文章编号:1007－6735(2021)06－0597－09　DOI: 10.13255/j.cnki.jusst.20210102001

基于双通道融合和 BiLSTM-attention 的评论文本情感分类算法

颜礼蓉，　朱小栋，　陈　曦
（上海理工大学 管理学院，上海 200093）

摘要：对在线商业评论文本的情感进行挖掘，融合评论文本不同特征为分类器提供更多的信息量，提出了一种新的在线电商情感分类算法。首先，针对传统词嵌入模型无法很好地融合词语情感信息特征的不足，考虑了词嵌入特征和词性特征的多特征融合方法；其次，在两种特征融合方法的基础上采用了双通道和单通道的对比来比较分类的准确性，提出了并行的 CNN 和 BiLSTM-Attention 双通道神经网络模型；最后，使用真实的京东电商评论数据集对所提模型进行了评估，并且在实验中与不同分类算法进行对比。实验结果表明，新的混合方法具有更好的分类准确率、召回率和 F1 指标。

关键词：情感分析；电商评论；多特征融合；神经网络
中图分类号：TP 391　　**文献标志码：**A

Emotional classification algorithm of comment text based on two-channel fusion and BiLSTM-attention

YAN Lirong,　ZHU Xiaodong,　CHEN Xi
(*Business School, University of Shanghai for Science and Technology, Shanghai 200093, China*)

Abstract: The emotion of online business review texts was mined, and different features of review texts were fused to provide more information for the classifier. A new online e-commerce emotion classification algorithm was proposed. Firstly, aiming at the deficiency that the the emotional information features of words cannot be fused well in the traditional word embedding model, the multi-feature fusion method of word embedding features and part-of-speech features was considered in this study. Secondly, on the basis of two feature fusion methods, the classification accuracy based on both two-channel and single-channel methods was compared. A parallel CNN and BiLSTM-attention two-channel neural network model was proposed. Finally, the real JD.COM e-commerce review data set was used to evaluate the proposed model, and compared with different classification algorithms in experiments. Experimental results show that the new hybrid method has better classification accuracy,

收稿日期: 2021−01−02
基金项目: 国家自然科学基金资助项目（71871144）；上海高校智库内涵建设计划（战略研究）项目；上海市教委上海市级新农科研究与改革实践项目
第一作者: 颜礼蓉（1997−），女，硕士研究生．研究方向：数据挖掘、电子商务．E-mail：ylryan@163.com
通信作者: 朱小栋（1981−），男，副教授．研究方向：大数据管理、数据挖掘、云计算、信息安全．E-mail：zhuxd@usst.edu.cn

recall rate, and F1 scores.

Keywords: *emotional analysis*; *e-commerce review*; *multi-feature fusion*; *neural network*

电子商务的在线评价作为由消费者主动发起的评论商品质量和服务等的言论，对潜在消费者作出购买决定起着一定的作用，也直接影响了电商平台的用户使用黏性。挖掘这些评论的褒贬态度，从而识别人们对某种商品的购买倾向，这个过程被称为情感分析(sentiment analysis，SA)。情感分析是一种使用自然语言处理(natural language processing, NLP)以及文本挖掘(text mining)技术的分析方法，通过分析大量信息来理解人们对一件事情的看法。在自然语言处理中，文本情感分析是一个重要的应用领域。

1 相关研究

为了更充分地挖掘网民的观点和立场，使用机器学习的方法进行情感分析成为了自然语言处理的一个新兴的研究方向。在早期，情感分析是通过使用传统的机器学习方法构造分类器来解决的。Pang 等[1]首次提出利用机器学习算法包括朴素贝叶斯(NB)，最大熵分类(ME)和支持向量机(SVM)来解决情感分类问题，他们利用 n-grams 模型和词性提取电影评论的特征。Liu 等[2]利用基于情感特征的细颗粒度情感分析方法，主要贡献是通过句法分析提取相应特征，并与 TF-IDF 基准模型进行对比实验，其提出的模型在积极/消极评价中的精确率(precision)、召回率(recall)和 F1-Score 上都有所提升。Lin 等[3]通过词嵌入 word2vec 方法提取中国酒店评论的特征，并放入分类器朴素贝叶斯(NB)、支持向量机(SVM)和卷积神经网络(CNN)中进行对比，其中支持向量机(SVM)在分类中表现最好。利用词嵌入的方式可以有效地从评论文本中提取到词语的信息(例如某个词在文本中出现与否或是出现的频数)和词语的层次信息(主要是指上下文的信息)，但是无法提取出词语中所表示情感的信息。因而将情感词典与词嵌入的方法融合可更全面地表达出评论中的信息。

