

Word Attention-Based BiLSTM and CNN Ensemble for Chinese Sentiment Analysis

Kai Sun

School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming Yunnan
Email: 892752842@qq.com

Received: Feb. 2nd, 2020; accepted: Feb. 17th, 2020; published: Feb. 24th, 2020

Abstract

Film comments as an important way of expressing emotions in audience viewing, and an important field of emotional mining research. This paper proposes a stacking model based on word attention mechanism BiLSTM and CNN (Word Attention-Based BiLSTM and CNN Stacking Model, WABCSM), and uses the Chinese commentary of the Douban platform as the research object to analyze the emotional tendency of fans to the film. The feasibility of the model. Firstly, the word vector is used to train the text, and the word vector representation is obtained. Then, the fusion model based on word attention mechanism BiLSTM and CNN is used to perform emotion mining to better extract the emotional words in the text to achieve the purpose of correct classification. The model of this paper is valid. Compared with the traditional LSTM model and CNN model, the experimental results show that the classification accuracy and recall are improved. In particular, the addition of the attention mechanism can make the model effectively extract the emotional features, thus effectively overcoming the difficult problem of colloquial short text emotional polarity judgment.

Keywords

Sentiment Analysis, Word2vec, WABCSM, Word Attention

基于词注意力的BiLSTM和CNN集成模型的中文情感分析

孙 凯

云南财经大学统计与数学学院, 云南 昆明
Email: 892752842@qq.com

收稿日期: 2020年2月2日; 录用日期: 2020年2月17日; 发布日期: 2020年2月24日

摘要

电影评论作为观众观影的重要情感流露方式,是情感挖掘研究的一个重要领域。本文提出了一种基于词注意力机制的BiLSTM和CNN集成模型(Word Attention-Based BiLSTM & CNN Stacking Model, WABCSM),以豆瓣平台的电影中文评论为研究对象,分析影迷对电影的情感倾向,来论证本文模型的可行性。首先利用word2vec训练文本,得到文本的词向量表示,然后使用基于词注意力机制的BiLSTM和CNN集成模型来进行情感挖掘以更好的提取文本中的情感词,达到正确分类的目的。本文模型相比较传统的LSTM模型和CNN模型,实验结果显示在分类精准率和召回率上都有很大提升。尤其是注意力机制的添加,使得本文模型对情感特征也可以进行有效的提取,从而有效克服了口语化短文本情感极性判断的难点问题。

关键词

情感分析, Word2vec, WABCSM, 词注意力机制

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着互联网技术的发展,越来越多的网友更愿意在网上发表自己的观点,比如在微博上发表对某件事的看法或是观看电影后书写自己的观影体验。尤其是近年来国内电影行业的快速发展,更多的人选择在休闲时间和家人或朋友去电影院观看电影,对影评进行情感挖掘显得越来越重要。这些影视评论一方面影响着其他用户体验,同时也会将影迷观感反馈给影视公司,这对于后面的电影制作会有很大的帮助。电影评论主要是针对故事情节和演员的演技,所以评论会有比较大的多变性,更加偏向口语化和无规则性,尤其是一词多意,这就让问题变得复杂。很多的实验表明,针对中文影评的情感分析模型准确率相对较低。本文通过集成思想设计一种融合模型,并且爬取豆瓣平台的电影评论作为研究对象,来论证本文模型的有效性。

2. 相关工作

情感分析,也称为意见挖掘,可以从书面语言中分析出人们的情感、观点和态度。情感分析方法主要包括两种:基于规则和基于机器学习[1]。

基于规则的方法依赖于情感词典的构建,即将分词后的文档与情感词典的词进行匹配,并统计匹配成功的情感词数量。肖江等为了能够更加准确的识别情感词,在基准情感词的前提下构建了相关领域情感词典[2]。Paltoglou 等于 2012 年采用基于情感词典的方法进行情感分析,利用否定词、程度词、情感极性 etc 对微博语料进行情感极性判别[3]。对于不同领域的情感分析,使用领域情感词典具有更好的效果。Jo 等将主题和情感同时标签,利用句子的主题代替词的主题标签进行采样,以此缩小词语之间的主题关联[4]。基于情感词典的方法,可以节省很多时间,但是在不同领域的情感分类,需要该领域下的情感词典,模型迁移上效果很差,准确率较低。

基于机器学习的方法可以达到更高的准确率,但是在机器学习训练模型前,需要对文本语料进行标

注。早期, Pang 等首次应用机器学习方法解决情感分析问题, 他们尝试使用 N-gram 模型提取特征, 并取得很好的效果[5]。国内的一些学者利用机器学习的方法进行情感分类, 比如 liu 等先利用协同过滤的方法对文本进行半标注, 而后利用支持向量机算法进行分类[6]。对于电影评论, Darger A L 等结合 bayes 和 KNN 两种算法, 并得出了 bayes 分类准确率高于 KNN 算法的分类准确率的结论[7]。随着机器学习的发展, 词向量被提了出来, one hot 方法得到广泛的应用, 但是 one hot 方法具有无法表示词语之间的相关性和维度矩阵的稀疏性的问题。而后 Hiton 提出了词向量(Word Embedding)方法, 可以很好的将文本语法结构和句法结构映射到向量的不同纬度, 从而解决了向量稀疏问题, 并且向量空间也可以很好地反映文本语义之间的相关性[8]。Le 等基于词向量的方式进行优化, 构建句子或短语的词向量表示, 从而克服了 one hot 本身的缺点[9]。

