基于 RoBERTa-WWM 的多特征情感分析模型

张宏伟[1] 李晓晔[1]
ZHANG Hongwei LI Xiaoye

摘 要 针对传统词向量模型无法有效解决多义词表征、经典网络模型存在文本特征提取不足和无法捕捉特征的关键信息等问题，提出了一种基于 RoBERTa-WWM 的多特征情感分析模型。首先采用 RoBERTa-WWM 模型动态生成词向量，解决了多义词表征问题，丰富了语义特征；接着将多个 BIGRU 进行堆叠形成 Multi-Layer BIGRU，并与 TextCNN 结合提取局部和全局特征，让模型从不同角度学习语义特征；最后通过多头注意力机制，将局部和全局特征进行融合赋予不同的权重，以突出关键信息和文本的情感极性。在酒店评论数据集和商品评论数据集上进行了实验，结果显示，所提出的模型与其他模型相比，准确率和 F_1 值都有显著提升。

关键词 情感分析；RoBERTa-WWM；卷积神经网络；多层双向门控循环单元；多头注意力机制

doi：10.3969/j.issn.1672-9528.2024.05.008

0 引言

随着移动互联网的高速发展，越来越多的互联网用户喜欢主动发表自己的观点，其中往往带有个人情感观点的文本信息，诸如微博评论、商品评论、直播评论等。因此，对于主观性言论进行情感分析的研究越来越受重视，通过情感分析能够获得更多有价值的信息。

文本情感分析利用了文本挖掘、机器学习和深度学习等技术，旨在处理带有情感色彩的主观性文本，以实现情感抽取和分类的目标 [1]。基于情感词典法 [2] 的效果主要取决于情感词典的质量，而机器学习方法 [3] 在面对大量的数据时，泛化能力有待提高。目前主流的做法是深度学习，它以出色的自动信息表示学习能力和优异的识别效果而备受关注，在情感分析领域有着重要的地位。常见的深度学习模型有卷积神经网络和循环神经网络等。Kim[4] 最早提出了将卷积神经网络（CNN）应用于文本情感分析领域，他将预训练的词向量输入到不同大小卷积核的卷积神经网络中来提取特征，并在句子级别上进行分类。相对于传统机器学习算法，这种方法明显提高了情感分析的准确度。Mikolov[5] 将 RNN 应用到情感分析领域中，相比于 CNN，RNN 在捕获长距离依赖上更具优势，但随着输入量的增大，RNN 在感知初始输入的能力方面有所衰弱，会导致梯度弥散或梯度爆炸等问题。随着 LSTM[6] 和 GRU[7] 等变种的提出，梯度弥散或梯度爆炸问题得到有效改善。Xiao 等人 [8] 提出了双向 LSTM 模型，应用于情感分析领域，利用双向神经元的循环计算方法，提取文本所蕴含的信息。吴贵珍等人 [9] 将 CNN 和 BIGRU 结合起来，并融合了注意力机制，充分利用同一层次上的特征信息，增强句子体系特征的表示，从而提升模型的特征学习能力。

文本词向量表示对于情感分析任务同样重要。Word2Vec 和 GloVe 方法的基本思想是通过预测词语在上下文中出现的概率和向量值来训练词向量，使得相似的词在向量空间中距离更近。这些方法经过大量的实验和应用证明，对于文本分类、句子相似度计算等自然语言处理任务都有非常好的效果。最早使用 one-hot 编码对文本进行离散化表示，后来有了新的词嵌入方法诞生，比如 Word2Vec[10] 和 GloVe[11] 等模型，它们都属于静态词向量模型，无法解决一词多义的问题。Devlin 等人 [12] 提出了 BERT 模型，通过对大规模的语料进行训练，实现对不同语义环境下词向量的动态表示，它的优秀性能为情感分析任务提供了极大的便利。Liu Y 等人 [13] 对 BERT 进行改进，提出了 RoBERTa 模型、使用全词掩码策略进行预训练的 RoBERTa-WWM 模型 [14] 等。