近几年，由于深度学习方法在自然语言处理方面取得了较好的研究成果，许多学者对于深度学习模型应用于情感分析十分热衷，将不断改进的经典深度学习模型应用于这个领域，甚至为了解决这个领域中的问题而提出新的深度学习模型。Kim[4]较早提出了 textCNN 模型，利用卷积神经网络(CNN)算法处理句子分类的问题，在 7 项任务中，有 4 项得到了比以往研究更好的结果(包括情感分类任务)。Zhao 等[5]利用 Glove 提取特征，放入非常深的 CNN 模型中做 Twitter 的情感分类实验，实验结果表明其模型准确性和 F1 值都比基准模型要高。除了 CNN 以外，深度学习中的长短期记忆网络(LSTM)算法被认为能更好地学习文本的上下文信息，也被学者们应用于情感分类这个问题上。Hassan 等[6]利用 LSTM 替换 CNN 中的池化层，进行二分类以及五分类的实验，结果比之前提出的模型准确率更高。

除此之外，现有许多学者将文本中的多种特征融合起来，达到更好的分类效果。Lazib 等[7]通过句法分析得到文本的特征，将 CNN 和 BiLSTM 模型结合到同一个模型进行分类。相比其他同类型实验来说，提出特征的方法更节省时间并且有更好的效果。Abdi 等[8]融合了词嵌入特征、情感信息(词典)特征和语言知识特征，并根据相关的策略克服了词嵌入的缺点，其模型与其他经典方法相比具有优势。

以往的研究证明了对文本不同特征的融合可以为分类器提供更多的信息量，然而，要如何通过构造词嵌入方式和分类器的结构，并融合不同的特征，使提出的模型有更好的分类效果，这是本文研究的一个重点。也就是说，在判断句子的情感极性时，不仅要考虑结合更多的文本词信息，还要考虑构造分类器结构，更好地提取出文本中的不同特征。

在此基础上，首先提出了一种构造词嵌入拼接机制，这种机制不仅能够将文本中的词向量表示出来，还能将词性标注信息作为句子表征的一个特征嵌入其中。此外，提出了一种信息并行注意机制，使用 BiLSTM 计算的并行结构时序信息特征，通过门控机制更新存储单元。CNN 模型能够更好地提取到某一类的词性(比如形容词、副词、名词等)，这些词性对于情感的表达具有更显著的作用。在此基础上，加入 attention 机制进一

步强化分类结果，可以理解为在一句话中，有一些词为这句话的重点词，而在这里 attention 机制就是为了帮助提取这一类重点，更好地将特征中的重点“表现”出来。实验结果表明，与其他基准模型相比，该方法具有更好的分类准确率和 F1 指标。

2　模　型

由于词嵌入模型无法很好地包含词语的情感信息，本文提出了 PWCNN 模型和 PW2CNN 模型来应用词性特征相结合的方式使特征包含的信息更加丰富。而单一卷积神经网络模型无法很好地捕捉到一句话的时序信息，本文使用循环神经网络模型中的双向长短期记忆模型进行实验，加入注意力机制，提出并行分类算法 PW2CNN&BiLSTMatt 模型。

2.1　Word2vec + CNN 模型

TextCNN 经典模型是由 Kim[4] 提出的，每一个词是一条一维的向量，一句话组成一个矩阵。通过不同的卷积核进行卷积操作，再通过池化操作进行降维，最后利用 sigmoid 函数进行二分类。

早先，词袋模型（bag of words, BOW）表示每句话的词频向量，该方法是把文档中所有的词做成一个词袋并构建对应的词典（dictionary），使生成的向量与词典相互对应且能够把词的频率表现出来。在较长的句子集中，这种方法会使生成的矩阵非常稀疏，而且丧失了语序关系的信息。词嵌入模型 word2vec 能够更好地考虑到词语位置的关系，并且解决用 One-Hot 编码对词进行向量化时过于稀疏的问题。

词嵌入模型的提出是为了更好地在一个大型数据集上快速、准确地学习出词表示，利用将高维词向量嵌入到低维空间的想法，使相邻意思的词具有更近的空间距离。本文模型中使用的是连续词袋模型（continuous bag-of-word model, CBOW），生成的每条文本矩阵为 256×128 维。每个词设置使用 128 维表示，而词与词之间拼接起来便形成了特征向量矩阵，假如一句话不超过 256 个词则用 0 进行填充。

Word2vec + CNN 模型对 TextCNN 经典模型进行了修改，模型的 CNN 结构对每句话的特征矩阵上通过多个不同的 4×4 卷积核进行卷积操作。输入文本矩阵 $\boldsymbol{X}\in\mathbb{R}^{N\times N}$ 和滤波器 $\boldsymbol{W}\in\mathbb{R}^{U\times V}$，二维卷积的公式如下：

$$c_{ij}=\sum_{u=0}^{U}\sum_{v=0}^{V}w_{uv}x_{i-u+1,j-v+1} \tag{1}$$

式（1）表示用一个窗口大小为 $U\times V$ 的滤波器 $\boldsymbol{W}$，作用在 $x_{i,j}$ 到 $x_{i-u+1,j-v+1}$ 上，将文本矩阵 $\boldsymbol{X}$ 中元素 x_{ij} 和卷积核矩阵 $\boldsymbol{W}$ 中元素 w_{uv} 相乘，得到特征 c_{ij} 的过程，最后得到多个特征矩阵。

