文章编号：2095-2163(2024)01-0063-07 中图分类号：TP391 文献标志码：A

基于情感融合和层次注意力机制的情感分析

邵 清，张文双，李 颖，王少军
（上海理工大学 光电信息与计算机工程学院，上海 200093）

摘 要：针对两个反义词在相似语境下转化成词向量后空间距离相近，容易造成情感信息丢失，循环神经网络等的特征提取方式容易导致网络依赖增强，难以充分提取局部性特征。针对第一个问题，本文提出情感嵌入模块，在词嵌入的过程中加入情感向量与语义信息作为网络的输入层；针对第二个问题，本文提出层次注意力机制，将融合后的词向量切片形成两个子序列，将单词的词向量输入到双向门控循环网络，利用注意力机制对隐藏层进行加权计算，获得子序列文本信息，通过多个网络层获得整个文本序列信息；最后，经过 softmax 函数输出文本情感极性。在 NLPIR 微博语料库和 NLPCC2014 的微博公开数据集进行实验，表明该情感分析模型在准确率上有所提高，证明了模型的有效性。

关键词：情感嵌入；层次注意力；双向门控循环网络；情感分析

Sentiment analysis based on sentiment fusion and hierarchical attention mechanism

SHAO Qing, ZHANG Wenshuang, LI Ying, WANG Shaojun
(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: For two antonyms converted into word vectors in similar contexts, the spatial distances are similar, which is easy to cause loss of emotional information, and feature extraction methods such as recurrent neural networks can easily lead to network dependence enhancement and it is difficult to fully extract local features. For the first question, this paper proposes a sentiment embedding module, which adds emotion vector and semantic information as the input layer of the network in the process of word embedding. Aiming at the second problem, this paper proposes a hierarchical attention mechanism to slice the fused word vector into two subsequences. The word vector of the word is input into the bi-directional gating recurrent network, the attention mechanism is used to perform weighting calculation on the hidden layer to obtain the subsequence text information, and the entire text sequence information is obtained through multiple network layers. Finally, the sentiment polarity of the text is output through the softmax function. Through experiments on the NLPIR Weibo corpus and the NLPCC2014 Weibo public data set, the accuracy of the sentiment analysis model has been improved, which proves the effectiveness of the model.

Key words: sentiment embedding; hierarchical attention; bidirectional gated recurrent network; sentiment analysis

0 引 言

微博是一个全员参与的社交平台，微博用户发帖话题广泛，包括日常生活分享、评论社会事件、电影等。微博帖子中包含大量的情感信息和观点信息，利用社交媒介中的数据进行情感分析，能更了解事件的发展趋势，帮助人们做出决策。

文本的情感分析属于自然语言处理的重要部分，现有的研究中许多情感分类的深度学习模型都是基于循环神经网络（Recurrent Neural Networks，RNNS）提出的。RNNS 专门用于顺序建模，但无法以并行的方式提取特征，对于长数据序列，传统的 RNNS 会产生梯度爆炸以及梯度消失的问题。长短期记忆（Long Short-Term Memory，LSTM）是一种具有长期和短期特性的特殊 RNNS 体系结构，存储单元作为隐藏单元，有效地解决了梯度消失和梯度爆

基金项目：国家重点研发计划项目（2018YFB1702601）。

作者简介：邵 清（1970-），女，博士，副教授，CCF 会员，主要研究方向：网络智能与自然语言处理；张文双（1996-），女，硕士研究生，主要研究方向：自然语言处理；李 颖（1997-），女，硕士，主要研究方向：自然语言处理；王少军（1998-），男，本科生，主要研究方向：自然语言处理。

收稿日期：2023-01-01

炸问题。对于自然语言处理来说,LSTM 可以提取高层次的文本信息。双向长短期记忆(Bidirectional Long Short-Term Memory, BiLSTM)是 LSTM 的进一步发展,BiLSTM 将前向隐藏层和后向隐藏层结合起来,可以访问正向和后向的上下文序列。目前 LSTM 和 BiLSTM 已经被应用到文本分类中,并取得了一些成果。