基于以上工作，本文提出一种基于 RoBERTa-WWM 的多特征情感分析模型，主要工作如下。

（1）使用 RoBERTa-WWM 模型做词嵌入，得到文本向量表示。

（2）将得到的向量分别输入到 TextCNN 与 Multi-Layer BIGRU 模型中进行局部和全局特征的充分提取。

1. 齐齐哈尔大学计算机与控制工程学院 黑龙江齐齐哈尔 161000
[基金项目] 黑龙江省省属高等学校基本科研业务费科研项目（145209124）

（3）将局部和全局特征进行融合，通过多头注意力机制为其赋予不同的权重，突出关键信息。

（4）通过全连接层对其进行情感分类。

1 相关工作

1.1 RoBERTa-WWM 模型

文献 [12] 中的 RoBERTa-WWM 模型是在 BERT 模型基础上进行改进的先进模型。它不仅继承了 BERT 模型的优点，还对模型结构和数据层面进行了改进。在预训练阶段，使用了中文全词遮掩技术。以“交通比较方便，周围的小饭店比较多”为例，如表 1 所示。这样做的好处是生成的语义表示不仅包含整个句子的信息，还包含单词级别的信息。为了遮掩组成同一个词的汉字，模型引入了动态掩码机制，通过这一机制，模型可以学习到不同的语言表征。这样，模型能够更好地理解语言的结构和语义，提高对汉字级别的处理能力。

表 1 不同 Mask 策略产生的结果

Mask 策略	例句
原始 mask	交通比较 [mask]，周围的小饭店比较多
WWM mask	交通比较 [mask][mask]，周围的小饭店比较多

1.2 卷积神经网络

卷积神经网络由于其强悍的局部特征提取能力，在情感分析领域也同样适用。它包括以下几个部分。卷积层是 CNN 模型中最重要的部分之一，它使用多个卷积核对输入数据进行扫描，从而提取出数据中的各种特征。池化层则用于对卷积层输出数据进行下采样，缩小数据规模，并减少数据中的噪声和冗余信息。其模型结构图如图 1 所示。

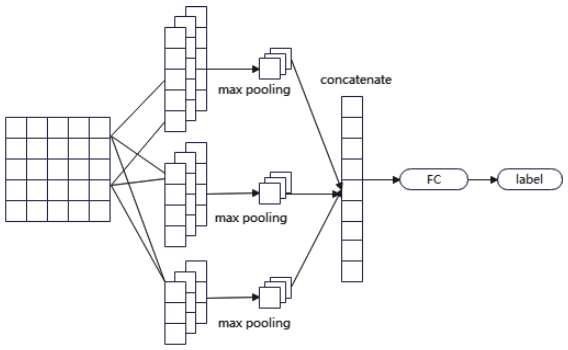

图 1 卷积神经网络结构图

1.3 BIGRU

GRU 是 RNN 的一种改进形式，其采用了独特的门控结构，可以解决传统的 RNN 中存在的梯度弥散和梯度爆炸的问题。与 LSTM 相比，GRU 模型结构简洁，参数量更少，可以有效提高训练效率。GRU（gated recurrent unit）结构包含了更新门、重置门和隐藏层，通过这些门的控制，GRU 可以很好地处理长序列输入。其中，重置门可以控制历史信息在当前状态中的占比，决定有多少过去的信息需要被保留在当前状态中。更新门能够控制历史信息在当前状态中的占比，决定有多少新信息需要被融合在当前状态中。模型结构图如图 2 所示。

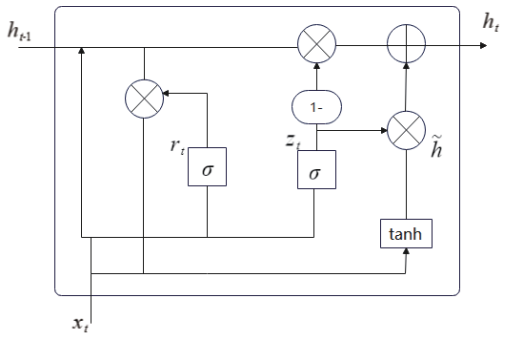

图 2 循环门控单元网络结构图

图 2 中，x_t 表示 t 时刻的输入，r_t 表示 t 时刻的重置门，z_t 表示 t 时刻的更新门，σ 表示 *Sigmoid* 激活函数，h_{t-1} 和 h_t 分别表示上一时刻隐层的输出和 t 时刻的隐层的状态，$\tilde{h}_t$ 表示时刻候选激活状态。具体计算过程如下：