图 1 中第一层为输入层，最后一层为输出层，而夹在中间的几层代表的是卷积神经网络模型的结构。这个卷积神经网络模型有两个卷积层，一个池化层。第一个卷积层有 32 个卷积核，都是 4×4 的矩阵，激活函数为 ReLu 函数。第二个卷积的卷积核有 16 个，卷积核的大小为 4×4，同样使用 ReLu 函数。采用 ReLu 函数的原因是因为其计算量较其他激活函数更小，并能起到较好的防止过拟合作用。这里的第三层是最大池化层，设置的池化核大小为 2×2。池化层和全连接层之间加入了 dropout，其取值为 0.25。该全连接层与输出层之间同样加了 dropout，值为 0.5，最后使用的是 sigmoid 函数进行二分类。

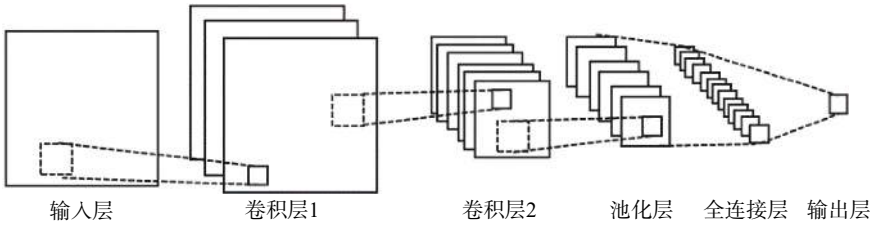

图 1　Word2vec + CNN 模型中卷积神经网络模型

Fig.1　Convolution neural network in Word2vec+CNN model

2.2　PWCNN 模型

在这个实验中，Word2vec + CNN 基准模型是单以 word2vec 作为输入的单通道卷积神经网络模型，并且与 PWCNN 模型双通道模型中的卷积神经网络模型的结构一样，不同的是特征的输入。两个特征分别是词嵌入 word2vec 模型矩阵（256×

128×1 维)和词性标注(part-of-speech, POS)特征输入矩阵(220×56×1 维)。

词性标注(part-of-speech，POS)是在分词的基础上，判断每个词在该句话中的词性标注。使用词性特征可以有效地消除歧义，例如“工作”这个词既可以做名词也可以做动词，在中文里面这样的词还有很多。并且词嵌入模型无法很好地表达一句话的情感信息，因而这里利用词性特征作为补充，能够较好地提供语句中的情感信息。模型比较关注的重要词性有形容词、副词和动词，因为这些词性的词能够比较好地表达出评论者的主观感受。另外，一句话中标点符号也很重要，起着判断情感极性的作用，因而在进行分词并给出对应分词的词性时，不考虑去除标点符号。模型词性特征提取是先由分词后的每一句话生成一个词性的列表，再根据词性字典，生成对应的特征矩阵(原理和词袋模型一样)。

PWCNN 模型将词嵌入 word2vec 模型训练的词向量与词性向量(POS)进行了“上下”拼接，包括两种拼接方式，如图 2 和式(2)所示。

$$\boldsymbol{X} = [\boldsymbol{X}_W, \boldsymbol{X}_P] \tag{2}$$

式中：$\boldsymbol{X}_W$表示词嵌入 word2vec 模型矩阵；$\boldsymbol{X}_P$表示词性标注特征输入矩阵。

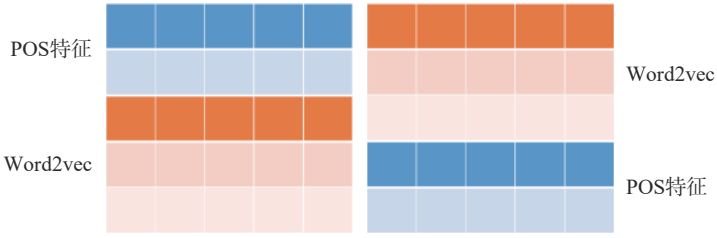

图 2　词嵌入模型训练的词向量与词性向量拼接图

Fig.2　Mosaic diagram of word vector and part-of-speech vector trained by the word embedding model

另外，每句话的 POS 特征和 word2vec 特征是一一对应拼接起来的，这样拼接出来的整个特征矩阵才具有意义。然后再将拼接的特征向量放入和 word2vec + CNN 模型一样的单通道卷积神经网络模型中进行训练，这主要是为了将特征融合的单通道模型和基准模型的实验结果直接进行对比。考虑到 CNN 模型结构与经典的 textCNN 模型结构的不同，选择这种“上下”拼接而不是“左右”拼接的方式，但是“左右”拼接也非常值得尝试。通过实验结果证明，拼接的特征含有更加丰富的信息，因此分类效果更好。与此同时，通过对比实验，发现这两种拼接方式对实验结果没有显著的影响，对于卷积神经网络来说无差别。

