



文章编号:1001-9081(2019)07-1931-05

DOI:10.11772/j.issn.1001-9081.2018112375

## 融合情感与语义信息的情感分析方法

孟仕林<sup>1</sup>, 赵蕴龙<sup>1,2\*</sup>, 关东海<sup>1,2</sup>, 翟象平<sup>1,2</sup>

(1. 南京航空航天大学 计算机科学与技术学院, 南京 211106; 2. 软件新技术与产业化协同创新中心, 南京 210023)

(\*通信作者电子邮箱 zhaoyunlong@nuaa.edu.cn)

**摘要:**在使用词嵌入法进行词转向量时,两个反义词会转换成相近的向量。如果这两个词是情感词,将会导致词的情感信息的丢失,这在情感分析任务中是不合理的。为了解决这个问题,提出了一种在词嵌入的基础上增加情感向量来获取情感信息的方法。首先利用情感词典资源构建情感向量,将其与词嵌入法得到的词向量融合在一起;然后采用双向长短期记忆(BiLSTM)网络获取文本的特征;最后对文本的情感进行分类。在4个数据集上分别对该方法与未融合情感向量的方法进行了实验。实验结果表明所提方法分类准确度与F1值都高于未融合方法,说明了加入情感向量有助于提高情感分析的性能。

**关键词:**情感分析;词嵌入;情感词;情感信息;双向长短期记忆网络

**中图分类号:** TP389.1    **文献标志码:**A

### Sentiment analysis method combining sentiment and semantic information

MENG Shilin<sup>1</sup>, ZHAO Yunlong<sup>1,2\*</sup>, GUAN Donghai<sup>1,2</sup>, ZHAI Xiangping<sup>1,2</sup>

(1. School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing Jiangsu 211106, China;

2. Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing Jiangsu 210023, China)

**Abstract:** When using word embedding method for word-to-vector, two antonyms are converted into similar vectors. If they are sentiment words, it will lead to the loss of sentimental information, which is unreasonable in sentiment analysis task. To solve this problem, a method of adding sentiment vectors to obtain sentimental information based on word embedding was proposed. Firstly, the sentiment vector was constructed by using sentiment lexicon, and combined with word vector obtained by word embedding method. Then, a bidirectional Long Short Term Memory (BiLSTM) network was used to obtain the characteristics of text. Finally, the sentiment of text was classified. Experiments of the proposed method and the method without fusing sentimental vector were carried out on four datasets. The experimental results show that the classification accuracy and F1 score of the proposed method are higher than those of the method without fusion, which indicates that adding sentimental vectors is beneficial to improve the performance of sentiment analysis.

**Key words:** sentiment analysis; word embedding; sentiment word; sentiment information; bidirectional LSTM (BiLSTM)

## 0 引言

随着社交网络与电子商务的发展,网民在上网的过程中会产生大量带有情感的评论数据。这些海量的评论数据引起了越来越多学者的关注,他们可以使用网民对热点事件的评论来了解大众的舆论倾向<sup>[1]</sup>,进一步管理、引导和控制事件的发展。也可以使用购物者对商品的评论数据<sup>[2]</sup>来了解商品的质量,对未来的购物行为起到一定的参考。目前对文本评论的情感分析主要分为两种方法:基于情感词典<sup>[3]</sup>的方法与基于机器学习<sup>[4]</sup>的方法。基于情感词典的方法是利用句子中的情感词与一些影响情感的词的组合来判断句子的情感倾向,它能很好地利用情感词来反映句子的情感倾向,不过这种方法非常依赖情感词库的构建,并且同一个词在不同的领域内所表达的情感倾向可能是不同的,所以这种方法有一定

的局限性。基于机器学习的方法是利用标记好的文本训练一个分类器,其中重要的两个步骤是构造特征与选择分类器。构造特征是将词转化为能表示情感的向量,主要的方法有词袋法与词嵌入法。分类器主要有支持向量机(Support Vector Machine, SVM)、朴素贝叶斯(Naive Bayes)、卷积神经网络(Convolution Neural Network, CNN)与长短期记忆(Long Short Term Memory, LSTM)网络<sup>[5-8]</sup>。