在自然语言处理领域,Bahdanau 等[1]首次将注意力机制应用于自然语言处理领域,在机器翻译任务上将翻译和对齐同时进行,实验证明了注意力机制在自然语言处理(Natural Language Processing, NLP)任务中的有效性,注意力机制很好地把源语言端的每个词和当前要预测翻译的词联系起来,效果有很大提升。目前,注意力机制已经应用于解决文本分析中的问题。例如文本生成、机器翻译、机器阅读理解。Wang 等[2]将多层注意力机制和卷积神经网络结合用于句子关系分类任务中,模型在多个数据集上的实验结果表明,使用注意力机制的模型比未使用的模型有更高的分类精度;M.E. Basiri 等[3]提出一种基于注意力的双向 CNN-RNN 情感分析深度模型,将注意力机制用于 BiLSTM 和 BiGRU 分支的输出,使模型能够或多或少地关注不同单词和句子,然后使用卷积层提取信息特征并减少输入数据的维数,同时全局和平均池化层堆叠在 CNN 层的输出处,对其特征图进行下采样,使得生成的特征图对特征的位置变化更具有鲁棒性。

在训练词向量时相同的语境中具有相反意思的词语,词向量的空间距离很近。比如“这家饭店的服务很好”和“这家饭店的服务很差”,这两个句子中“很差”和“很好”是两个情感相反的形容词,但余弦相似度计算显示两个词构建的空间向量很近,说明这两个形容词相关度很高。ZOUH 等[4]将情感词结合卷积神经网络提高情感分类的结果;Yang 等[5]提出了一种情感语义增强的 CNN,该 CNN 可以从表情嵌入和单词级文本向量中捕获情感语义特征,从而提高 CNN 在微博情感分类中的性能;M.A. Ullah 等[6]提出一种同时使用文本和表情符的情感分析算法,分别在机器学习和深度学习对算法进行分析,通过对比实验证明了表情符与语义信息的结合对模型的有效性;陶永才等[7]提出多特征融合情感分析,在语义向量中加入情感符号、词性特征、情感标签特征,构建多种情感特征组合,使得情感分类能力得到有效提升;王行甫等[8]提出了一种结合词性、位置和单词情感的网络模型实现方面级情感分析。

大多数的神经网络模型可以对句子进行编码,使得输出的隐藏层中包含单词的上下文信息。对于较长文本,神经网络模型不能完全掌握句子间的序列关系,在构建语言时序特征中,词信息是依次反馈的,耗费时间较长,难以实现并行计算,句子中单词间的依赖关系会随着距离的增加和时间的增强而减弱。此外,结合注意力机制的神经网络模型只能进行单次注意力计算,导致模型无法对句中单词间的依赖关系和局部特征实现深层次抽取。Yang 等[9]使用的单词级别的注意力权重计算方式提高了模型的分类效果;Yu 等[10]提出切片递归神经网络,通过将输入的序列切片成多个子序列来并行化,通过多个层来获取更高级信息的能力,在情感分类任务中实现了更好的性能。

本文提出了一种基于情感融合和层级注意力机制的微博情感分析模型(Emotion and Hi-erarchical Attention Network, E-HAN),利用情感词的情感信息构造出情感向量,与评论生成的语义向量进行相关计算来表示文本,并将文本切片形成子序列,在词向量级和子序列级独立进行双向门控循环单元(Gated Recurrent Unit,GRU)特征提取,增加语义提取的深度,提高训练效率,优化了训练时间,并在每个子序列上增加注意力层,充分提取局部性特征,突出文本关键信息和文本序列中的关键信息。模型在 NLPCC2014 的微博公开数据集和 NLPIR 微博语料库数据集上进行实验,并与一些经典模型进行对比实验,本文的 E-HAN 模型在各评估指标下表现良好,证明该模型分类的有效性。