$$r_t = \sigma(W_r[h_{t-1}, x_t]) \quad (1)$$

$$z_t = \sigma(W_z[h_{t-1}, x_t]) \quad (2)$$

$$\tilde{h}_t = \tanh(W \bullet [r_t h_{t-1}, x_t]) \quad (3)$$

$$h_t = (1 - z_t)h_{t-1} + z_t \tilde{h}_t \quad (4)$$

式中：W_r 和 W_z 都是权重矩阵，$\bullet$ 表示矩阵相乘。BiGRU 是由向前和向后的 GRU 组合而成的，能够同时捕捉到文本中正向和逆向的语义信息。结合上下文，BiGRU 可以深层次地提取文本中所蕴含的情感特征。而 Multi-Layer BIGRU 利用堆叠方式完美地提取文本全局特征。

1.4 多头注意力机制

注意力机制是一种能够针对关键信息进行优先关注，并聚焦于重要信息、淡化非重要信息的机制。该机制通过模拟人类大脑中注意力资源的分配方式，对不同特征向量进行加权计算，以实现高质量特征提取。

Bahdanau 等人 [15] 将 Attention 机制应用于机器翻译任务，通过建立源语言端单词与目标语言端待预测单词之间的对应关系，显著提高了模型的准确性。王锐等人 [16] 融合双向 LSTM 和注意力机制进行情感分析研究，随着输入文本长度的增加，编码过程中产生的中间向量可能会带来信息的丢失，从而导致分类精度的急剧下降。然而，注意力机制可以有效地缓解这种问题。通过引入注意力机制，模型能够更加准确地捕捉输入文本中的重要信息，从而提高分类的准确性。

多头注意力机制是自注意力机制的进一步改进。它不仅可以减少对外部信息的依赖，同时还能提高对数据内部相关性的捕捉能力。多头注意力机制还可以将输入数据分成多个

头，分别计算每个头所对应的注意力权重，并将多个头的输出结果进行拼接和处理，从而将多个头的互补信息整合在一起，以更好地捕捉文本的内部相关性和语义特征。本文通过使用多头注意力机制，增强了对文本情感极性相关特征的关注程度，从而在分类任务中取得了更高的准确性。

2 本文模型

本文提出的基于 RoBERTa-WWM 与多头注意力机制的多特征情感分析模型，主要分为输入层、特征提取层、多头注意力层、输出层四个部分。该模型采用 RoBERTa-WWM 模型动态生成词向量，先将向量输入到 TextCNN 中，经过不同的卷积核和最大池化提取局部特征；然后再将 RoBERTa-WWM 模型得到的特征表示输入 Multi-Layer BIGRU，它通过多个 BIGRU 堆叠的方式，提取更深层次全局特征；最后通过多头注意力机制将局部和全局特征进行融合赋予不同的权重，以突出关键信息和文本的情感极性。其模型结构图如图 3 所示。

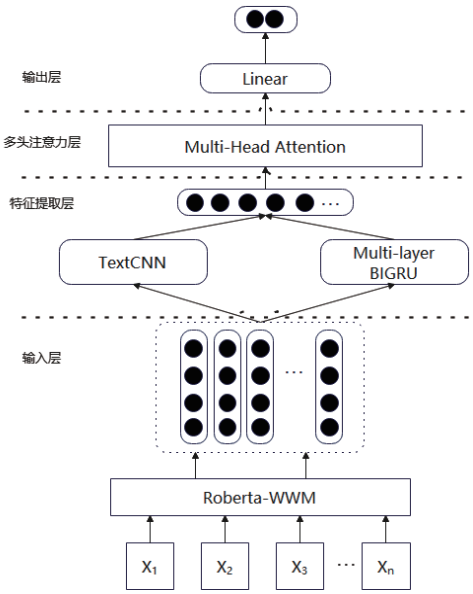

图 3 本文模型结构示意图

2.1 输入层

通过 RoBERTa-WWM 做词嵌入，对于一句话 $S=\{X_1, X_2, X_3, \cdots, X_n\}$，将其利用 Tokenizer 转换成文本向量输入到 RoBERTa-WWM 中，得到字向量集合 $\boldsymbol{T}=\{t_1, t_2, t_3, \cdots, t_n\}$，然后将向量集合进行拼接得到句子 S 向量矩阵 $\boldsymbol{R}_{n\times s}$，其拼接方式为：

$$R_{n\times s}=t_1 \oplus t_2 \oplus \cdots \oplus t_n \tag{5}$$

式中：$\oplus$ 为向量拼接运算符；n 表示每条语句的最大长度，本文设置的文本最大长度为 150，即 n=150；s 表示词向量的维度，RoBERTa-WWM 模型的词向量维度为 768，即 s 的值为 768。