将融合特征的词向量放入与上文相同结构的卷积中进行实验，并且其中激活函数和参数的设定都保持一致。通过两层的卷积层把特征向量中的信息提取出来，再通过池化层进一步降维、汇总信息，这里通过随机 dropout 进行全连接处理，为了进一步防止过拟合(减少参数)，在该 flatten 层后面加入 dropout，最后进行分类。实验模型如图 3 所示。

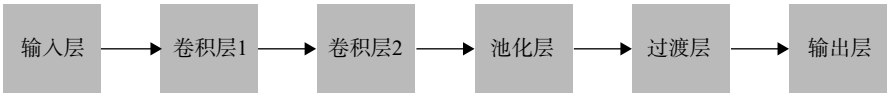

图 3　PWCNN 模型

Fig.3　PWCNN model

该实验与 word2vec + CNN 模型情况作对比，可以清楚地判断 POS 特征在这个分类实验中是否能够起到一定的作用。实验证明加入 POS 特征的输入矩阵的确有更好的分类效果。下一步就是在此单通道的基础上，进一步考虑双通道的 CNN 模型，探索相同的特征分别进入相同结构的 CNN 模型时是否受到影响。

2.3　PW2CNN 模型

双通道模型(并行)与单通道模型的输入不同，单通道模型的输入是通过拼接不同的特征而形成的融合特征，而并行模型中则是通过分别输入不同的特征进行相关实验的。并且不同的特征

可以进入到不同的神经网络模型中，最后可以通过矩阵拼接的方式达到融合的效果。双通道 CNN 模型中一边输入的是 POS 特征的矩阵，另一边输入的是 word2vec 模型的矩阵。两个特征矩阵分别经过 CNN 处理(两个卷积层、一个池化层以及设置了 dropout 层，值为 0.25)，在过渡层中将处理后的两个特征进行“压平”操作(将多维数据变成一维的过程)，“左右”拼接在一起，形成一条矩阵，这样做的目的也是为了同时保留两边的信息。

$$\boldsymbol{C}=[\boldsymbol{C}_1,\boldsymbol{C}_2] \tag{3}$$

式中：$\boldsymbol{C}_1$表示输入为 POS 特征的卷积特征矩阵；$\boldsymbol{C}_2$表示输入为 word2vec 模型的矩阵卷积特征矩阵。

将二维数据变成一维后拼接，将拼接层以全连接的方式连接到最后的输出层，其中设置 dropout 层，值为 0.5。整体的模型的结构如图 4 所示。

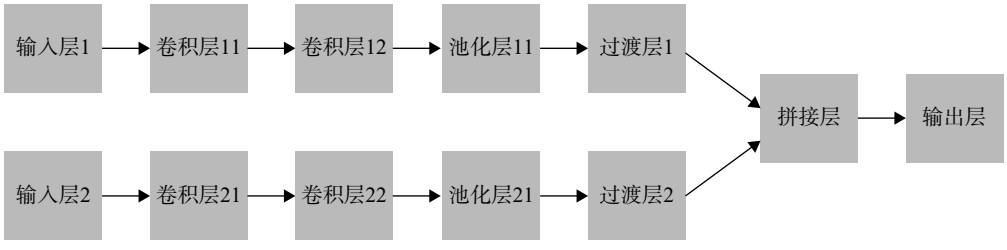

图 4　PW2CNN 模型

Fig.4　PW2CNN model

为了确保模型之间的可对比性，双通道与单通道的 CNN 结构基本保持一致，通过实验结果的对比发现，其验证集上的正确率和 F1 值改进的效果并不明显。但是，当把模型双通道一边的 CNN 换成 BiLSTM 时，结果发生了显著的变化，说明影响整体结果的不仅仅在于特征的选取(信息的选取)，还在于分类器的选取。当一边输入的是 POS 特征时，利用 CNN 能够很好地提取局部的特性。而当另一边输入为 word2vec 模型时，使用能够考虑上下文语境的 BiLSTM 模型，因为词嵌入特征本身就具备词本身的信息。综上所述，这样的分类器有更好的实验结果，也是本文提出该模型的原因。

2.4　PW2CNN&BiLSTMatt 模型

本文提出的 PW2CNN&BiLSTMatt 模型词嵌入部分使用 word2vec 模型，分类器选用 CNN 和 BiLSTM 模型，并且加入注意力机制。

由于 RNN(recurrent neural network)在输入时间序列不断增加的情况下，经过多次传播后会出现梯度消失或梯度爆炸的现象，从而丧失学习长期信息的能力，即存在长期依赖问题。为此，在模型设计中引入长短时记忆网络(long short-term memory, LSTM)，这是一种特殊类型的 RNN，能够学习长期的依赖关系。LSTM 通过门结构来实现向细胞状态中移除或添加信息。