1 基本理论

1.1 Glove 词嵌入模型构建

由于神经网络模型只能接受数值输入,所以就需要将字符文本进行数值化表示。通常利用词向量模型将句子进行分词处理后将词语映射到多维空间中。Glove 的开发者表示 word2vec 只是一个与局部上下文窗口相关的模型,很少在语料库中使用一些统计信息。因此,Jeffrey Pennington 等[11]引入了 Glove,结合了全球统计信息和本地上下文信息,主要特点是引入了共现概率矩阵。共现矩阵用 $\boldsymbol{X}$ 表示,字典维数用 n 表示,X_{ij} 是整个语料库中词 i 和词 j 同时出现在一个窗口的次数,使用 x_i 来表示所有单词(窗口中的)和单词(中心单词)出现的总数,P_{ij} 用来表示顺序为 j 的单词出现在单词 i 窗口中的概率,式(1)和式(2):

$$X_i = \sum_k X_{ik} \tag{1}$$

$$P_{ij} = p(j \mid i) = \frac{X_{ij}}{X_i} \tag{2}$$

P_{ik}/P_{jk} 的比率取决于 i,j,k 的位置，利用 P_{ik}/P_{jk} 去拟合未知函数 F，式(3)

$$F(w_i, w_j, \tilde{w}_k) = \frac{P_{ik}}{P_{jk}} \tag{3}$$

其中，$w_i, w_j \in R^d$ 表示中心词的词向量，$\tilde{w}_k \in R^d$。

经过一系列推导，得到代价函数 J，式(4)：

$$J = \sum_{i,j=1}^{n} f(X_{ij})(w_i^{\mathrm{T}} \tilde{w}_j + b_i + b_j - \log X_{ij})^2 \tag{4}$$

$f(x)$ 是一个权重函数，式(5)：

$$f(x) = \begin{cases} \left(\dfrac{x}{x_{\max}}\right)^{\alpha}, & \text{if } x < x_{\max} \\ 1, & \text{otherwise} \end{cases} \tag{5}$$

其中，α 的值为 3/4。

1.2　GRU 时序捕捉

上下文关系的捕捉在获取评论文本深层特征时尤为重要，如果使用单个卷积神经网络模型，则句子的语义特征信息可能会被丢弃。如果使用单个的循环神经网络模型，句子中单词间的依赖关系会随着距离的增加和时间的增强而减弱。LSTM 能够体现文本序列特征，但结构复杂性很高，计算代价大。GRU 在 LSTM 基础上做了改进，减少了“门”的数量，包含重置门 r 确定新输入和之前信息的组合，更新门 z 判断要传递先前的多少信息，因此结构更加简单，GRU 模型结构如图 1 所示。

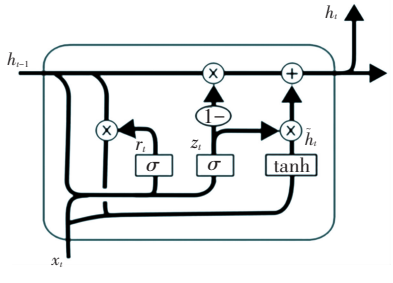

图 1　GRU 结构图

Fig. 1　GRU structure diagram

获取当前词的词序特征时不仅要考虑历史信息，还要和下文密切相关，所以本文采用 BiGRU 充分考虑当前词的上下文特征，结构如图 2 所示。

在 t 时刻将语义矩阵输入到 BiGRU 中，输出为该文本的语义特征，BiGRU 对应的各状态值式(6)～式(9)所示：

$$r_t = \sigma(w_t \cdot [h_{t-1}, x_t]) \tag{6}$$

$$z_t = \sigma(w_z \cdot [h_{t-1}, x_t]) \tag{7}$$

$$h_t' = \tanh(w_h' \cdot [r_t * h_{t-1}, x_t]) \tag{8}$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * h_t' \tag{9}$$

其中，r_t 和 z_t 分别表示重置门和更新门，用来控制和有选择地挑出需要的信息；x_t 表示在 t 时刻的输入；h_t' 为候选激活函数，用来计算 t 时刻的隐藏层状态 h_t。