近几年,深度神经网络在图像识别与语音识别领域有了较大的突破,于是大家也将这一方法应用到情感分析的研究上。Bengio等<sup>[9]</sup>利用神经网络语言模型学习每个词的分布式表示与词序列的概率函数,将每个词转化成低维、连续的向量。Mikolov等<sup>[10]</sup>通过对频繁出现的词进行子抽样来加速词向量的训练,同时也提出了负采样的方法。通过语言模型获得的词向量在一定程度上能表达句子的语义信息,它能很好地代

收稿日期:2018-12-03;修回日期:2019-02-17;录用日期:2019-02-19。    基金项目:国防基础科研计划资助项目(JCKY2016605B006);江苏省“六大人才高峰”高层次人才项目(XYDXXJS-031);中央高校基本科研业务费专项(90YAH16042)。

**作者简介:**孟仕林(1994—),男,安徽界首人,硕士研究生,CCF会员,主要研究方向:数据挖掘、情感分析;赵蕴龙(1975—),男,黑龙江哈尔滨人,教授,博士,CCF会员,主要研究方向:无线网络、群体计算、说话人识别、数据挖掘、可穿戴计算技术;关东海(1981—),男,黑龙江哈尔滨人,副教授,博士,CCF会员,主要研究方向:语音识别、数据挖掘、人工智能;翟象平(1984—),男,山东淄博人,副教授,博士,CCF会员,主要研究方向:普适计算、蜂群无人系统、群体智能、数据挖掘。



替词袋法来解决自然语言处理中的一些问题。不过针对情感分析这个问题,词向量并不能反映出一个词的情感倾向,比如“好”与“坏”这两个词的词向量非常相似,这是不太合理的,因此,有一些学者开始研究如何利用句子中表达情感信息的成分来提高情感分类效果,比如句子中包含的表情符号和一些情感词<sup>[1,6,11]</sup>。

循环神经网络(Recurrent Neural Network, RNN)是一种具有记忆功能的神经网络,适合序列数据的建模,它在语音识别、自然语言处理等领域取得了成功。Yang 等<sup>[12]</sup>提出了基于注意力机制的网络,改进了 6 种文本分类任务的效果;Wang 等<sup>[13]</sup>通过使用基于注意力的长短期记忆网络来处理 Aspect-level 的情感分析问题;Alec 等<sup>[14]</sup>通过使用 CNN 与 LSTM 层的多个分支的组合核来描述一种新的情感分析方法,在互联网电影数据库(Internet Movie DataBase, IMDB)评论情绪数据集中取得了最高的准确度。本文使用双向长短期记忆(Bidirectional Long Short Term Memory, BiLSTM)网络分类器进行情感的预测,它用两个不同的循环层分别从正向和反向对数据进行扫描,不仅利用了过去的数据,还利用了未来的数据,进一步提高了网络性能。

综上所述,本文提出了一种融合情感与语义信息的情感分析方法,即利用情感词的情感信息构造出情感向量,与语言模型生成的向量(称之为语义向量)相结合来表示文本,使用双向 LSTM 作为分类器,充分学习文本的序列特征,对商品评论与电影评论数据进行情感分析。