双向长短期记忆网络(bi-directional long short-term memory, BiLSTM)是由两个普通的 LSTM 所组成，一个正向的 LSTM 可以利用句子中过去的信息，一个逆向的 LSTM 可以利用句子中未来的信息。这样一来，模型就可以实现对上下文信息的提取了，因此，双向 LSTM 的预测结果会更加精确。对于文本的情感分类问题来说，该模型非常的适用，因为其包含了句子中的前向和后向的所有信息。

$$i_t=\sigma(W_i[h_{t-1},x_t]+b_i) \tag{4}$$

$$f_t=\sigma\left(W_f[h_{t-1},x_t]+b_f\right) \tag{5}$$

$$\tilde{c}_t=\tanh(W_c[h_{t-1},x_t]+b_c) \tag{6}$$

$$c_t=f_t\odot c_{t-1}+i_t\odot\tilde{c}_t \tag{7}$$

$$o_t=\sigma(W_o[h_{t-1},x_t]+b_o) \tag{8}$$

$$h_t=o_t\odot\tanh C_t \tag{9}$$

式中：σ 表示使用 sigmoid 函数作为激活函数；W_i，W_f，W_c，W_o 表示函数训练出来的权重；b_i，b_f，b_c，b_o 表示函数训练出来的偏置。LSTM 网络引入一个新的内部状态(internal state)$c_t\in\mathbb{R}^D$进行信息传递，同时输出信息给隐藏层的外部状态 $h_t\in\mathbb{R}^D$，其中 $f_t\in[0,1]^D$、$i_t\in[0,1]^D$ 和 $o_t\in[0,1]^D$ 为 3 个门(gate)来控制信息传递的路径；$\odot$为向量元素乘积；c_{t-1}为上一时刻的记忆单元；$\tilde{c}_t\in\mathbb{R}^D$是通过 tanh 函数得到的候选状态。

注意力机制(attention mechanism)的灵感来自于人本身的认知功能，在大量信息中提取接收小部分重要的信息，忽略其他信息，这样的能力叫

作人的注意力。类似地，注意力机制的本质就是关注输入的某些关键部分，给予更高的权重。基于 attention 机制的 LSTM 模型在任一时刻的隐藏状态不仅取决于当前时刻的隐藏层状态和上一时刻的输出，还依赖于上下文的特征，这个上下文特征是通过所有时刻的隐藏状态加权平均得到的。计算公式为：

a. 计算注意力分布。

$$a_i = p(z=i|\boldsymbol{X},\boldsymbol{q}) = \mathrm{softmax}(s(x_i,q)) = \frac{\exp(s(x_i,q))}{\sum_{j=1}^{N}\exp(s(x_i,q))} \tag{10}$$

b.根据注意力分布来计算输入信息的加权平均。

$$att(\boldsymbol{X},\boldsymbol{q}) = \sum_{i=1}^{N}\alpha_i x_i = E_{z\sim p(z|\boldsymbol{X},\boldsymbol{q})}[x] \tag{11}$$

计算注意力分布就是计算在给定查询向量 $\boldsymbol{q}$ 和输入 $\boldsymbol{X}$ 下，选择第 i 个输入向量的概率 a_i。其中：$z\in[1, N]$ 为注意力变量，表示被选择信息的索引位置，即 $z=i$ 表示选择了第 i 个输入向量；a_i 为注意力分布；$s(x_i,\boldsymbol{q})$ 为注意力打分函数。注意力分布 a_i 可以解释为在给定任务相关的查询 $\boldsymbol{q}$ 时，第 i 个输入向量受关注的程度，根据 a_i 来计算输入信息的加权平均，对输入信息进行汇总。

模型分类器分别选用 CNN 和 BiLSTM 作为并行的结构，通过 CNN 模型能够很好地提取到某一类的词性（比如形容词、副词、名词等），这些词性对于情感的表达具有更显著的作用。而 BiLSTM 更适合抓取时序信息特征。加入 attention 机制进一步强化分类结果，可以理解为在一句话中，有一些词为这句话的重点词，而在这里 attention 机制就是为了帮助提取这一类重点，更好地将特征中的重点“表现”出来。整体的模型结构如图 5 所示。

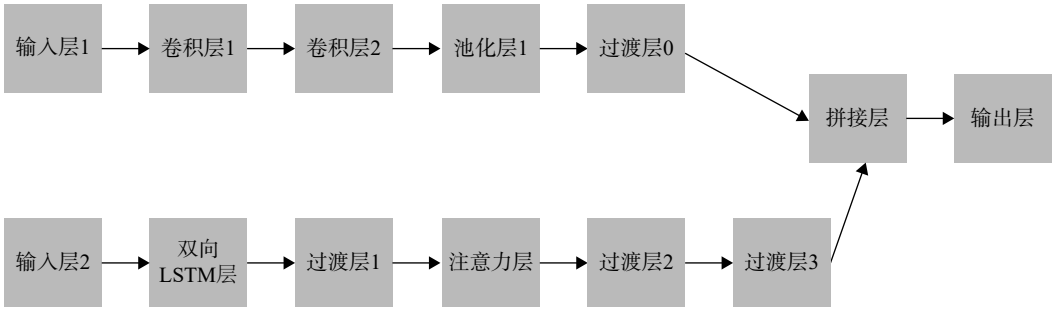

图 5 PW2CNN&BiLSTMatt 模型

Fig.5 PW2CNN&BiLSTMatt model

CNN 模型用到的设定参数，与前文保持一致，这样可以更好地对比这些模型结构性能的优劣。而另一边是加入注意力机制的长短期记忆模型，其中，在 BiLSTM 上使用 attention 机制的示意图如图 6 所示。