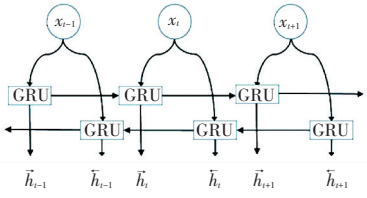

图 2　BiGRU 结构图

Fig. 2　BiGRU structure diagram

1.3　Attention 特征提取

判定一个语句的情感倾向时并非所有的单词在句子中对情感判断都有贡献。利用注意力机制能够捕获一个语句中单词之间的一些句法或者语义特征，更容易捕获长距离的依赖特征。对一个句子中不同的词分配不一样的权重，使得远距离之间依赖特征会被大大地缩短。具体实现步骤如式(10)～式(12)所示：

$$u_{ij} = \tanh(w_t h_{ij} + b_t) \tag{10}$$

$$a_{ij} = \frac{\exp(u_{ij} \boldsymbol{u}_t)}{\sum_j \exp(u_{ij} \boldsymbol{u}_t)} \tag{11}$$

$$k_i = \sum_j a_{ij} h_{ij} \tag{12}$$

其中，h_{ij} 表示输入第 i 个子序列中的第 j 个单词；u_{ij} 表示经过一个全连接神经网络模型后的隐含表示；a_{ij} 表示 u_{ij} 与一个包含上下文信息向量 $\boldsymbol{u}_t$ 做相似度计算后的权重值；k_i 表示在对原始输入进行加权后的特征表示。

2　算法框架

E-HAN 模型结构如图 3 所示，共有 3 部分构成：情感语义构建层、特征提取层、情感分类层。该模型首先在语义向量中加入情感信息，构建情感语义特征矩阵；通过层次注意力深层次提取局部语义特征，屏蔽无用的噪声特征，对提取的特征进行情感分类。

2.1　构建情感语义矩阵

对微博语料进行预处理包括分词、去停用词、去除话题、标点以及特殊字符，之后使用 Glove 词嵌入

工具获取语义向量。共现窗口设置为 15,输出向量维数设置为 100,经过 Glove 工具训练之后每个词被转化成 100 维的语义向量。由于微博设定评论最多输入 140 个字符,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,得到微博文本原始语义矩阵 $\boldsymbol{K}=[\alpha_1,\alpha_2,\cdots,\alpha_n]_{d\times n}$,$n$ 表示微博中分词的个数,d 表示维数。

情感词典选用大连理工大学信息检索实验室整理标记的中文本体资源 DUTIR,以同样的方式训练情感词典,得到矩阵 $\boldsymbol{M}=[\beta_1,\beta_2,\ldots,\beta_m]_{d\times m}$,$m$ 为情感词的个数。利用情感矩阵的每一行和语义矩阵每一列做余弦相似度计算,构建情感融合语义矩阵 $\boldsymbol{C}$,使语义矩阵中包含情感信息,式(13):

$$\boldsymbol{C}=\cos(\alpha_i,\beta_j)=\frac{\alpha_i,\beta_j}{|\alpha_i||\beta_j|};i\in[1,n],j\in[1,m] \tag{13}$$