2.2 特征提取层

在上一层获取到文本信息，分别输入到 TextCNN 和 Multi-Layer BIGRU 提取局部与全局特征，然后对这些特征进行融合得到最终特征。

提取局部特征时，本文使用了大小为 2、3、4、5 四种卷积核，每种卷积核的输出通道都为 128，保证卷积输出向量的维度一致。记不同卷积核输出的结果为 $C_i(i$=2,3,4,5)，通过最大池化的方法，将特征更加清晰化，并将不同卷积核池化后的特征进行拼接。具体公式为：

$$C_{out}=concat(\max\_polling(C_i)) \tag{6}$$

提取全局特征时，本文使用 Multi-layer Bigru 进行更深层次上下文信息。本文使用了两层的一个双向 GRU，用 H_i 表示得到的结果，其中 h_{i1} 表示第一层单向 GRU 的结果，i=1,2，H 表示最终结果。计算过程为：

$$\overrightarrow{h_{il}}=\overrightarrow{GRU}(R) \tag{7}$$

$$\overleftarrow{h_{il}}=\overleftarrow{GRU}(R) \tag{8}$$

$$H_1=[\overrightarrow{h_{il}},\overleftarrow{h_{il}}] \tag{9}$$

同理可得 H_2，最终这个通道所得特征为：

$$H=[H_1, H_2] \tag{10}$$

将局部和全局特征进行融合，可得：

$$M=[C_{out}, H] \tag{11}$$

2.3 多头注意力层

在注意力得分计算时，注意力机制的公式为：

$$Attention(Q,K,V)=softamx(\frac{QK^T}{\sqrt{d}})V \tag{12}$$

$$MultiHead(M)=Concat(head_1,\cdots,head_i)W^O \tag{13}$$

$\boldsymbol{M}$ 是输入矩阵。然后对每个头计算得分，再沿着头的维度进行拼接（即将每个头的输出沿着列的维度进行拼接），得到最终的多头注意力的输出。其中，head_i 表示第 i 个头的输出，$\boldsymbol{W}^O$ 是一个用于将拼接得到的输出向量进行线性变换的矩阵。

2.4 输出层

首先在通过全连接层前加入 dropout 层，可以有效缓解模型出现过拟合，然后输入到全连接层分类。本文在计算损失时，使用交叉熵损失函数，在此函数中已经对结果做了 softmax 处理，故本层的公式如下：

$$out=f(Wx+b) \tag{14}$$

式中：x 是上一层的输出特征，$\boldsymbol{W}$ 表示权重矩阵，b 表示偏置，out 为分类结果。

3 实验与结果分析

3.1 实验环境与数据集

本文采用的是情感分析领域两个公开的中文数据集，

Online_Shopping_10_Cat 和中国科学院计算所学者谭松波收集的酒店评论这两个数据集来进行实验。

数据集 Online_Shopping_10_Cat 包含 10 个类型（书籍、平板、手机、水果、洗发水、热水器、蒙牛、衣服、计算机、酒店）的线上交易，共 6 万多条评论数据，其中正、负向评论各约 3 万条。本文将这些数据打乱后按 8:1:1 的比例分成训练集、验证集和测试集。具体的信息如表 2 所示。

表 2 实验中使用的 Online_Shopping_10_Cat 数据集

	总数	正向	负向
Train	50 218	25 302	24 916
Dev	6277	3204	3073
Test	6278	3221	3057

谭松波收集的酒店评论共有 10 000 条评论数据，其中包括 3000 条负向情感倾向的评论，7000 条正向情感倾向的评论。本文用同样的方法将其按照 8:1:1 的比例分成训练集、验证集和测试集，具体的信息如表 3 所示。