与前文所提到的基于 attention 机制的 LSTM 模型类似，双向的 LSTM 中当前状态也应与上下文特征有关，通过 attention 机制将所有时刻的隐藏状态加权平均后输入当前状态，从而影响输出。根据输出结果与真实情况的差异，调整 attention 机制中的权值。实际上，通过本实验可以证明，与基准模型相比，这个并行模型的分类效果最好，它不仅充分利用了不同的特征信息，也发挥了不同神经网络模型的优势。而相比起双通道的 CNN 模型，该模型的分类结果更佳，说明加入 attention 机制的 BiLSTM 模型在其中发挥了重要作用。

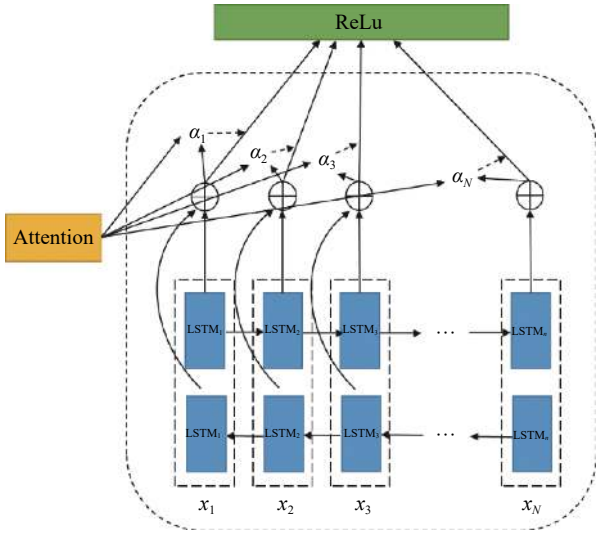

图 6 BiLSTM 上使用 attention 机制的示意图

Fig. 6 Schematic diagram of applying attention mechanism on BiLSTM

3　实验研究

3.1　数据集

实验所用数据集为海信、小米、飞利浦、TCL 和创维 5 个品牌的电视机评论数据，来源于京东电商平台。将网页中的好评和差评分开爬取，从而解决文本数据的标签问题。表 1 给出了本文爬取的文本评论数据的数量信息。

表 1　爬取的文本数据信息

Tab.1　Crawled text data information

品牌	好评	差评	总计	好/坏(比)
海信	4 282	1 218	5 500	3.52
小米	4 002	2 592	6 594	1.54
飞利浦	4 624	1 124	5 750	4.11
TCL	5 007	1 949	6 956	2.60
创维	4 000	2 332	6 332	1.72
汇总	21 915	9 215	31 130	2.38

模型使用的文本数据是所有的好评论(21 790 条)和差评论(9 340 条)，在训练过程中的训练集和验证集的比例设为 7∶3。换句话说，训练集中好评的数量为 15 253 条，差评的数量为 6 538 条。验证集中好评的数量为 6 537 条，差评的数量为 2 802 条。

3.2　实验环境

为了有效验证模型的性能指标，对比实验环境设置如下：操作系统为 Windows 64，内存为 32 GB，处理器为 Intel(R)Xeon(R)CPU E5-2 650 v4 @ 2.20 GHz(2 处理器)，模型使用的是 keras 深度学习框架。

3.3　损失函数(loss function)

损失函数是一个非负实数函数，用来量化模型预测和真实标签之间的差异。本文实验中的损失函数用的是交叉熵损失函数(binary cross entropy)，公式如下：

$$loss = -\sum_{i=1}^{n} y_i \lg \hat{y}_i + (1-y_i)\lg(1-\hat{y}_i) \qquad (12)$$

式中：y_i为真实的离散类别；$\hat{y}_i$为预测类别标签的条件概率分布。

3.4　实验参数

本文的卷积神经网络模型上的超参数设置如表 2 所示。

PW2CNN&BiLSTMatt 模型将长短期记忆模型放入并行结构中，并且加入了 attention 机制，该并行结构中的长短期记忆模型中 attention 层的激活函数为 Softmax 函数，中间输出层的激活函数为 Relu 函数，attention 层的 dropout 值为 0.3。本文的并行结构模型，拼接之后是全连接层以及输出层，全连接层及输出层激活函数分别为 Relu 函数和 Sigmoid 函数，其中全连接层设置的 dropout 值为 0.5。

表 2　卷积神经网络的参数设置

Tab.2　Parameter setting of convolutional neural network

卷积层	卷积核个数	卷积核大小	激活函数
Conv1	32	4×4	Relu函数
Conv2	16	4×4	Relu函数
池化层	**池化方法**	**池化核大小**	**dropout值**
Pool1	最大池化	2×2	0.25