## 1 词嵌入法

词嵌入是将词映射到低维密集向量的方法,解决了使用词袋法带来的稀疏问题。它是从大量文本语料中以无监督的方式学习语义知识的一种模型,通过学习文本来用词向量的方式表征词的语义信息,即通过一个嵌入空间使得语义上相似的单词在该空间内距离很近。在众多词嵌入方法中,skim-gram 模型是使用较广泛的一种,skim-gram 使用输入层、隐含层、softmax 层三层神经网络训练出一个模型,根据词共现法将句子中的词构造成(输入词、输出词)的单词对。再将单词对转成 one-hot 编码的形式输入到神经网络中训练,模型采用了负采样的方法选择性地更新权重矩阵,降低了训练代价并且改善所得到词向量的质量。网络训练完成后,获得隐含层的权重矩阵  $W^{n \times d}$ , 权重矩阵的每一行就代表一个词的词向量,这里本文称为语义向量,  $n$  表示语料中词的个数,  $d$  表示词向量的维数。经过词嵌入方法的转化,词和文本的语义向量表示如下:

$$w_i = \text{row} (W, i) \quad (1)$$

$$S_w = [w_1, w_2, \dots, w_n] \quad (2)$$

其中: $w_i$  表示第  $i$  个词的语义向量,  $S_w$  表示文本文本的语义向量,  $\text{row}$  代表取行向量。通过 skim-gram 模型得到的词向量可以很好地捕捉语义单元之间潜在的语义关系,且训练语料越多,词向量蕴含意义越丰富。

## 2 LSTM 网络

长短期记忆(LSTM)网络是循环神经网络的一种,它既保留了循环神经网络处理序列数据的能力,又解决了序列学习过程时存在的长期依赖问题,避免梯度消失和梯度爆炸现象。

LSTM 通过“记忆单元”和“门”控制器的巧妙设计实现序列数据学习能力的提升,它由记忆单元、输入门、遗忘门和输出门四个主要元素组成。其中:遗忘门  $f$  决定记忆单元前一个状态信息的留存,输入门  $i$  控制记忆单元中当前时刻信息的输入,记忆单元  $c$  根据当前输入信息更新记忆状态,再由输出门  $o$  判断记忆单元对下一个状态的输出结果。下面是 LSTM 的结构和计算公式:

$$f_t = \sigma(W_f[\mathbf{h}_{t-1}, \mathbf{x}_t] + b_f) \quad (3)$$

$$i_t = \sigma(W_i[\mathbf{h}_{t-1}, \mathbf{x}_t] + b_i) \quad (4)$$

$$\tilde{c}_t = \tanh(W_c[\mathbf{h}_{t-1}, \mathbf{x}_t] + b_c) \quad (5)$$

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \quad (6)$$

$$o_t = \sigma(W_o[\mathbf{h}_{t-1}, \mathbf{x}_t] + b_o) \quad (7)$$

$$\mathbf{h}_t = o_t * \tanh(c_t) \quad (8)$$

其中: $\mathbf{h}_t$  表示  $t$  时刻记忆单元的状态,  $\mathbf{x}_t$  表示  $t$  时刻的信息输入,  $W$  是权重矩阵,  $b$  是函数的偏置项,  $\sigma$  是 sigmoid 函数。

双向长短期记忆模型由两层 LSTM 模型组成:一层学习顺序输入数据的特征,另一层学习逆序输入数据的特征,这种结构使得双向 LSTM 可以更充分学习输入序列数据中词语的上下文信息,在后续内容中 BiLSTM 即代表双向 LSTM。

## 3 情感与语义信息融合方法

情感与语义信息融合方法是将语义向量与情感向量结合作为情感分析任务的输入,利用双向 LSTM 分类器对标记好的数据进行训练,生成情感分类器。主要包括:生成语义向量、生成情感向量、语义向量融合情感向量。

### 3.1 生成语义向量

本文使用 Google 开源的工具 Word2vec<sup>[15]</sup> 获取语义向量,以新闻数据作为训练语料,共现窗口设置为 5,输出向量维数设为 100。经过 Word2vec 工具训练之后每个词被转化成 100 维的语义向量。在对文本进行数字化表示的时候,由于文本的长度长短不一,为了保持文本序列的一致性,本文取前 100 个词的语义向量,如果文本的词汇数量小于 100,不足的词数使用等维度的全 0 向量补充,这样一段文本就可以用 100 个语义向量序列表示。