传统的机器学习算法具有良好的性能, 前提是需要对文本做一定的处理, 即传统的机器学习算法对特征有一定的要求。因此, 情感分析的表现很大程度上依赖于文本特征的选择, 这一过程需要大量的人力物力。深度学习通过构造深层次的神经网络结构来学习数据特征和数据分布, 从而可以在文本中自动去学习特征, 这相对于传统机器学习, 会节约很多时间。深度学习已经在许多自然语言处理任务中取得了巨大的成功。Collobert 等首次将卷积神经网络(convolutional neural network, CNN)引入了自然语言处理的领域, 并取得了较好的结果[10]。王煜涵等利用卷积神经网络(convolutional neural network, CNN)对 Twitter 文本进行情感分析, CNN 算法在提取文本的局部特征上表现了良好的性质, 实验表明 CNN 模型在情感分类上的表现要比传统的机器学习算法更好[11]。

注意力机制目前已经得到广泛运用, 在文本处理领域, 注意力机制的可以使得模型有选择的对最优特征进行提取, 提高分类效率。Quanzeng You 等在 2016 提出了一种基于注意力机制的 LSTM 的情感分析机制, 通过提取长短文本特征, 并结合注意力机制进行特征筛选来达到分类的目的[12]。关鹏飞等运用 LSTM 模型结合注意力机制应用于情感分析, 提高了模型训练的准确率[13]。王盛玉等结合 CNN 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[14]。

虽然有很多深度学习模型应用在情感分析领域, 但是在情感分析领域的大部分的模型都是单一的。卷积神经网络最早被应用于图片处理领域, 提取局部特征具有好的表现, 并取得了巨大的成功。实验表明 CNN 在文本处理领域也具有很好的效果, 但 CNN 模型在捕捉文本的语义关系时表现较差, 对于电影短文本, 对语义的理解很重要, 所以为了提高 CNN 算法的准确度, 本文在 CNN 模型的基础上加入了对文本语义特征提取具有较好效果的加有词注意力机制的 BiLSTM 模型, 提出基于词注意力机制的 BiLSTM 和 CNN 模型的集成模型(Word Attention-based BiLSTM & CNN Stacking Model, WABCSM)。

本文模型主要完成以下任务: (1) 词嵌入层。利用 word2vec 将文本结构化, 作为模型的输入。(2) BiLSTM 层。考虑到文本中情感词不仅与前面出现的内容相关, 还与后面的内容相关, 所以采用 BiLSTM。BiLSTM 模型对文本的语义理解有很好的效果, 将词向量的输入进行文本语义的理解和记忆。(3) 注意力机制。注意力机制的引入, 用来加强文本中情感词的权重。(4) CNN 层。卷积操作用来获得 n-gram 特征, 用于最终的情感分类。本文通过 WABCSM 模型在中文影评数据集上做了验证, 并且与传统的深度学习模型进行比较, 取得了很好的结果。

3. 基于词注意力机制的 BiLSTM 和 CNN 集成模型的中文情感分析

本文的模型采用集成的思想将 BiLSTM 模型与 CNN 模型进行模型融合, 一方面通过 BiLSTM 进行文本的语义提取, 结合 CNN 模型, 提取文本关键词。为了提高模型的准确度, 在 BiLSTM 层后加入了 Attention 层, 目的是增加文本关键词的权重, 使得模型可以更好的提取关键词, 来达到情感分类的目的。本文模型结构如图 1 所示。