对于不是并行结构的模型，在卷积神经网络后面也同样跟着如上参数设置的全连接层和输出层，主要是为了保持多个模型中的一致性。除了以上提到的参数设置外，在上述 4 个不同的对比实验中，训练过程中所设置的细节如下：损失函数为交叉熵函数，优化器选择 Adam 方法，Batch size 设置为 64。前期进行多个 30 次 epoch 的实验后往往在前 10 次 epoch 已经收敛了，而训练过多 epoch 次数只会使模型趋向于过拟合的状态，因此最终对比实验 epoch 设置为 10。而实验次数越多越具有科学性，为避免偶然性的误差，本文的实验结果都是以 10 次实验的平均数记录并对比的。

3.5　实验结果及分析

通过前文提到的方法，对 4 个模型 10 次实验结果进行比较。实验结果如图 7 和图 8 所示。图中 BASELINE 基准模型为 word2vec + CNN 模型，PWCNN 为特征融合的 CNN 模型，PW2CNN 模型为双通道 CNN 模型，PW2CNN&BiLSTMatt 模型为

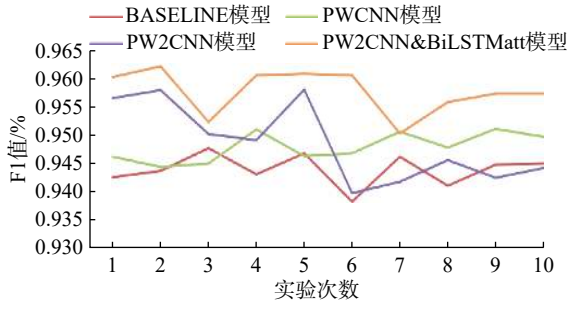

图 7　不同实验模型的 F1 值

Fig. 7　F1 values of different experimental models

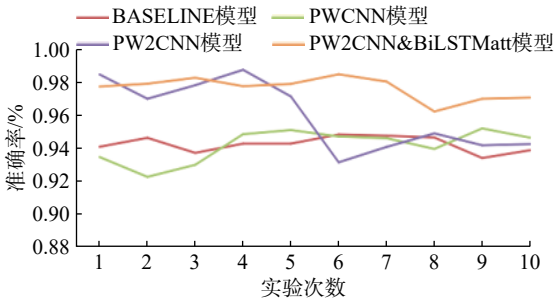

图 8 不同实验模型的准确率

Fig. 8 Accuracy of different experimental models

并行的 CNN 和 BiLSTM 模型(加入了 attention)模型。

3.6 不同模型之间的显著性检验

t 检验是一种显著性检验的方法，以小概率反证法进行逻辑推理，判断假设是否成立。这种检验方法可以用于检验两个样本平均值差异程度，其主要是通过 *t* 分布理论来推断差异发生的概率，进而判断两个平均数是否存在显著差异。这里使用 *t* 检验的原因是用来检验不同模型的实验结果是否存在显著差异。表 3 为 4 个模型之间的 *t* 检验值。

表 3 4 个模型间的 *t* 检验

Tab.3 *t* test among four models

模型	BASELINE	PWCNN	PW2CNN	PW2CNN&BiLSTMatt
PWCNN	0.00339*	–	–	–
PW2CNN	0.03128	0.36937	–	–
PW2CNN&BiLSTMatt	2.8009×10^{-9}**	3.5968×10^{-7}**	0.00024**	–

注: *表示小于显著性水平α = 0.05, 存在显著性差异; **表示小于显著性水平α = 0.001, 存在显著性差异

当显著性水平为 0.001 时，PW2CNN&BiLSTMatt 模型对 BASELINE 模型、PWCNN 模型和 PW2CNN 模型的 *t* 检验的 *p*-value 值(2.800 9×10^{-9}、3.596 8×10^{-7}和 0.000 24)均小于 0.001。因此，PW2CNN&BiLSTMatt 模型与 BASELINE 模型、PWCNN 模型和 PW2CNN 模型之间均存在显著性差异。换句话说，在显著性水平为 0.001 的条件下，PW2CNN&BiLSTMatt 模型显著优于 BASELINE 模型、PWCNN 模型和 PW2CNN 模型。此外，当显著性水平为 0.05 时，PWCNN 模型对 BASELINE 模型 *t* 检验的 *p*-value 值(0.003 39)均小于 0.05。因此，PWCNN 模型与 BASELINE 模型之间存在显著性差异。换句话说，在显著性水平为 0.05 的条件下，PWCNN 模型显著优于 BASELINE 模型。

3.7 实例分析

对实验结果(分类过的评论)进行具体的分析，从不同角度对好评和差评进行数据分析。评论数据的标签是通过 PW2CNN&BiLSTMaat 模型分类出来的结果(好评/差评)，对分类后的评论进行相应的分词处理并且统计其出现的次数，再将好评和差评中有意义的高频词选出来进一步分析。可以利用去停用词的方法将“电视”/“电视机”或品牌名称等无意义的词去掉，再通过统计数量的多少进行从大到小的排列，将前面的高频词和词频数保留下来。好评论和坏评论的词频统计图如图 9 和图 10 所示。