### 3.2 生成情感向量

使用 Word2vec 工具生成的语义向量能很好表示词语的语义信息,但是缺少了词的情感信息。为了弥补这一缺点,本文针对此情感倾向构造出情感向量。受到基于情感词典方法的启发,本文发现影响文本情感的词不仅仅有积极情感词与消极情感词,还包括一些否定词、转折词与主张词,这些词按照一定的语法结构综合地影响文本的情感倾向,因此本文将词分成 6 大类,分别是积极词、消极词、中性词、否定词、转折词与主张词,采用类似 one-hot 编码的方式构造一个 6 维的向量,这 6 维代表着 6 类词语,如果一个词属于其中的一类,则相应位置的数值置为 1,其余置为 0。比如“喜欢”这个词的词性是属于积极的,那它的情感向量就表示为[1,0,0,0,0,0]。

上述方法生成的情感向量包含词的情感信息,然后将其应用到情感分析任务中。在对词进行分类的步骤中,本文参考了台湾大学的 NTUSD (National Taiwan University Sentimental Dictionary) 简体中文情感词典中的积极词词典、消极词词典,并构造了常用的否定词词典、转折词词典和主张词词典,一共 5 种词典。利用上述 5 种词典将文本中的词进行



分类,如果一个词不存在于5种词典中的任何一种,则该词被认为是中性词。一段文本经过上述的方法可以表示成情感向量序列:

$$e_i^n = \begin{cases} 1, & w \in dict_n \\ 0, & w \notin dict_n \end{cases} \quad (9)$$

$$S_E = [e_1, e_2, \dots, e_n] \quad (10)$$

其中: $e_i$ 表示第*i*个词*w*的情感向量, $dict_n$ 表示第*n*个情感词典, $S_E$ 表示文本的情感向量序列。同语义情感向量的操作一样,选取序列中的前100个情感向量表示该文本,不足100的使用等维度的全0向量补齐。进一步地考虑每一类词语对情感倾向的影响效果有所不同,比如转折词的影响程度就大于主张词的影响程度,因此对每一类词都赋予不同的权重,并且对于情感向量中为0的值使用 $[-0.2, 0.2]$ 区间的一个随机值来代替,降低情感向量的稀疏性。

### 3.3 语义向量融合情感向量

生成了词的语义向量与情感向量之后,如何融合这两种向量是一个非常重要的问题。本文提出了两种融合方法:一种是在神经网络的输入层将语义向量和情感向量串联起来构成一个词的词向量;一种是在神经网络的双向LSTM层将语义向量的输出结果与情感向量的输出结果串联在一起。这两种融合方法分别对应不同的网络结构,在下面的公式与图中,“+”表示串联操作。

第一种在输入层将语义向量与情感向量融合,每一词就由100维的语义向量和6维的情感向量串联表示,一共106维,一段评论文本*S*就由100个106维的向量序列表示。将融合后的向量序列输入到双向LSTM网络中进行情感分类:

$$S = S_W + S_E = [w_1 + e_1, w_2 + e_2, \dots, w_{100} + e_{100}] \quad (11)$$

第二种是将语义向量序列与情感向量序列分别输入到双向LSTM网络中,网络的输出结果进行融合再进入全连接层中,最后经过输出层输出情感分析结果。融合的方式也有许多种,本文经过实验探讨了串联、求和与求平均三种融合方式的效果,结果发现采用串联的融合效果最好。

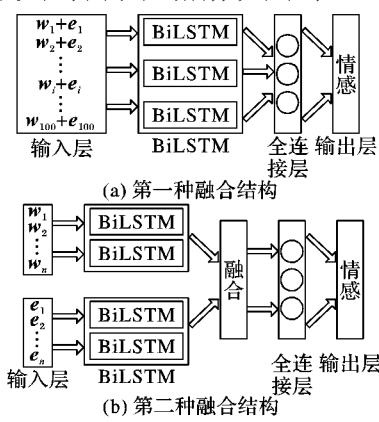


图1 两种融合方法的网络结构

Fig. 1 Network structure of two fusion methods

## 4 实验

### 4.1 实验数据

本次实验的数据集共有4个,如下所示:

1)商品评论数据。从淘宝商场平台上爬取,涉及到五类商品,评论的积极性与消极性在平台上已经作好了标注。

2)电影影评数据。从豆瓣网上爬取,本文将影评中评分为5星的评论设置为积极评论,评价为1星的设置为消极评论。

3)酒店评论数据。谭松波<sup>[16]</sup>收集整理了一个较大规模的酒店评论语料,语料从携程网上自动采集,并经过整理而成。

4)自然语言处理及中文计算会议(Natural Language Processing and Chinese Computing, NLPCC)2014任务2数据。该语料共包含中文和英文两种语言,本文使用其中的中文数据,主要是商品评论,可以被应用于篇章级或者句子级的情感分析任务。

每个数据集都按2:8的比例分成训练集与验证集,训练集用于训练情感分类模型,测试集用于评价模型效果。4个数据集的详细信息如表1所示。

表1 实验数据信息

Tab. 1 Experimental data information

数据集	情感		总计	划分	
	积极	消极		训练集	测试集
商品评论	12 003	12 000	24 003	19 202	4 801
电影评论	8 211	6 580	14 791	11 832	2 959
酒店评论	7 000	3 000	10 000	8 000	2 000
NLPCC2014	5 000	5 000	10 000	8 000	2 000

在构建词的情感向量时,本文构造了5个词典,其中积极词词典、消极词词典参考了NTUSD台湾大学简体中文情感词典,本文方法将词典中较长的短语去除,加入了一些常用的带有情感倾向的评论词。否定词词典、转折词词典与主张词词典是从常用的中文文本中总结得到的。这些词典的词汇量如表2。

表2 词典词汇量

Tab. 2 Vocabularies of lexicons

词典	词汇量	举例
积极词	3 452	美观、自信、轻巧
消极词	8 201	丑、生硬、失望
否定词	33	不是、没有、并非
转折词	13	可是、不过、然而
主张词	38	感觉、发现、认为

### 4.2 实验设计

本文采用语义向量+SVM、语义向量+双向LSTM、语义向量+CNN、语义+情感向量+双向LSTM、语义向量+分层注意力网络HN-ATT(Hierarchical ATTention Networks)这几个模型进行情感分析实验。

1) SVM。支持向量机是传统机器学习算法中较为常用的一种,它通过寻找一个超平面将数据进行分类。本文将文本序列的语义向量的平均值作为特征,SVM模型作为分类器,来研究传统机器学习算法用于情感分类的效果。

2) CNN。卷积神经网络在图像识别领域获得较大的成果,它能够更好地获取图像的局部特征。本文采用文本的语义向量序列为特征,CNN为分类器,研究卷积神经网络在本文情感分类的效果。

3) 双向LSTM。长短期记忆网络作为RNN的一种改进,对处理序列数据有较好的性能。本文使用文本的语义向量序列为输入,双向LSTM网络为分类器,通过实验研究循环神



经网络对文本序列的处理效果。

4) 语义 + 情感向量 + 双向 LSTM。作为本文提出的情感信息融合方法,实验中将情感向量与语义向量分别作为特征输入到模型中,采用 3.3 节中的两种方法进行融合,双向 LSTM 网络为分类器,通过实验研究“情感融合”的有效性。

5) 语义向量 + HN-ATT。文献[11]中提出的层次注意力网络,通过注意力机制对每个词赋予不同的权重,实验中将语义向量作为特征输入。

#### 4.3 实验流程

##### 4.3.1 文本预处理

文本是由词组成的,首先要将文本分割成词的序列。实验中使用 jieba 分词工具对文本进行分词,之后对每一个词进行词转向量操作,经过上述两步之后,一个评论文本就转化成一段词向量序列,取前 100 个词向量表示这个文本,将其输入到模型中去。