表 3 实验中使用的酒店评论数据集

	总数	正向	负向
Train	8000	5603	2397
Dev	1000	697	303
Test	1000	700	300

3.2 实验参数设置

除去模型本身之外，超参数的设置也对模型的性能有着至关重要的作用。本文词嵌入层使用的是由哈工大讯飞联合实验室研制的 hfl/chinese-roberta-wwm-ext[17]，该模型采用 12 层 Transformer，隐藏层的维度为 768，激活函数为 GELU。特征提取层和特征融合层的参数设定如表 4 所示。

表 4 实验参数配置

Hyperparameter	Configuration
max_len	150
batch_size	6
num_heads	8
epochs	10
learn_rate	1e-5
dropout	0.5
filter_sizes	[2,3,4,5]
num_filters	128
Optimizer	AdamW

3.3 模型的评价指标

为验证模型的有效性，本文采用准确率（accuracy）和 F_1 值作为评价指标。准确率是模型预测正确的样本数与总样本数的比值，而 F_1 值则是一种综合衡量模型性能的指标。具体的计算公式为：

$$\text{Accuracy} = T_{true} / S_{sum} \tag{15}$$

$$F_1 = \frac{2 \times \text{pre} \times \text{rec}}{\text{pre} + \text{rec}} \tag{16}$$

式中：T_{true} 代表预测正确的样本数；S_{sum} 代表总的样本数；pre 表示精准率，即预测为正例的样本中实际也是正例的比例；rec 表示召回率，即预测的样本中有多少个正例被预测为正例的比例 [18]。

3.4 对比实验与结果分析

3.4.1 词向量模型对比实验

为了验证 RoBERTa-WWM 对文本的向量化表示具有更好的表征能力，本文将其与随机 Embedding 和 Word2Vector 做对比实验，在相同的实验环境下，使用谭松波收集的酒店评论数据集进行对比分析。实验结果如表 5 所示。

表 5 词向量模型对比实验结果

	accuracy/%	F_1 值 /%
Embedding	79.20	86.67
Word2Vec	84.20	88.84
BERT	92.20	94.43
RoBERTa	92.30	94.47

结果说明，随机 Embedding 和 Word2Vec 的表征能力都逊色于 RoBERTa-WWM，而 RoBERTa-WWM 词嵌入模型的 Accuracy 和 F_1 值分别达到了 92.30% 和 94.47%，都高于其他词嵌入模型，说明 RoBERTa-WWM 模型具有更出色的文本向量化表征能力。

3.4.2 模型对比实验分析

为验证本文模型的有效性，本组实验设计了对比实验和消融实验，在上述两个数据集下进行对比实验分析。以下给出这些模型的简介。

（1）BERT-CNN：使用 BERT 生成字向量，送入 TextCNN 网络提取特征，通过交叉熵损失函数计算损失，然后通过全连接层进行情感分类。

（2）ATT-BIGRU-CNN[19]：首先使用 BIGRU 模型提取文本的序列化特征信息；接着利用自注意力机制对特征进行初步筛选，将经过处理的特征信息传入具有不同卷积核的卷积神经网络中；然后使用自注意力机制对局部特征进行动态权重调整，以突出关键特征的提取；最后通过应用 Softmax 函数得到文本情感极性的预测结果。

（3）ATT-BiLSTM[20]：使用注意力机制对 BILSTM 网络的输出进行不同权重的区分。

（4）ROBERTA-CNN：消融实验，先使用 RoBERTA-WWM 获取字向量，再送入 TextCNN 进行特征提取，最后通过全连接层进行情感分类。

（5）ROBERTA-CNN-ATT：消融实验，先使用 RoBERTa-WWM 获取字向量，将其送入 TextCNN 进行特征提取；然后利用多头注意力机制进行动态权重调整；最后通过全连接层进行情感分类。

（6）ROBERTA-CNN-BIGRU：消融实验，先使用 RoBERTa-WWM 获取字向量，将其送入 TextCNN 进行局部特征提取；然后再利用多层 BIGRU 提取全局特征；最后将两种特征相结合，通过全连接层进行情感分类。

各模型的实验结果如表 6 和表 7 所示。

表 6 不同模型在 Online_Shopping_10_Cat 实验结果对比

	Acc/%	F_1 值 /%
Bert-cnn	92.59	92.77
Att-bigru-cnn	92.94	—
ATT-BiLSTM	92.38	—
Roberta-cnn	92.96	93.17
Roberta-cnn-att	93.26	93.32
Roberta-cnn-M-bigru	93.49	93.62
本文模型	94.14	94.21