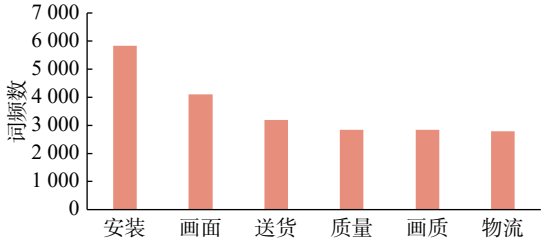

图 9 好评高频词统计图

Fig.9 Statistical chart with high word frequency in positive reviews

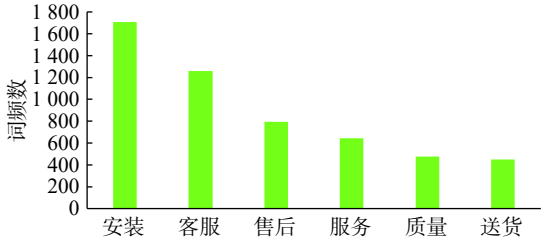

图 10 差评高频词统计图

Fig.10 Statistical chart with high word frequency in negative reviews

可以看出，“安装”对于顾客满意度有重要影响，其次就是对于“画面”/“画质”的满意程度，也是顾客关注的重点。另外还有“送货”/“物流”方面，也是顾客所重点关注的问题。因此，商家在实际情况下，为了得到更多的好评，这几个方面是不能忽视的。

4 结束语

对评论文本的情感进行挖掘和分析，提出一种对在线电商评论情感分类的 PW2CNN&BiLSTMatt模型，同时使用了词向量以及 POS 特征向量。这两个特征是本文针对文本数据信息提取的两种方式，主要考虑到了词本身的特性和含情感信息的特征特性。通过 BASELINE 模型和 PWCNN 模型的对比，验证了词性特征在模型中的

作用。与此同时，对这两个特征的融合方式提出了两种方案，一种是两个向量直接拼接的融合方式，另一种是通过并行结构神经网络模型的方式进行“融合”。

对卷积神经网络和加入 attention 机制的双向长短期记忆网络进行了对比，并通过实验结果证明了哪个模型更适合。卷积网络擅长抓取局部特征，而长短期模型更适合含有“时间”信息的特征，因此什么样的特征用什么样的深度学习算法模型至关重要。

本文不足之处在于未进行多种参数的设定尝试。此外，关于并行结构也可以进一步作不同尝试，对此类研究将在未来继续展开。

参考文献：

[1] PANG B, LEE L, VAITHYANATHAN S. Thumbs up?: sentiment classification using machine learning techniques[C]//Proceeding of 2002 Conference on Empirical Methods in Natural Language Processing. Philadelphia, PA, USA: ACM, 2002: 79–86.

[2] LIU L Z, SONG W, WANG H S, et al. A novel feature-based method for sentiment analysis of Chinese product reviews[J]. China Communications, 2014, 11(3): 154–164.

[3] LIN Y O, LEI H, WU J, et al. An empirical study on sentiment classification of Chinese review using word embedding[C]//Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation. Shanghai, China, 2015: 258–266.

[4] KIM Y. Convolutional neural networks for sentence classification[C]//Proceedings of 2014 Conference on Empirical Methods in Natural Language Processing. Doha, Qatar: Association for Computational Linguistics, 2014: 1746–1751.

[5] ZHAO W, GUAN Z Y, CHEN L, et al. Weakly-supervised deep embedding for product review sentiment analysis[J]. IEEE Transactions on Knowledge and Data Engineering, 2018, 30(1): 185–197.

[6] HASSAN A, MAHMOOD A. Convolutional recurrent deep learning model for sentence classification[J]. IEEE Access, 2018, 6: 13949–13957.

[7] LAZIB L, QIN B, ZHAO Y Y, et al. A syntactic path-based hybrid neural network for negation scope detection[J]. Frontiers of Computer Science, 2020, 14(1): 84–94.

[8] ABDI A, SHAMSUDDIN S M, HASAN S, et al. Deep learning-based sentiment classification of evaluative text based on Multi-feature fusion[J]. Information Processing & Management, 2019, 56(4): 1245–1259.

[9] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278–2324.

[10] 陈龙, 管子玉, 何金红, 等. 情感分类研究进展 [J]. 计算机研究与发展, 2017, 54(6): 1150–1170.

[11] LI W, ZHU L Y, SHI Y, et al. User reviews: sentiment analysis using lexicon integrated two-channel CNN–LSTM family models[J]. Applied Soft Computing, 2020, 94: 106435.

[12] TANG F L, FU L Y, YAO B, et al. Aspect based fine-grained sentiment analysis for online reviews[J]. Information Sciences, 2019, 488: 190–204.

[13] 陈佳. 在线评论对消费者态度演化和购买行为的影响研究 [D]. 成都: 电子科技大学, 2018.

（编辑：丁红艺）