表 3 不同模型的实验结果

Tab. 3 Experimental results of different models

情感分析模型	商品评论数据		电影评论数据		酒店评论数据		NLPCC2014 数据	
	accuracy	F1	accuracy	F1	accuracy	F1	accuracy	F1
SVM	0.9044	0.9032	0.7854	0.8103	0.8558	0.8994	0.8519	0.7492
CNN	0.9119	0.9149	0.8297	0.8439	0.8879	0.9221	0.9158	0.9177
BiLSTM	0.9317	0.9313	0.8645	0.8809	0.8914	0.9230	0.9294	0.9322
M-BiLSTM1	<b>0.9379</b>	<b>0.9368</b>	<b>0.8777</b>	<b>0.8903</b>	0.9044	0.9327	<b>0.9390</b>	<b>0.9397</b>
M-BiLSTM2	0.9367	0.9360	0.8723	0.8843	0.9069	0.9336	0.9360	0.9373
HN-ATT	0.9365	0.9359	0.8732	0.8851	<b>0.9114</b>	<b>0.9357</b>	0.9384	0.9390

另外,为了研究情感向量维数和情感资源对分类效果的影响,又分别设计了以下两种实验:1) 构建三维的情感向量,每一维表示积极词、消极词和中性词,除去否定词、转折词与主张词这三维信息。模型采用“语义 + 情感向量”的双向 LSTM 模型。2) 情感向量仍然使用 6 维,但是情感词典的容量减少为原来的一半,使用的模型不变。

#### 4.4 实验结果与分析

实验中使用 Keras 深度学习框架搭建网络模型,分别采用商品评论和电影评论作为实验数据来测试 4 种模型的效果。

图 2 展示的是几种神经网络模型在电影影评数据上测试的 accuracy-epochs 变化。图中 M-BiLSTM1、M-BiLSTM2 表示融合了情感向量与语义向量的两种模型,可以看出本文提出的两种融合模型的测试效果要优于 BiLSTM 模型与 CNN 模型。

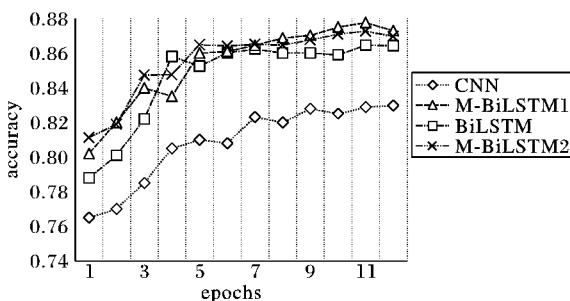


图 2 不同模型的 accuracy-epochs 变化

Fig. 2 Accuracy-epochs changes of different models

从表 3 中可以发现:SVM 模型与 CNN 和 BiLSTM 模型相

#### 4.3.2 模型构建与测试

实验一共构建 5 种分类模型,SVM 模型接收语义向量的平均值,使用 RBF(Radial Basis Function)核函数构建超平面进行分类。CNN 采用卷积提取数据的特征,实验中卷积核大小为  $3 \times 3$ ,卷积核个数为 200。LSTM 网络的输出单元为 200,双向 LSTM 输出结果的融合方式为取平均值。HN-ATT 网络中门控制单元 GRU(Gated Recurrent Unit)的维数为 100。在 CNN、双向 LSTM 模型和 HN-ATT 模型中都加入了 maxpooling 层与 dropout 层,maxpooling 层的池化大小为 3,步长为 3,dropout 的参数设置为 0.5;这三个模型训练时采用的损失函数为交叉熵损失函数,优化器为 Adam,训练的 batch-size 为 64,epochs 为 12。实验将数据按 8:2 的比例分成训练集与测试集,训练集用于训练模型,测试集测试模型的分类性能,采用的评价指标为模型分类的准确度 accuracy 与 F1 得分值(如表 3)。