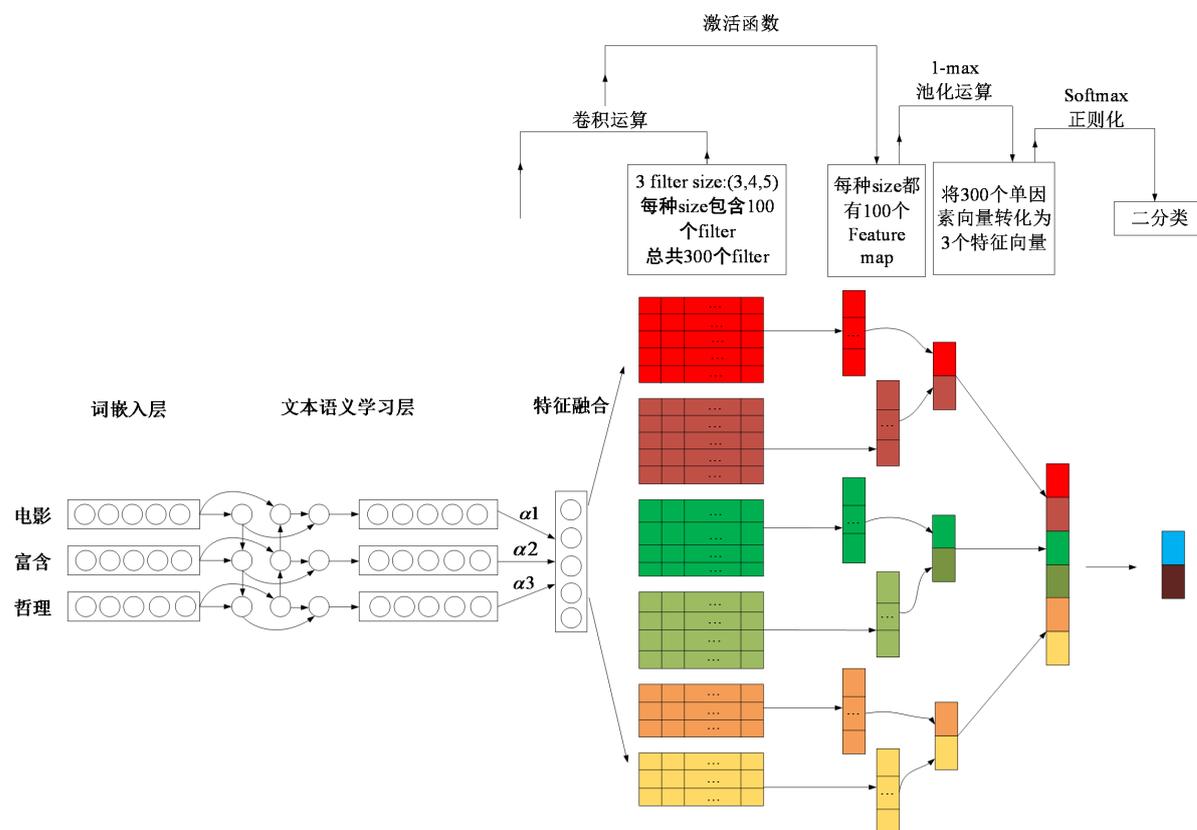


Figure 1. Integration model architecture
图 1. 集成模型架构

3.1. 词嵌入层

本文词嵌入层的输入是经过 jieba 分词后的结果，包括词和字。如何正确的分词对模型的训练有很大的影响。比如“电影富含哲理”可以分成“电影”，“富含”，“哲理”，也可以分成“电”，“影”，“富”，“含”，“哲”，“理”。这就加大了难度，本文考虑到中文的特殊性，采用第一种分词方式。对于分词之后的文本要进行结构化。传统的稀疏向量表示存在矩阵维度过大的缺点，无法表示出文本语义之间的联系。Hinton 于 1986 年提出分布式表示(distributed representation)的方法很好地解决了 One hot 的不足[8]。目前比较常用的 word2vec 方法就是属于分布式表示。Word2vec 可以将文本通过低纬度稠密的向量来表示，不仅解决了稀疏性的问题，而且分布式方法可以很好的表示语义之间的关系。word2vec 将语料训练成 $R^{D \times s}$ 规模的词向量矩阵， D 表示字典的大小， s 表示词向量的维度。假如 x_i 表示文本的第 i 个词，并且文本长度为 L ，则输入的向量可以表示为

$$S = [x_1; x_2; \dots; x_L] \in R^{L \times d} \quad (1)$$

例如本文语料库为“超级，好看，准备，二刷”，其数字表示就是{1: 超级 2: 好看 3: 准备 4: 二刷}，文本即转为数字表示，结构化(1, 2, 3, 4)就表示文本“超级好看准备二刷”。将结构化语言(1, 2, 3, 4)进行 word2vec 训练，词向量维度是预定好的。word2vec 分为 CBOW 模型(continuous bag-of-words model)和 skip-gram 模型(continuous skip-gram model)，本文的词向量表示是基于 skip-gram 模型，Skip-Gram 模型的网络结构分为：输入层、投影层和输出层，原理是以当前词 $w(t)$ 对其上下文 $Context(w)$ 中的词进行预测。

本文 Skip-Gram 模型更新参数的伪代码:

```

 $\varepsilon = 0$ 
FOR  $u \in \text{Context}(w)$  DO
{
  FOR  $j = 2 : l^u$  DO
  {
     $q = \sigma(v(w)^T \theta_{j-1}^u)$ 
     $g = \eta(1 - d_j^u - q)$ 
     $\theta_{j-1}^u := \theta_{j-1}^u + gv(w)$ 
     $e := e + g\theta_{j-1}^u$ 
  }
}
 $v(w) := v(w) + e$ 

```

3.2. 文本语义学习层

BiLSTM 层。循环神经网络(Recurrent neural network, RNN)是传统神经网络的变体,神经元的设计使得网络具有“记忆性”,所以 RNN 是深度学习的一种常用框架。对于短文本的处理上 RNN 具有良好的性质,但对于长文本,RNN 网络的“记忆”功能越来越差。为了解决这个问题,Hochreiter & Schmidhuber 于 1997 年提出了长短时记忆网络(Long Short-Term Memory, LSTM) [15]。本文利用的 BiLSTM 模型是其变种,为双向长短时记忆网络,因为情感词不仅和之前出现的文本相关,还会和后面的文本相关。