情感语义矩阵建立过程如图 4 所示。

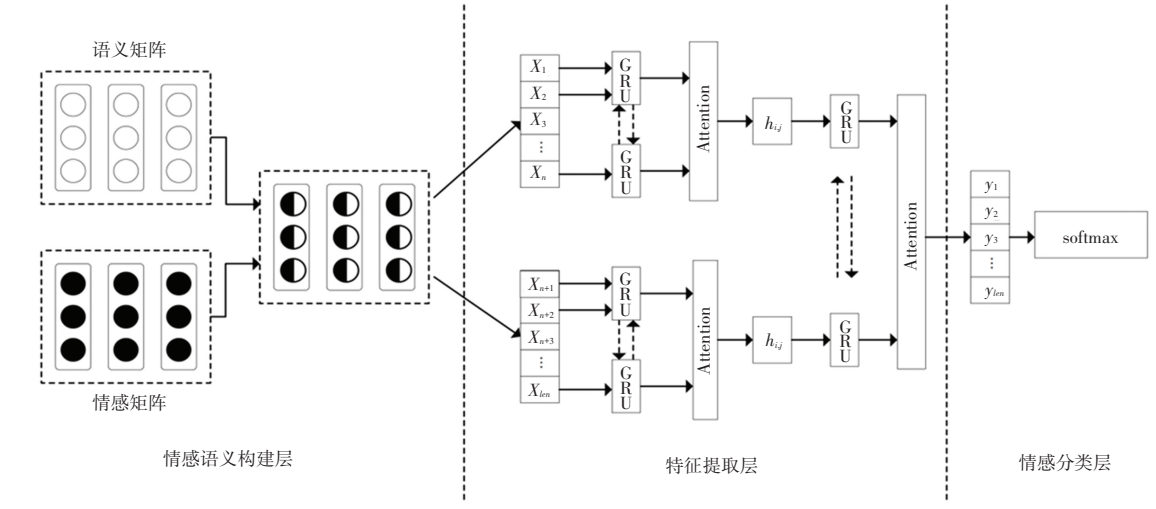

图 3 E-HAN 模型结构

Fig. 3 E-HAN model structure

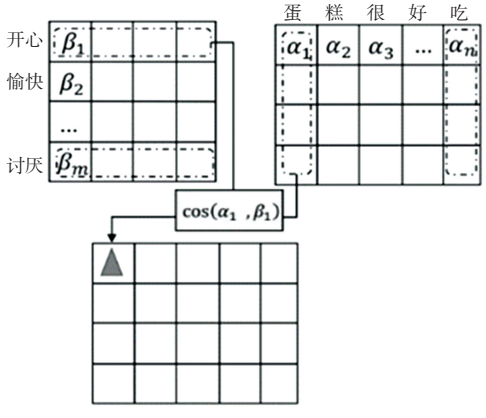

图 4 情感语义矩阵建立过程

Fig. 4 The process of establish the sentiment semantic matrix

2.2 层次注意力计算

传统的注意力机制中源端和目标端的内容是不一样的,得到的结果即源端的每个词与目标端每个词之间的依赖关系。为挖掘微博语料中更加深层次和各分词多样有效的特征,本文采用层次注意力机制。将矩阵 $\boldsymbol{C}$ 作为输入序列并划分成两个长度相等的子序列。利用句子的层次结构特性,先使用单词的词向量组成表示子序列,在此基础上以子序列向量构建长句的信息表示。具体操作步骤如下:$\boldsymbol{C}$ 作为层次注意力机制的输入,使用 python 中的切片操作,将矩阵 $\boldsymbol{C}$ 分裂为 2 个等长的子序列,那么每个子序列的长度 len 表示为式(14):

$$len=\frac{T}{2} \tag{14}$$

其中,T 表示每条数据集序列总长度。

每个子序列 N 可以表示为式(15):

$$N=[x_1,x_2,x_3,\cdots,x_{len}] \tag{15}$$

2.2.1 基于单词级的注意力

2.2.1.1 GRU 层

对于子序列表示中的单词并不是每一个单词对分类任务都是有用的,如在做文本的情绪分类时可能会比较关注“很好”、“伤心”这些形容词。为了使神经网络能自动将“注意力”放在这些词汇上采用了基于单词的注意力模型。将两个子序列作为输入序列,并行地输入到 BiGRU 中,在相反方向上分别进行处理,式(16)和式(17),设置 BiGRU 的隐藏单

元为 128。

$$\overrightarrow{h_{it}} = \overrightarrow{GRU}(x_{it}),t \in [1,len] \tag{16}$$