比,其分类的准确度与 F1 得分值都较低,说明传统的机器学习方法在情感分析中表现的性能要低于深度学习方法。CNN 模型与 BiLSTM 模型相比较起来,BiLSTM 的准确度与 F1 得分值都略高于 CNN,说明双向 LSTM 网络能更好地获取到文本的上下文信息,更适合于处理序列数据。融合了情感向量与语义向量的模型与 BiLSTM 模型相比,性能略高。其中在商品评论数据集的测试中,M-BiLSTM1 比 BiLSTM 准确度高了 0.62 个百分点,F1 提高了 0.55 个百分点;在电影评论数据集的测试中,M-BiLSTM1 比 BiLSTM 准确度提高了 1.32 个百分点,F1 提高了 0.94 个百分点;在酒店评论数据的测试中,M-BiLSTM2 比 BiLSTM 准确度提高了 1.55 个百分点,F1 提高了 1.06 个百分点;在 NLPCC2014 数据的测试中,M-BiLSTM1 比 BiLSTM 准确度提高了 0.96 个百分点,F1 提高了 0.75 个百分点;由此可以看出加入了情感向量对分类性能有了一定的提升。在本文的方法与 HN-ATT 方法的对比中,商品评论、电影评论、NLPCC2014 三个数据集上的实验效果高于 HN-ATT,酒店评论数据的实验效果低于 HN-ATT;说明本文提出的方法在大部分数据上取得了较好的实验效果。

在研究情感向量维数和情感资源对分类效果的实验时,使用的数据集为电影评论数据,模型的融合方式为第一种融合方式。得到的实验结果如表 4 所示。从中可以看出:1) 使用 3 维情感向量的分类效果略低于 6 维,缺少了否定词、转折词与主张词这 3 维的信息,情感分类的效果有所降低。说明否定词、转折词与主张词都会对情感分类产生影响。2) 将情感词典容量减半会降低分类的效果,由于情感词典缺少对词的分类,生成的情感向量就丢失了情感信息。如果情感词典的资源足够丰富和准确,那么情感向量就能更好地提高分类



效果。

表4 情感资源与向量维数的实验结果

Tab. 4 Experimental results of sentimental resources and vector dimensions

情感向量构造	accuracy	F1
3维向量 + 全部情感资源	0.8729	0.8882
6维向量 + 一半情感资源	0.8696	0.8830
6维向量 + 全部情感资源	<b>0.8777</b>	<b>0.8903</b>

## 5 结语

在情感分析的研究任务中,采用词嵌入法+深度神经网络是目前常用的做法。本文提出的融合情感与语义信息的方法是在词嵌入法生成词向量的基础上,根据词的情感信息构建了情感向量,将二者融合一起应用到情感分析任务中。通过两个数据集的实验验证,该方法能够提高情感分析的性能,说明词的情感信息对文本的情感极性判断是有影响的;同时实验中还比较了卷积神经网络与长短时记忆网络在情感分析中的性能,结果得出长短时记忆网络更适合处理类似文本的序列数据。通过实验本文发现词的情感信息对情感分析任务有一定影响,在今后的工作中将研究基于主题和目标的细粒度的情感分析方法。