LSTM 单元结构的设计体现在 3 个“门”结构上,这三个“门”结构使得 LSTM 可以实现对历史信息的重要部分的记忆,这体现在网络中的单元(cell)状态的不断更新。首先记 t (当前)时刻的候选单元值为 \check{c}_t , 输入值的权重为 w_{xc} 以及 $t-1$ 时刻 cell 输出的权重为 w_{hc} , 则 t 时刻的候选单元值为

$$\check{c}_t = \tanh(w_{xc}x_t + w_{hc}h_{t-1} + b_c) \quad (2)$$

激活函数 \tanh 创建的候选单元值 \check{c}_t 将会加到 cell 状态中。

输入门的作用是 t 时刻的输入对 cell 状态的影响,这体现在“遗忘门”机制。并且 3 种“门”机制的运算都受到来自当前输入、前一时刻的 cell 输出以及前一时刻 cell 记忆值的共同影响,即输入门 i_t 表示为

$$i_t = \sigma([w_{xi}, w_{hi}, w_{ci}, 1][x_t, h_{t-1}, c_{t-1}, b_i]^T) \quad (3)$$

遗忘门由 Sigmoid 层实现,控制当前输入值对 cell 状态的影响,可以表示为

$$f_t = \sigma([w_{xf}, w_{hf}, w_{cf}, 1][x_t, h_{t-1}, c_{t-1}, b_f]^T) \quad (4)$$

所以当前时刻 t 的 cell 状态为

$$c_t = f_t \otimes c_{t-1} + i_t \otimes \check{c}_t \quad (5)$$

公式(5)中运算符 \otimes 表示点积运算,从最终的 cell 状态可以得出, LSTM 单元状态的更新由遗忘门、输入门、前一时刻的单元状态和当前候选单元共同决定。

对于输出门, o_t 的表示形式为

$$o_t = \sigma\left([w_{x_o}, w_{h_o}, w_{c_o}, 1][x_t, h_{t-1}, c_{t-1}, b_o]^T\right) \quad (6)$$

最后的输出可以表示为

$$h_t = o_t \otimes \tanh(c_t) \quad (7)$$

本文采用的 BiLSTM 模型的输出表示为

$$\bar{h}_t = \overline{LSTM}(h_{t-1}, x_t, c_{t-1})$$

$$\underline{h}_t = \overline{LSTM}(h_{t+1}, x_t, c_{t+1})$$

$$h_t = [\bar{h}_t, \underline{h}_t]$$

取其平均值作为输出，即输出为 $\text{mean}(h_t)$ 。对于长短文本传入 BiLSTM 层，“门”结构将实现信息的更新与遗弃，这体现在 sigmoid 函数，其输出为 0~1 之间，1 指完全保留，0 指完全遗弃。

词注意力机制层。电影评论为长短文本，评论中的情感词对于评论情感倾向至关重要，加强这些情感词的权重将可以更好的对文本进行情感挖掘。采用词注意力机制，用来学习评论语句中情感词的权重分布，如表 1 所示。

Table 1. Keyword distribution
表 1. 关键词分布

评论文本	关键词
剧本做作，演技尴尬	做作，尴尬
尴尬的对白，尴尬的剧情，尴尬的逻辑，姜文想把电影拍的有趣，可是忽视了角色内在的感情，叙事的内在逻辑，花了这么长时间，这么多钱，可惜了	尴尬，忽视，可惜
非常温暖的一个片，看完心情很好	温暖，很好
单纯美好的小幸福，看的怦然心动	美好，幸福，怦然心动

词注意力机制最早是由 google mind 团队在 RNN 模型上运用，来进行图像分类。Bahdanau 等第一次将词注意力机制运用到 NLP 领域，被应用到机器翻译领域的 sequence to sequence 模型[16]。而后被广泛应用于 CNN 模型，用于图像处理。Wang 等在 LSTM 模型中应用注意力机制来获取重要的评论信息，原因在于注意力机制可以自动学习关键词的权重分布[17]。

本模型中，注意力机制的加入用来自动学习权重分布，自动学习的结果为

$$T = \tanh(Wx + b) \quad (8)$$

其中 W , b 分别为权重和偏置项。将输出处理后，即得注意力向量矩阵。且权重分布为

$$\alpha_i = \frac{\exp(T)}{\sum_{i=1}^S \exp(T)} \quad (9)$$

其中 S 为文本的长度。假设 BiLSTM 层的输出为 $S = [x_1; x_2; \dots; x_L]$ ，则经过 Attention 层后的输出为

$$\tau = \sum_{i=1}^L \alpha_i x_i \quad (10)$$

注意力机制下，权重 α_i 越大，其对应的词越重要，这对文本的关键词提取具有很好的效果。

3.3. CNN 层

卷积神经网络(convolutional neural network, CNN)的出现最早应用于图像处理领域, Kim 等人在 2014 年对传统 CNN 网络进行了一些调整, 使得 CNN 第一次用于文本处理[18]。Santos 等人提出的 CharSCNN 模型, 结合两个卷积核实现语义特征的识别, 在情感分析问题上表现很好[19]。李昊璇等利用词向量结合 CNN 的方法, 对书籍评论进行情感分析, 取得了很好的效果[20]。李慧等将 CNN 用于细粒度情感分析, 相比较传统分类模型具有更好的结果[21]。下面介绍 CNN 模型的主要结构。