$$\overleftarrow{h_{it}} = \overleftarrow{GRU}(x_{it}),t \in [len,1] \tag{17}$$

其中，x_{it} 表示子序列 i 中的第 t 个词向量。

2.2.1.2　Attention 层

Attention 层为子序列每个单词分配不同权重。步骤如下：

(1)将 BiGRU 输出的隐藏层序列 $\boldsymbol{h}_{ij}$ 输入到单层全连接神经网络模型，获得隐含表示 u_{ij}；

(2)计算 u_{ij} 和一个表示上下文向量的 u_w 的相似度，再通过 softmax 函数归一化处理矩阵权重值 a_{ij}；

(3)对原始矩阵 $\boldsymbol{h}_{ij}$ 进行加权求和，得到 k_i 子序列的局部重要特征提取值。

2.2.2　基于子序列的注意力

将加权后的子序列 k_1 和 k_2 作为 GRU 的输入，在两个方向上分别获取句子的前向和后向隐藏序列，式(18)和式(19)：

$$\overrightarrow{h_i} = \overrightarrow{GRU}(k_i) \tag{18}$$

$$\overleftarrow{h_i} = \overleftarrow{GRU}(k_i) \tag{19}$$

对于 Attention 层做同样的处理，输入单层全连接神经网络模型获得 h_i 的隐含表示 u_i，之后通过相似度来衡量单词的重要性，最后通过词向量加权求和的方式得到序列的最终表示 $\boldsymbol{Y}$。

2.3　情感极性分类

将前向状态 $\overrightarrow{Y}$ 和后向状态 $\overleftarrow{Y}$ 进行拼接形成 $\boldsymbol{Y}$。由于 $\boldsymbol{Y}$ 是一个向量，不能根据 $\boldsymbol{Y}$ 直接判断出文本属于哪一类情感，因此使用 softmax 进行归一化处理操作。使用 softmax 函数计算每个类别的条件概率，选取条件概率最大值对应的类别作为预测 d 输出类别。计算过程如式(20)~式(22)所示：

$$\boldsymbol{Y} = \text{concat}(\overrightarrow{Y},\overleftarrow{Y}) \tag{20}$$

$$p(y' \mid S) = \text{softmax}(\boldsymbol{W}_o\boldsymbol{Y} + \boldsymbol{b}_o) \tag{21}$$

$$y = \text{argmax}p(y' \mid S) \tag{22}$$

其中，$\boldsymbol{W}_o$ 表示可学习权重向量，$\boldsymbol{b}_o$ 表示偏置。

本文模型训练的目标为最小化分类交叉损失熵，式(23)：

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m} t_i \log(y'_i) \tag{23}$$

3　实验研究

3.1　实验数据集

本文采用 NLPIR 微博语料库和 NLPCC2014 的微博公开数据集作为情感评测任务的数据集。NLPIR 微博内容语料库由北京理工大学张华平博士从新浪微博公开采集与抽取获得，原始的 NLPCC2014 数据集每一条微博语句都标注情感标签，分别为“厌恶”、“生气”、“伤心”、“害怕”、“高兴”、“喜爱”、“惊喜”和“无情感”8 类标签；NLPIR 微博语料库并未对每条文本语句进行情感分类。

由于这 8 类情感标记的频率有很大差别，因此对文本进一步分类，将“高兴”、“喜爱”、“惊喜”作为积极情感标 1，将“厌恶”、“生气”、“伤心”、“害怕”作为消极情感标 2，“无情感”作为中性标记 0。具体实验数据见表 1。

表 1　实验使用的训练数据统计

Table 1　Statistics of training data used in the experiment

数据集	积极	消极	中性
NLPIR	12 826	14 069	13 105
NLPCC2014	14 189	12 790	13 021

3.2　实验测评

3.2.1　环境参数设置

深度学习模型的参数设置非常关键，本文实验主要参数与参数值见表 2。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。

表 2　实验主要参数与参数值

Table 2　Main parameters and parameter values of the experiment

参数	参数描述	参数值
Learning_rate	学习率	0.001
Batch_size	每批次处理数据量	128
Epochs_size	迭代轮数	20
Embedding_size	词向量长度	100
Hidden_unit	隐藏层单元数	128
dropout	神经元丢弃率	0.5

3.2.2　实验评价指标

评价分类模型的评测指标主要包括正确率、召回率以及 $F1$。正确率是指预测为正的样本中真正的正样本所占的比率。预测为正有两种可能，一种是把正类预测为正类(TP)，另一种是把负类预测为正类(FP)，正确率可以用式(24)表达：

$$P = \frac{TP}{TP + FP} \tag{24}$$

召回率是指样本中正例被预测正确的比率。也有两种情况，一是预测正确，即把原来的正类预测成正类(TP)，另一种是预测失败，即把原来的正类预测为负类(FN)，式(25)：