### 参考文献 (References)

- [1] 吴鹏,应杨,沈思. 基于双向长短期记忆模型的网民负面情感分类研究[J]. 情报学报, 2018, 37(08): 845–853. (WU P, YING Y, SHEN S. Negative emotions of online users' analysis based on bi-directional long short-term memory[J]. Journal of the China Society for Scientific and Technical Information, 2018, 37(8): 845–853.)
- [2] 赵刚,徐赞. 基于机器学习的商品评论情感分析模型研究[J]. 信息安全研究, 2017, 3(2): 166–170. (ZHAO G, XU Z. Research on the sentiment analysis model of product reviews based on machine learning[J]. Journal of Information Security Research, 2017, 3(2): 166–170.)
- [3] TETSUYA N, JEONGHEE Y. Sentiment analysis: Capturing favorability using natural language processing [C]// Proceedings of the 2nd International Conference on Knowledge Capture. New York: ACM, 2003: 70–77.
- [4] WEI J, HUNG H H, ROHINI K S. OpinionMiner: a novel machine learning system for Web opinion mining and extraction [C]// Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2009: 1195–1204.
- [5] ZOU H, TANG X, XIE B, et al. Sentiment classification using machine learning techniques with syntax features [C]// Proceedings of the 2015 International Conference on Computational Science and Computational Intelligence. Piscataway, NJ: IEEE, 2015: 175–179.
- [6] YIN R, LI P, WANG B. Sentiment lexical-augmented convolutional neural networks for sentiment analysis [C]// Proceedings of the 2017 IEEE 2nd International Conference on Data Science in Cyberspace. Piscataway, NJ: IEEE, 2017: 630–635.
- [7] 於雯,周武能. 基于LSTM的商品评论情感分析[J]. 计算机系统应用, 2018, 27(8): 159–163. (YU W, ZHOU W N. Sentiment analysis of commodity reviews based on LSTM[J]. Computer Systems and Applications, 2018, 27(8): 159–163.)
- [8] 丁晨春,吴靓婵媛,李红梅. 基于SVM的中文微博观点倾向性识别[J]. 情报学报, 2016, 35(12): 1235–1243. (DING S C, WU J C Y, LI H M. SVM-based Chinese microblogging viewpoint orientation recognition[J]. Journal of the China Society for Scientific and Technical Information, 2016, 35(12): 1235–1243.)
- [9] BENCIO Y, DUCHARME R, VINCENT P, et al. A neural probabilistic language model[J]. Journal of Machine Learning Research, 2003, 3: 1137–1155.
- [10] 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.
- [11] 何炎祥,孙松涛,牛菲菲,等. 用于微博情感分析的一种情感语义增强的深度学习模型[J]. 计算机学报, 2017, 40(4): 773–790. (HE Y X, SUN S T, NIU F F, et al. An emotional semantic enhanced deep learning model for microblog emotion analysis [J]. Chinese Journal of Computers, 2017, 40(4): 773–790.)
- [12] 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. Stroudsburg, PA: Association for Computational Linguistics, 2016: 1480–1489.
- [13] WANG Y Q, HUANG M L, ZHAO L, et al. Attention-based LSTM for aspect-level sentiment classification [C]// Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: Association for Computational Linguistics, 2016: 606–615.
- [14] ALEC Y, ABHISHEK. V. Deep CNN-LSTM with combined kernels from multiple branches for IMDB review sentiment analysis [C]// Proceedings of the 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference. Piscataway, NJ: IEEE, 2017: 540–546.
- [15] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space [C]// Proceedings of the 2013 International Conference on Learning Representations. Scottsdale, Arizona: [s. n.], 2013: 1–12.
- [16] 谭松波. 酒店评论数据[EB/OL]. [2018-10-20]. <http://www.datatang.com/data/11970>. (TAN S B. Hotel review data [EB/OL]. [2018-10-20]. <http://www.datatang.com/data/11970>.)

This work is partially supported by the National Defense Basic Scientific Research Program (JCKY201605B006), the Six Talent Peaks Project of Jiangsu Province (XYDXJS-031), the Fundamental Research Funds for the Central Universities (90YAH16042).

**MENG Shilin**, born in 1994, M. S. candidate. His research interests include data mining, sentiment analysis.

**ZHAO Yunlong**, born in 1975, Ph. D., professor. His research interests include wireless network, crowd computing, speaker recognition, data mining, wearable computing technology.

**GUAN Donghai**, born in 1981, Ph. D., associate professor. His research interests include speech recognition, data mining, artificial intelligence.

**ZHAI Xiangping**, born in 1984, Ph. D., associate professor. His research interests include ubiquitous computing, swarm unmanned system, collective intelligence, data mining.