卷积层。记输入的字长度数为 l , 窗口大小为 n , 输入值到卷积单元的参数记为 w , 那么卷积操作后的输出值 c_i , 即卷积后得到的特征为

$$c_i = f(w \cdot x'_{ii+n-1} + b) \quad (15)$$

其中, $1 \leq i \leq l - n + 1$ 表示步长, f 为非线性激活函数, b 为偏置项。

池化层。通过一定的策略进行最优特征选择, 常用的方法是平均值法和最大值法, 本文选用最大值方法将卷积后的输出特征 c_i 进行池化操作来选取最优特征。

$$p_i = \max(C) \quad (16)$$

其中, $C = (c_1, c_2, \dots, c_{l-n+1})$ 。

全连接层。池化后的结果 P 通过全连接层后传入输出层, 利用 softmax 函数进行分类, 返回的是情感分类的预测概率, 输出的值在 0~1 之间, 0 表示完全消极, 1 表示完全积极, 公式为

$$p(y|P, w_s, b_s) = \text{softmax}(P \cdot w_s + b_s) \quad (17)$$

其中 w_s , b_s 分别为输出层的权重和偏置项。

4. 实验

4.1. 实验环境

本实验环境下的模型是基于深度学习库 keras, 操作系统为 windows7 旗舰版 64 位, CPU 为 i3, 显卡是 Inter(R) HD Graphics, 开发工具是 pycharm。由于本文数据量不大, 所以不需要太高的配置。

4.2. 实验数据

本文通过爬虫技术爬取了豆瓣网 60775 条评论, 包括 30445 条消极评论, 30330 条积极评论。本文将数据按照 8:2 的比例构造训练集和测试集, 即训练集个数为 48620, 测试集个数为 12155。积极评论和消极评论例子对比见表 2。

Table 2. Comparison of negative text and positive text
表 2. 消极文本和积极文本对比

消极文本	积极文本
中国不适合拍这种片	比想象中好, 导演擅长处理这种题材, 换别的导演或许会非常矫情
怀疑我跟评论的其他人看的是不是同一部电影	被特出的胡歌惊艳到, 这样的周迅真美
看得我十分尴尬	生活, 很暖

从表 2 中可以看出, 消极文本中充斥着消极情感词, 例如消极文本中的“不合适”, “怀疑”, “尴

尬”，积极文本中的“好，擅长”，“惊艳，真美”，“很暖”，情感词直接表示评论文本的情感极性。对文本进行数据预处理，过程如下：

- (1) 评论文本和标签的获取；
 - (2) 去除评论文本标点符号及分词处理；
 - (3) 将处理后的文本构建字典，并将字典中的词进行 token 转化，提取出现频数大于 5 的词来创建词典；
 - (4) 词向量训练，将句子中存在而字典中不存在的词用随机值代替词向量；
 - (5) 将数据打乱，用于模型训练。
- 本文模型的超参数如表 3 所示。

Table 3. Hyperparameter settings

表 3. 超参数设置

参数	数量	释义
Batch_size	256	批量大小，每次输入的样本数量
Sentence_length	82	评论文本的平均长度
Drop_prob	0.5	Dropout 比例
Vocab_size	18082	字典长度
Embedding_dim	300	词向量维度
Num_filters	100	Filter 个数
Hdden_dim	50	全联接层单元数
Num_epochs	10	迭代次数

4.3. 性能评价指标

本文模型的评价指标采用精准率、召回率以及 F1 值。先说明一些指标。

TP 表示正类预测为正类，TN 表示负类预测为负类，FP 表示负类预测为正类，FN 为正类预测为负类。对于模型好坏的评价方式一般采用精准率(Precision)、召回率(Recall)和 F1 值。

精准率：

$$P = \frac{TP}{TP + FP}$$

召回率：

$$R = \frac{TP}{TP + FN}$$

F1 值：

$$F1 = \frac{2TP}{2TP + FP + FN}$$

4.4. 实验对比及分析

本文单模型 CNN 的窗口大小设置为(3, 4, 5)，窗口维度大小为 100，激活函数为 ReLu，池化层采用 Max。单模型 LSTM 的优化函数为 Adam，激活函数为 ReLu。

对比实验。将以下四种模型分别在 word2vec 和随机向量填充两种词向量方法下进行实验。本文实验

不仅讨论了词向量对模型的影响，而且也比较了模型之间的好坏。本文的重要贡献在于将单模型进行集成，得到新的模型，并加入了注意力机制来优化模型。实验如下，实验结果如表 4 所示。

(1) 基于随机向量的 LSTM 模型。本模型仅利用随机化的向量来作为样本输入，用于 LSTM 模型训练。

(2) 基于词向量的 LSTM 模型。本模型利用 word2vec 对文本训练词向量作为输入，用于 LSTM 模型训练。

(3) 基于随机向量的 CNN 模型。本模型仅利用随机化向量来作为样本输入，用于 CNN 模型训练。

(4) 基于词向量训练的 CNN 模型。本模型利用 word2vec 对文本训练词向量并作为输入，用于 CNN 模型训练。

(5) 基于随机向量的 BiLSTM_CNN 模型。该模型是对两种单模型的 Stacking 融合，将样本的随机向量作为模型的输入。

(6) 基于词向量的 BiLSTM_CNN 模型。该模型是将样本进行 word2vec 训练并作为输入。结合 BiLSTM_CNN 模型进行训练。

(7) 本文模型。基于随机向量作为输入，并加入词注意力机制的 BiLSTM_CNN (rand & Word Attention-based BiLSTM & CNN Stacking Model, WABCSM + rand)。