表 7 不同模型在酒店数据集上的实验结果对比

	Acc/%	F_1 值 /%
Bert-cnn	92.20	94.31
Att-bigru-cnn	92.75	—
ATT-BiLSTM	88.22	—
Roberta-cnn	91.40	93.86
Roberta-cnn-att	93.20	95.05
Roberta-cnn-M-bigru	92.60	94.79
本文模型	94.20	95.85

通过表 6 和表 7 可知，本文提出的基于 RoBERTa-WWM 的多特征情感分析模型在 Online_Shopping_10_Cat 和酒店评论数据集中的效果都优于其他模型。在 Online_Shopping_10_Cat 数据集上的准确率和 F_1 值分别达到了 94.14% 和 94.21%，在酒店评论数据集上的准确率和 F_1 值分别达到了 94.20% 和 95.85%。实验结果证明，本文模型在中文语料的数据中具有普遍性的优异性能。

同时，为了验证模型的合理性，本文对其做了消融实验。实验结果表明，使用了 TextCNN 与 Multi-Layer BIGRU 对文本的特征表示更加全面，尤其是文本数据较长时，对结果的提升有着更加明显的效果。

在对本文模型除去多头注意力机制进行消融实验后发现，在加入多头注意力机制后，准确率和 F1 值在两个数据值上都有着不同程度的提升，说明本文所提出的模型用多头注意力机制对局部和全局特征进行权重的动态分配有着至关重要的作用。

3.4.3 回调参数分析

对模型训练过程中的一些参数进行分析，如图 4 所示，绘制了交叉熵损失函数值随着训练轮次逐步下降的过程，模型逐渐趋于收敛。

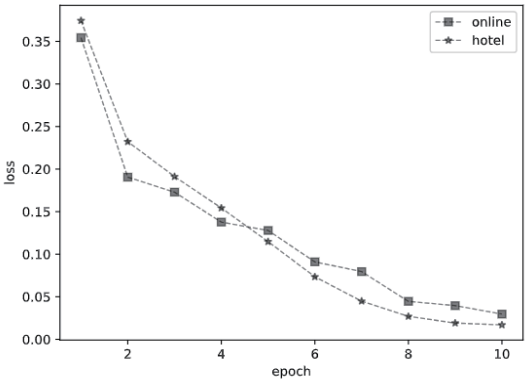

图 4 模型的训练损失函数变化

通过观察模型在 Online_Shopping_10_Cat 和酒店评论数据集上的训练表现发现，在第 6 轮次时，loss 已经下降到 0.10 以下，在第 8 轮次后 loss 下降缓慢，此时已经达到 0.050 以下，模型趋于收敛，对数据集的拟合效果出色。

通过绘制模型准确率随训练轮次 epoch 的变化图，可以更加直观地看到模型的拟合情况。如图 5 绘制了模型在 Online_Shopping_10_Cat 和酒店评论数据集中准确率随 epoch 变化的折线图。

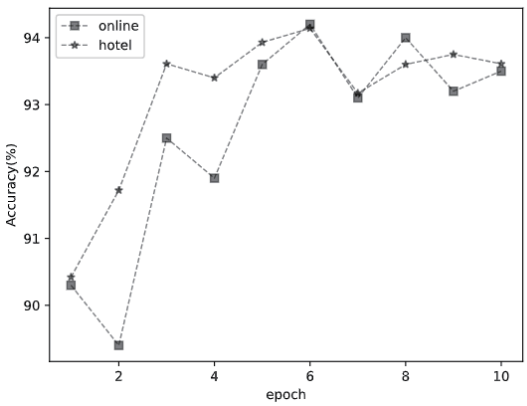

图 5 模型在数据集上的准确率变化

从图 5 可以看出，模型在前 5 轮中准确率稳步上升，在第 6 轮达到最高，在第 7 轮后模型的性能趋向稳定。

综上实验表明，本文模型在较少的训练轮次下就可以达到理想状态，收敛速度快且性能较优。

4 结语

本文提出的基于 RoBERTa-WWM 与多头注意力机制的多特征情感分析方法，用 RoBERTa-WWM 模型进行文本向量化表示，结合 TextCNN 与 Multi-Layer BIGRU 分别提取