$$R = \frac{TP}{TP + FN} \tag{25}$$

有时准确率和召回率会出现相互矛盾的情况，不管是准确率还是召回率均不能比较全面地衡量模型的性能，所以引入 F-Measure（又称 F-score）方法。F-Measure 是精确率和召回率的加权调和平均，本文采用最常见的计算方法，式(26)：

$$F1 = \frac{2 * P * R}{(P + R)} \tag{26}$$

3.3 实验过程

具体实验步骤如下：

Step 1 标注并分类本次实验的数据集；

Step 2 使用 jieba 工具包进行分词，之后使用 Glove 工具训练词向量，文本中的停用词和特殊符号均当作普通单词保留；

Step 3 将句子转换成索引表示并固定长度为 100；

Step 4 每个索引对应一个词向量，将词向量拼接成矩阵，长度不够 100 的使用 0 填充，每个词向量维数为 100；

Step 5 将情感词典中的词表示成向量矩阵，与语义向量融合计算；

Step 6 将情感语义矩阵切片分为两个子序列作为 BiGRU 的输入；

Step 7 文本子序列索引表示输入嵌入层，再输入 BiGRU 提取文本深层特征；

Step 8 将提取到的深层次特征输入到 Attention 层，对每个词分配相应的权重值；

Step 9 将子序列的中间状态输入到层次 BiGRU 提取局部特征，再进入 Attention 层加权得到整条语句特征表示；

Step 10 利用测试数据集对模型进行验证。

3.4 对比实验

为了研究 E-HAN 模型对情感分类的效果，本文设计了两种实验：

(1) 6 种模型在 2 个不同数据集上进行实验，对比其他 5 种模型和 E-HAN 模型的分类效果；

(2) 使用 E-HAN 模型，情感词典中的情感词减少为原来的一半，之后与语义向量融合。

本文通过与 CNN、LSTM、BiLSTM、BiGRU-Attention、E-HAN 和 MHA-SSA 6 个模型的实验结果进行对比分析，结果见表 3。由表 3 可知，E-HAN 模型与其他模型相比准确率和 $F1$ 值得到了提高。通过对比模型 CNN、LSTM 和 BiLSTM 算法的实验结果可知，BiLSTM 由于增加了后向传播单元，可以同时考虑上下文信息，实验结果比 LSTM 更好、准确率更高。

通过对比模型 BiLSTM 和 BiGRU-Attention 的实验结果可知，加入注意力机制后模型能够获取句子中更多的关键特征，提高模型的准确率、召回率和 $F1$ 值。相较于 BiGRU-Attention 模型，E-HAN 模型的准确率、召回率和 $F1$ 值分别提高了 0.97%、0.78% 和 0.87%，因为 E-HAN 模型中使用的层次注意力机制能够比传统注意力机制提取文本更多的局部特征。

通过对比模型 MHA-SSA 和 E-HAN 的实验结果可知，E-HAN 模型 $F1$ 值提高了 0.41%，因为 E-HAN模型在词嵌入层将语义信息与情感信息相结合，不仅提供了更多的词语特征，还加强了句子的情感强度，除此之外还使用层次注意力机制更充分地获取局部隐藏信息。

表 3 NLPIR 实验结果

Table 3 NLPIR experiment results %

模型	P	R	$F1$
CNN	79.32	72.68	75.85
LSTM	78.46	75.09	76.74
BiLSTM	80.02	75.78	77.84
BiGRU-Attention	80.46	77.67	79.04
MHA-SSA	81.12	77.95	79.50
E-HAN	**81.43**	**78.45**	**79.91**

6 个模型在 NLPCC2014 数据集上的实验结果，见表 4。由表 4 可以看出，前 3 种单一神经网络模型中情感分析的准确率偏低，证明在单模型和复合模型的基础上添加注意力机制能有效提高模型分类的准确度，因为注意力机制可以分配不同的权重值给特征，让模型能更快地掌握重要特征。从对比结果可以看出 E-HAN模型比传统的神经网络模型准确率有明显提升。在 NLPCC2014 数据集上的分类效果比在 NLPIR 数据集中的好，通过分析数据集的特点可知，NLPIR 数据的平均句长要短于 NLPCC2014 数据的平均句长。