(8) 本文模型。基于词向量训练作为输入，并加入词注意力机制的 BiLSTM_CNN (word2vec & Word Attention-based BiLSTM & CNN Stacking Model, WABCSM + w2v)。

Table 4. Experimental results

表 4. 实验结果

Model	Recall (%)	Precision (%)	F1 (%)
LSTM + rand	79.9	80.3	80.1
LSTM + Word2vec	81.9	81.1	81.5
CNN + rand	80.2	80.4	80.6
CNN + Word2vec	81.4	81.4	81.5
BiLSTM_CNN + rand	81.8	81.3	81.6
BiLSTM_CNN + Word2vec	82.8	82.4	82.6
WABCSM + rand	82.1	81.5	81.8
WABCSM + Word2vec	83.6	83.4	83.5

实验结果分析：

基于词向量。四种模型下的结果显示，一方面 word2vec 训练下的词向量表示结果要优于随机变量填充的训练结果。这说明 word2vec 训练下的词向量很好的表示文本之间的语义关系，对情感分类具有积极的效果。而随机填充下的词向量达不到这个效果。另一方面，即使在随机填充下，模型训练的效果依然不差，这说明深度学习在特征学习上具有很好的效果。

基于模型。本文模型的实验结果要优于传统的单模型。尤其是在注意力机制下的模型，取得了很好的效果。对于单模型，CNN 模型要比 LSTM 模型取得更好的效果，这说明在提取文本的特征上，尤其是提取局部最优特征上(关键词)，CNN 模型要优于 LSTM 模型。比较本文的模型结果，融合模型 BiLSTM_CNN 的效果比传统 LSTM 以及 CNN 效果都要好，随机向量填充下，精准率分别提升了 1.0%和

0.9%；word2vec 训练的词向量下，精准率分别提升了 1.3%和 1.0%。在此基础上加入了词注意力机制，结果表明，词注意力机制下的模型 WABCSM 表现出更好的效果，随机向量填充下，精准率比 BiLSTM_CNN 提升了 0.2%，word2vec 训练的词向量下，精准率比 BiLSTM_CNN 提高了 1.0%。WABCSM 模型对测试集样本预测的结果混淆矩阵如表 5。两种词向量下的模型训练结果对比发现，word2vec 词向量下的训练对正负样本的预测比例与训练样本的比例相近，有 984 个积极样本被预测为消极样本，有 1034 个消极样本被预测为积极样本。而随机向量填充方法下的训练结果预测正确的正负样本都相对较少，并且有 1074 个正样本被预测为负样本，有 1174 个负样本被预测为正样本。

本文模型的实验结果的混淆举证如表 5 所示。

Table 5. Confusion matrix

表 5. 混淆矩阵

	WABCSM + rand		WABCSM_Word2vec	
	pos	neg	pos	neg
pos	4928	1074	5018	984
neg	1174	4979	1034	5119

本文模型的预测结果与真实结果对比见表 6。

Table 6. Forecast results (partial)

表 6. 预测结果(部分)

测试集文本	真实值	预测值
只能说我完全不纯情了	0	0
埃，又是这种，我看他们国家的爱情电影是没希望了	0	0
我看哭了说实话	1	0
小森湘太好看了吧！男二演的真好，水管其次	1	1
被一股神奇的力量恶心到了	0	0
合格的漫威作品。片头要看哭，从头到尾只记住斯坦李的一句话：Trust me, believers	1	1

表 6 中给出了本文模型对测试集的部分预测结果，其中 0 表示消极评论，1 表示积极评论，由于本文探索的是二分类问题，而评论会有部分属于中性评论，这使得模型对其情感极性的影响有所影响。比如“只能说我完全不纯情了”和“我看哭了说实话”这样的文本，文本太短，中文的一词多意也加重了机器学习的难度，缺乏直观的情感极性，使得机器只能根据文本中的单一情感词进行判断，“纯情”是积极情感词，机器就预测为积极文本，而“看哭”是消极情感词，机器就预测为消极文本。这也说明了一直以来对中文短文本进行情感挖掘是一个很难的课题。

图 2 给出了 BiLSTM_CNN 模型和 WABCSM 模型在两种词向量方法下的损失函数变化图。从四种模型在测试集上的损失变化图中可以看出，四种模型均收敛的比较快并在第 8 次迭代后达到收敛。对于模型 BiLSTM_CNN，两种词向量方法下的结果显示，在第二次迭代和第三次迭代时，随机向量填充下的损失下降要比 word2vec 方法下的更快，而在第四次迭代开始，word2vec 方法下的模型快速收敛并在第 7 次迭代开始达到收敛，而随机向量填充下的模型在第八次迭代开始收敛，并且结果要差一些。对于 WABCSM 模型，两种词向量下的结果都要比 BiLSTM_CNN 要好，无论在收敛速速上，还是在损失函数

上。尤其是 WABCSM_w2v 模型，在第 5 次迭代就开始趋向于收敛，WABCSM_rand 模型也在第六次开始收敛。除此之外，稳定后 WABCSM_w2v 模型的验证损失达到最低，优于其他模型。这显示出本文模型 WABCSM 具有很好的泛化能力，而且 word2vec 词向量训练方法有助于模型训练。

