 1301.3781, 2013.
- [5] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1780.
- [6] KIM Y. Convolutional Neural Networks for Sentence Classification [J]. Empirical Methods in Natural Language Processing, 2014, 1: 1746-1751.
- [7] ZHANG Y, GAN Z, FAN K, et al. Adversarial Feature Matching for Text Generation [C]// International Conference on Machine Learning. Sydney: ICAL, 2017; 4006-4015.
- [8] GRAVES A, JAITLY N, MOHAMED A, et al. Hybrid speech recognition with Deep Bidirectional LSTM [C]// IEEE Automatic Speech Recognition and Understanding Workshop. Olomouc, Czech Republic: IEEE, 2013; 273-278.
- [9] YANG Z, YANG D, DYER C, et al. Hierarchical Attention Networks for Document Classification [C]// Northamerican Chapter of the Association for Computational Linguistics. San Diego, California: ACL, 2016; 1480-1489.
- [10] PANG B, LEE L, VAITHYANATHAN S, et al. Thumbs up? Sentiment Classification using Machine Learning Techniques [J]. Empirical Methods in Natural Language Processing, 2002, 1(1): 79-86.
- [11] TURNEY P D. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews [J]. Meeting of the Association for Computational Linguistics, 2002, 10: 417-424.
- [12] CONNEAU A, SCHWENK H, BARRAULT L, et al. Very deep convolutional networks for text classification [J]. Conference of the European Chapter of the Association for Computational Linguistics, 2017, 1: 1107-1116.
- [13] SHEN D, MIN M R, LI Y, et al. Adaptive convolutional filter generation for natural language understanding [J]. arXiv: 2017, 1709.08294.
- [14] SHI B, FU Z, BING L, et al. Learning Domain-Sensitive and Sentiment-Aware Word Embeddings [J]. Meeting of the Associ-

- ation for Computational Linguistics, 2018, 1(1):2494-2504.
- [15] ADI Y, KERMAN Y, BELINKOV Y, et al. Fine-grained Analysis of Sentence Embeddings Using Auxiliary Prediction Tasks [J/OL]. International Conference on Learning Representations, 2017. <https://openreview.net/group?id=ICLR.cc/2017/conference>.
- [16] KALCHBRENNER N, GREFFENSTETTE E, BLUNSON P, et al. A Convolutional Neural Network for Modelling Sentences [J]. Meeting of the Association for Computational Linguistics, 2014, 1:655-665.
- [17] JOHNSON R, ZHANG T. Effective Use of Word Order for Text Categorization with Convolutional Neural Networks [OL]. <http://xxx.tau.ac.il/pdf/1412.1058.pdf>.
- [18] ZHANG X, ZHAO J, LECUN Y. Character-level convolutional networks for text classification [C]// Advances in Neural Information Processing Systems. Montreal, Canada; NPIS, 2015: 649-657.
- [19] LI J, LUONG T, JURAFSKY D, et al. When Are Tree Structures Necessary for Deep Learning of Representations [J]. Empirical methods in Natural Language Processing, 2015, 1(1): 2304-2314.
- [20] LE Q V, MIKOLOV T. Distributed Representations of Sentences and Documents [OL]. <http://people.ee.duke.edu/~lcarin/ChunyuLi4.17.2015.pdf>.
- [21] TANG D, QIN B, LIU T. Document Modeling with Gated Recurrent Neural Network for Sentiment Classification [C]// Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal; ACL, 2015: 1422-1432.
- [22] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is All you Need [C]// Neural Information Processing Systems. Long Beach; NIPS, 2017: 5998-6008.
- [23] WANG Y, HUANG M, ZHU X, et al. Attention-based LSTM for Aspect-level Sentiment Classification [C]// Conference on Empirical Methods in Natural Language Processing. Austin, Texas; ACL, 2017: 606-615.
- [24] MNINH V, HEES N, GRAVES A. Recurrent models of visual attention [C]// Advances in Neural Information Processing Systems. Montreal, Canada; NPIS, 2014: 2204-2212.
- [25] BAHDANAU D, CHO K, BENGIO Y, et al. Neural Machine Translation by Jointly Learning to Align and Translate [OL]. <https://arxiv-vanity.com/papers/1409.0473>.
- [26] LUONG T, PHAM H, MANNING C D, et al. Effective Approaches to Attention-based Neural Machine Translation [J]. Empirical Methods in Natural Language Processing, 2015, 1: 1412-1421.
- [27] FAN R E, CHANG K W, HSIEH C J, et al. LIBLINEAR: A Library for Large Linear Classification [J]. Journal of Machine Learning Research, 2008, 9(9): 1871-1874.
- [28] KIRITCHENKO S, ZHU X, MOHAMMAD S M, et al. Sentiment analysis of short informal texts [J]. Journal of Artificial Intelligence Research, 2014, 50(1): 723-762.
- [29] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed Representations of Words and Phrases and their compositionality [J]. Advances in Neural Information Processing Systems, 2013, 26: 3111-3119.
- [30] TANG D, WEI F, YANG N, et al. Learning sentiment-specific word embedding for twitter sentiment classification [C]// Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. Baltimore, Maryland, USA; ACL, 2014, 1: 1555-1565.
- [31] 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, USA; ACL, 2013: 1631-1642.
- [32] LE Q, MIKOLOV T. Distributed representations of sentences and documents [C]// International Conference on Machine Learning. Beijing, China; ICAL, 2014: 1188-1196.