文本中蕴含的局部与全局情感特征，将这些特征融合后，通过建立多头注意力权重分配机制，经过输出层计算后，得出情感极性，有效解决了多义词表征、文本特征提取不足和无法捕捉特征的关键信息等问题。本文模型在 Online_Shopping_10_Cat 数据集和酒店评论数据集上表现出了较优的性能，具有较大的实用价值。但是它仍然存在许多不足之处，由于其参数量较大、模型深度较深，会表现出模型训练时间过久，针对这些问题，后续工作需要加以改进。

参考文献：

[1]PANG B, LEE L.Opinion mining and sentiment analysis[J]. Foundations and trends® in information retrieval, 2008, 2(1/2): 1-135.

[2]XU G, YU Z, YAO H, et al. Chinese text sentiment analysis based on extended sentiment dictionary[J]. IEEE access, 2019, 7: 43749-43762.

[3]PANG B, LEE L, VAITHYANATHAN S. Thumbs up? Sentiment classification using machine learning techniques[EB/OL].(2002-05-28)[2024-02-10].https://arxiv.org/abs/cs/0205070.

[4]KIM Y. Convolutional neural networks for sentence classification[EB/OL].(2014-08-25)[2024-02-10].https://arxiv.org/abs/1408.5882.

[5]MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality[EB/OL].(2013-10-16)[2024-02-10].https://arxiv.org/abs/1310.4546.

[6]ZIA T, ZAHID U. Long short-term memory recurrent neural network architectures for Urdu acoustic modeling[J]. International journal of speech technology, 2019, 22:21-30.

[7] 阿音嘎 , 杨珍 , 房胜男 , 等 . 一种基于 GRU-Glove 算法的复杂文本分类方法 [J]. 警察技术 , 2022(5):49-53.

[8]XIAO Z, LIANG P.Chinese sentiment analysis using bidirectional LSTM with word embedding[C]//Cloud computing and security,part 2.Cham:Springer,2016: 601-610.

[9] 吴贵珍 , 王芳 , 黄树成 . 基于词向量与 CNN-BIGRU 的情感分析研究 [J]. 软件导刊 ,2022,21(8):27-32.

[10]MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[EB/OL].(2013-01-16)[2024-02-11].https://arxiv.org/abs/1301.3781.

[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).Stroudsburg, PA:Association for Computational Linguistics,2014:1532-1543.

[12]DEVLIN J, CHANG M W, LEE K, et al. Bert: pre-training of deep bidirectional transformers for language understanding[EB/OL].(2018-10-11)[2024-02-11].https://arxiv.org/abs/1810.04805.

[13]LIU Y, OTT M, GOYAL N, et al. Roberta: a robustly optimized bert pretraining approach[EB/OL].(2019-07-26)[2024-02-11].https://arxiv.org/abs/1907.11692.

[14]CUI Y, CHE W, LIU T, et al. Pre-training with whole word masking for chinese bert[C]// IEEE/ACM Transactions on Audio, Speech, and Language Processing. New York: ACM,2021:3504-3514.

[15]BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[EB/OL].(2014-09-01)[2024-02-11].https://arxiv.org/abs/1409.0473.

[16] 王锐 , 郑新章 , 宗国浩 , 等 . 融合 BiLSTM 和注意力机制的卷烟消费者评价情感分类方法 [J]. 烟草科技 , 2022, 55(11): 106-112.

[17]XU Z. RoBERTa-wwm-ext fine-tuning for Chinese text classification[EB/OL].(2021-02-04)[2024-02-13].https://arxiv.org/abs/2103.00492.

[18] 刘欢 , 窦全胜 . 嵌入不同邻域表征的方面级情感分析模型 [J]. 计算机应用 ,2023,43(1):37-44.

[19] 胡艳丽 , 童谭骞 , 张啸宇 , 等 . 融入自注意力机制的深度学习情感分析方法 [J]. 计算机科学 ,2022,49(1):252-258.

[20]WANG Y, HUANG M, ZHU X, 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.

【作者简介】

张宏伟（1999—），男，山西大同人，硕士研究生，研究方向：文本情感分析、深度学习。

李晓晔（1981—），女，辽宁凤城人，博士，副教授，CCF 会员（O2574M），研究方向：隐私保护、社会网络、机器学习。

（收稿日期：2024-03-01）