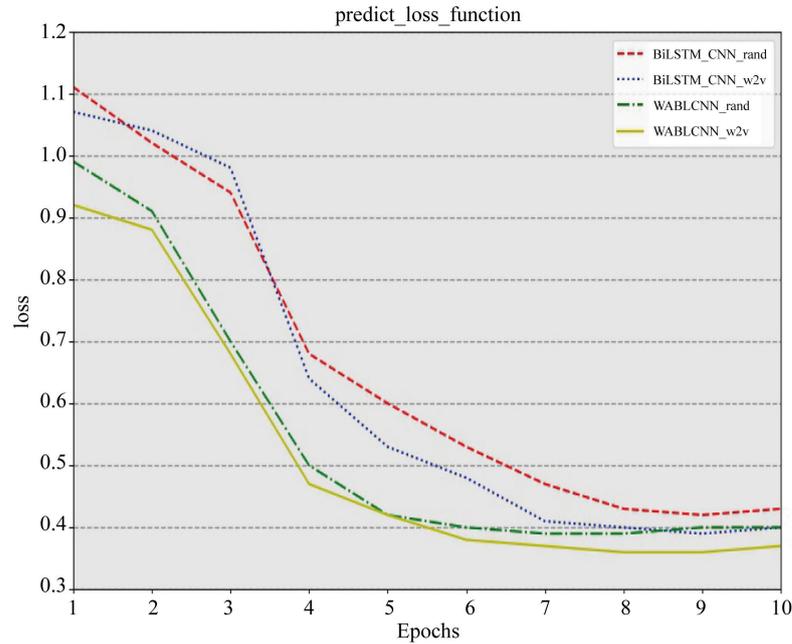


Figure 2. Comparison of changes in loss function
图 2. 损失函数变化对比图

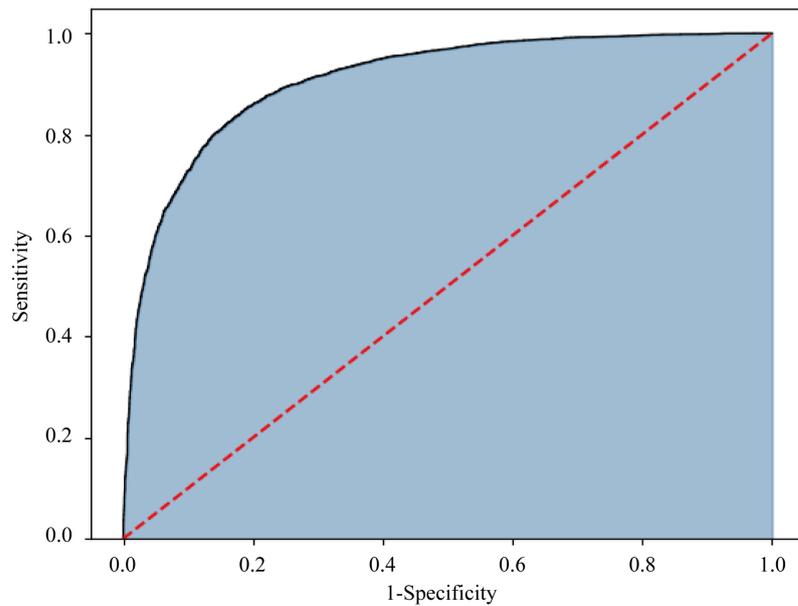


Figure 3. ROC curve of WABCSM model
图 3. WABCSM 模型 ROC 曲线图

本文模型的 ROC 曲线见图 3，ROC 曲线的横轴表示负正类率(False Positive Rate, FPR)，纵轴表示真

正类率(True Positive Rate, TPR)。TPR 值越大越好, 即 ROC 曲线越靠近左上角说明模型越好。本文模型首先利用 BiLSTM 模型对数据集进行长文本词语的更新和遗弃, 然后加入词注意力机制, 增大文本中情感词的权重, 使的模型更容易提取代表文本情感极性的情感词, 最后利用 CNN 模型进行局部最优特征提取来完成情感分类的任务。本文词向量的训练采用了两种方法, 分别为随机向量填充方法和 word2vec 训练方法。利用 LSTM 和 CNN 这两种传统单一模型以及本文设计的两种模型做对比。结果发现本文模型 BiLSTM_CNN 模型在两种词向量下的实验结果无论在收敛速度还是在准确度上都要优于 LSTM 模型和 CNN 模型。为了提升本文模型的效果, 在 BiLSTM_CNN 模型基础上加入了词注意力机制, 形成了本文第二种模型 WABCSM。WABCSM 模型对评论文本关键词的提取上表现的更加优异。与 BiLSTM_CNN 模型的实验结果比较发现, WABCSM 模型的收敛速度更快, 精准率更高, 性能更加鲁棒。