表 4 NLPCC2014 实验结果

Table 4 NLPCC2014 experiment results %

模型	P	R	$F1$
CNN	79.89	78.09	78.98
LSTM	80.46	78.98	79.71
BiLSTM	81.08	79.78	80.42
BiGRU-Attention	81.78	80.33	81.05
MHA-SSA	82.25	81.02	81.63
E-HAN	**82.76**	**81.35**	**82.05**

在研究情感资源的多少对实验的分类效果产生

的影响时，使用的数据集为 NLPCC2014，通过减少与语义向量融合的情感词，得到的实验结果见表 5。

表 5　情感词实验结果

Table 5　Experimental results of emotional words　%

模型	*P*	*F*1
三分之一情感词	86.96	88.74
一半情感词	87.29	88.83
全部情感词	**87.87**	**89.03**

由表 5 可知，使用全部的情感词得到的 *F*1 值比使用一半和三分之一的情感词得到的 *F*1 分别提高了 0.2%和 0.29%。如果将情感词典中的情感词数量减半，情感分类的性能将会降低，说明情感词的缺少，丢失了情感信息，会降低情感分类的效果。如果情感词典中的单词数量足够充足，情感足够丰富，那么融合情感向量到语义向量中能很好地提高分类效果。

4　结束语

本文提出了一种基于情感融合和层级注意力机制的情感分析模型。实验结果表明，与传统的深度学习模型相比，该模型进一步提高了算法的稳定性和有效性。

在研究过程中发现本研究的一些局限性。数据不仅仅是一个单词，而是更复杂的形式。例如文本混合了表达式、图片甚至短视频。此外，数据集的处理对于模型训练产生至关重要的影响，上述问题必然会导致数据预处理任务更加困难和复杂，潜在应用算法的整体复杂度也会更高。

参考文献

[1] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint arXiv: 1409.0473, 2014.

[2] WANG L, CAO Z, DE MELO G, et al. Relation classification via multi-level attentioncnns[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2016: 1298-1307.

[3] BASIRI M E, NEMATI S, ABDAR M, et al. ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis[J]. Future Generation Computer Systems, 2021, 115: 279-294.

[4] ZOU H, TANG X, XIE B, et al. Sentiment classification using machine learning techniques with syntax features[C]//2015 International Conference on Computational Science and Computational Intelligence (CSCI). IEEE, 2015: 175-179.

[5] YANG G, HE H, CHEN Q. Emotion-semantic-enhanced neural network[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2018, 27(3): 531-543.

[6] ULLAH M A, MARIUM S M, BEGUM S A, et al. An algorithm and method for sentiment analysis using the text and emoticon[J]. ICT Express, 2020, 6(4): 357-360.

[7] 陶永才，张鑫倩，石磊，等. 面向短文本情感分析的多特征融合方法研究[J]. 小型微型计算机系统，2020，41(6):7.

[8] 王行甫，王磊，苗付友，等. 结合词性、位置和单词情感的内存网络的方面情感分析[J].小型微型计算机系统，2019，40(2)：383-389.

[9] 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. 2016: 1480-1489.

[10] YU Z, LIU G. Sliced recurrent neural networks[J].arXiv preprint arXiv:1807.02291, 2018.

[11] PENNINGTON J, SOCHER R, MANNING C D. Glove: Global vectors for word representation[C] //Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014: 1532-1543.

(上接第 62 页)

[20] BEVILACQUA M, ROUMY A, GUILLEMOT C, et al. Low-complexity single image super-resolution based on nonnegative neighbor embedding[C]// British Machine Vision Conference. BMVA Press, 2012:1-10.

[21] ZEYDE R, ELAD M, PROTTER M. On single image scale-up using sparse-representations[C]// Proceedings of the 7th International Conference on Curves and Surfaces Avignon, France: Springer, 2010:1-20.

[22] MARTIN D, FOWLKES C, TAL D, et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics[C]// IEEE International Conference on Computer Vision. IEEE, 2002: 416-423.

[23] HUANG J B, SINGH A, AHUJA N. Single image super-resolution from transformed self-exemplars[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2015: 5197-5206.