5. 结论

本文提出了一种基于词注意力机制的 BiLSTM 和 CNN 的集成模型来进行中文情感分析, 本文模型首先利用 BiLSTM 模型对数据集进行长文本词语的更新和遗弃, 然后加入词注意力机制, 增大文本中情感词的权重, 使的模型更容易提取代表文本情感极性的情感词, 最后利用 CNN 模型进行局部最优特征提取来完成情感分类的任务。本文词向量的训练采用了两种方法, 分别为随机向量填充方法和 word2vec 训练方法。利用 LSTM 和 CNN 这两种传统单一模型以及本文设计的两种模型做对比。结果发现本文模型 BiLSTM_CNN 模型在两种词向量下的实验结果无论在收敛速度还是在准确度上都要优于 LSTM 模型和 CNN 模型。为了提升本文模型的效果, 在 BiLSTM_CNN 模型基础上加入了词注意力机制, 形成了本文第二种模型 WABCSM。WABCSM 模型对评论文本关键词的提取上表现的更加优异。与 BiLSTM_CNN 模型的实验结果比较发现, WABCSM 模型的收敛速度更快, 精准率更高, 性能更加鲁棒。

本文模型的设计源于集成思想, 综合利用模型的优点, 设计了表现力更好的集成模型, 无论对于研究还是应用都可以得到好的效果, 本文数据来源于实际生活, 使得本文的研究更有意义。本文的不足在于数据量不大, 利用深度学习模型训练出现了“过拟合”现象。本文尝试了比较数据量对模型训练结果的影响, 发现随着数据量的增大, 模型的性能越来越好, 但由于电脑配置的原因, 未找到最佳点。

参考文献

- [1] 周纯洁, 黎巛, 徐翼龙, 等. 文本情感分析研究[J]. 计算机科学, 2018, 10(45): 296-299.
- [2] 肖江, 丁星, 何荣杰. 基于领域情感词典的中文微博情感分析[J]. 电子设计工程, 2015, 6(12): 18-21.
- [3] Paltoglou, G. and Thelwall, M. (2012) Twitter, My Space, Digg: Unsupervised Sentiment Analysis in Social Media. *ACM Transactions on Intelligent Systems & Technology*, **3**, 1-19. <https://doi.org/10.1145/2337542.2337551>
- [4] Jo, Y. and Oh, A.H. (2011) Aspect and Sentiment Unification Model for Online Review Analysis. In: *ACM International Conference on Web Search and Data Mining*, ACM, New York, 815-824. <https://doi.org/10.1145/1935826.1935932>
- [5] Pang, B., Lee, L. and Vaithyanathan, S. (2002) Thumbs up? Sentiment Classification Using Machine Learning Techniques. *Proceedings of Annual Conference of the Association for Computational Linguistics*, July 2002, 79-86. <https://doi.org/10.3115/1118693.1118704>
- [6] Liu, S., Li, F., et al. (2013) Adaptive Co-Training SVM for Sentiment Classification on Tweets. In: *Proceeding of the 22nd ACM International Conference on Information & Knowledge Management*, ACM, New York, 2079-2088. <https://doi.org/10.1145/2505515.2505569>
- [7] Berger, A.L., Dellapetra, V.J., Pietra, S.A.D., et al. (1996) A Maximum Entropy Approach to Natural Language Processing. *Computational Linguistics*, **22**, 39-71.
- [8] Hinton, G.E. (1986) Learning Distributed Representations of Concepts. *Proceedings of the Eighth Annual Conference of the Cognitive Science Society*, **1**, 12.
- [9] Le, Q. and Mikolov, T. (2014) Distributed Representations of Sentences and Documents. *Proceedings of the 31st International Conference on Machine Learning*, **14**, 1188-1196.
- [10] Collobert, R., Weston, J., Bottou, L., et al. (2011) Natural Language Processing (Almost) from Scratch. *Journal of Machine Learning Research*, **12**, 2493-2537.
- [11] 王煜涵, 张春云, 赵宝林, 等. 卷积神经网络下的 Twitter 文本情感分析[J]. 数据采集与处理, 2018, 33(5): 921-927.
- [12] You, Q.Z., Chen, Y.X., Yuan, J.B. and Luo, J.B. (2018) Twitter Sentiment Analysis via Bi-Sense Emoji Embedding and Attention-Based LSTM. *Computer and Language*, **8**, 117-125.
- [13] 关鹏飞, 李宝安, 吕学强, 等. 注意力增强的双向 LSTM 情感分析[J]. 中文信息学报, 2019, 33(2): 105-111.

- [14] 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): 123-130.
- [15] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, **9**, 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [16] Bahdanau, D., Cho, K. and Bengio, Y. (2014) Neural Machine Translation by Jointly Learning to Align and Translate. *Computer Science*.
- [17] Wang, Y., Huang, M., Zhu, X. and Zhao, L. (2016) Attention-Based LSTM for Aspect-Level Sentiment Classification. *Proceedings of 2016 Conference on Empirical Methods in Nature Language Processing*, Austin, TX, 1-5 November 2016, 606-615. <https://doi.org/10.18653/v1/D16-1058>
- [18] Kim, Y. (2014) Convolutional Neural Networks for Sentence Classification. <https://doi.org/10.3115/v1/D14-1181>
- [19] Dos Santos, C.N. and Gatti, M. (2014) Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts. *Proceedings of the 25th International Conference on Computational Linguistics (COLING)*, Dublin, Ireland, 23-29 August 2014.
- [20] 李昊璇, 张华洁. 基于词向量和 CNN 的书籍评论情感分析[J]. 测试技术学报, 2019, 33(2): 165-171.
- [21] 李慧, 柴亚青. 基于卷积神经网络的细粒度情感分析方法[J]. 数据分析与知识实现, 2019, 3(1): 95-103.