

# BERT-WWM 融合双通道语义特征的 中文文本情感分析

赵雪峰, 狄恒西, 柏长泽, 仲兆满

(江苏海洋大学计算机工程学院, 江苏 连云港 222005)

[摘要] 针对传统文本分类模型存在中文文本语义特征的提取能力有限等问题, 提出一种 BERT-WWM 融合双通道语义特征的中文文本情感分析模型, 学习更深层次的文本语义特征表示. 首先采用 BERT-WWM 获取中文文本的动态特征向量表示并传入双通道卷积神经网络-双向长短期记忆网络(CNN-BiSLTM)中进行特征提取, 然后将 BiLSTM 通道提取文本的上下文特征向量经注意力(Attention)层动态权重调整后, 与 CNN 通道提取的文本局部语义特征向量进行融合, 以增强模型的文本特征提取能力. 最后, 将融合特征经过全连接层和 Softmax 函数得出文本的情感倾向. 实验结果表明, 相比传统的单通道模型和多通道混合模型, 所提模型在准确率 Acc 和综合评价指标 F1 上分别提高了 0.91% 和 1.40%, 证明了该模型在中文文本情感分析任务中的有效性和可行性.

[关键词] 文本情感分析, BERT-WWM, 双通道语义特征, 卷积神经网络, 双向长短期神经网络

[中图分类号] TP391 [文献标志码] A [文章编号] 1672-1292(2025)04-0018-10

## Chinese Text Sentiment Analysis Based on BERT-WWM with Dual Channel Semantic Features

Zhao Xuefeng, Di Hengxi, Bai Changze, Zhong Zhaoman

(School of Computer Engineering, Jiangsu Ocean University, Lianyungang 222005, China)

**Abstract:** Aiming at the problem of the limited extraction ability of Chinese text semantic features in traditional text classification models, a BERT-WWM text emotion analysis model is proposed to learn the deeper representation of text semantic features. Firstly, BERT-WWM is used to obtain the dynamic feature vector representation of Chinese text and it is passed into the two-channel convolutional neural network(CNN-BiSLTM) for feature extraction. Then, the context feature vector of text extracted by BiLSTM channel is adjusted by the dynamic weight of the Attention layer. The model is fused with the local semantic feature vector extracted by CNN channel to enhance the capability of text feature extraction. Finally, the fusion features are passed through the full connection layer and Softmax function to obtain the emotional tendency of the text. The experimental results show that compared with the traditional single-channel model and multi-channel mixed model, the accuracy of Acc and the comprehensive evaluation index F1 are improved by 0.91% and 1.40%, respectively, which proves the effectiveness and feasibility of the model in Chinese text sentiment analysis.

**Key words:** text sentiment analysis, BERT-WWM, dual-channel semantic feature, convolutional neural network, bidirectional long and short term memory network

文本情感分析(text sentiment analysis), 又称意见挖掘(opinion mining), 旨在识别和理解文本中表达的情感倾向<sup>[1-2]</sup>, 并进一步理解情感背后的主题和观点. 随着大数据时代的到来, 情感分析在市场营销、舆情检测、舆论引导和教育等多个领域扮演了日益重要的角色, 国内外学者对此展开了广泛的工作. 文本情感分析逐渐成为研究的热点, 具有广阔的前景和重要社会意义.

早期的文本情感分析主要基于情感词典<sup>[3]</sup>和机器学习<sup>[4]</sup>的方法. 近年来, 深度学习的发展为文本情

收稿日期: 2024-09-07.

基金项目: 国家自然科学基金项目(72174079)、江苏省“青蓝工程”优秀教学团队项目(2022-29).

通讯作者: 赵雪峰, 博士, 副教授, 研究方向: 多模态情感分析、图像处理. E-mail: zhaoxf@jou.edu.cn

感分析<sup>[3]</sup>任务提供了新的手段,较好地弥补了早期方法的不足.卷积神经网络(convolutional neural network, CNN)<sup>[4-8]</sup>最早应用于情感分类任务;陈珂等<sup>[9]</sup>提出一种针对微博文本的多通道卷积神经网络 MCCCNN 情感分析模型,实现了更高的情感分类准确率;Wu 等<sup>[10]</sup>利用 CNN 模型自动挖掘微博文本数据进行情感分析.基于 CNN 的文本情感分析可以有效捕捉文本的局部特征,能更好地理解文本的语义,但忽略了上下文语义信息.循环神经网络(recurrent neural network, RNN)<sup>[11]</sup>的使用可以弥补这一缺陷;Wang 等<sup>[12]</sup>提出堆叠残差长短期神经网络(long short term memory networks, LSTM)模型,使得深层的网络更容易优化;Dashtipour 等<sup>[13]</sup>提出了深度学习分类器的新架构,利用双向长短期神经网络(bidirectional long and short term neural networks, BiLSTM)进行情感分析,取得了较高的准确率.基于 RNN 的文本情感分析在处理文本序列任务中具有一定优势,但针对文本的局部特征提取存在不足. CNN 和 RNN 组成的神经网络框架有着不错的效果;Behera 等<sup>[14]</sup>提出 CNN 和 LSTM 两种深度学习架构的混合方法,对不同领域发布的评论进行情感分类;徐绪堪等<sup>[15]</sup>结合多个 CNN 和 BiLSTM 叠加的混合模型在微信推文的情感分类中取得了较好的效果.结合 CNN 和 RNN 的文本情感分析模型能够捕获局部和全局信息,实现更全面的特征提取,但增加了模型复杂度,导致参数调优困难.融合注意力机制(attention mechanism)<sup>[16]</sup>的情感分析模型也得到诸多研究者的关注;Basiri 等<sup>[17]</sup>提出了一种基于注意力的双向 RNN-CNN 模型,在文本情感分析任务上取得不错的效果;Yan 等<sup>[18]</sup>提出了一种注意力并行双通道深度学习混合模型 ADDHM,在舆情评论的情感分类方面具有更好的分类效果.近些年, BERT<sup>[19]</sup>等词嵌入技术的应用也取得了不错的效果.基于注意力机制<sup>[20]</sup>、CNN、RNN 和 BERT 的文本情感分析模型能有效关注重点信息,具有较强的灵活性和可行性,但针对中文文本语义特征<sup>[21]</sup>仍存在提取不充分、不全面等问题.

基于上述问题,本文利用 CNN、BiLSTM 和注意力机制,并结合哈工大讯飞实验室发布的中文 BERT-WWM(BERT with whole word masking)<sup>[22]</sup>预训练模型,提出一种 BERT-WWM 融合双通道语义特征的模型,用以解决中文文本情感分析问题,充分利用 BERT-WWM、CNN、BiLSTM 模型和注意力机制的优势,高效准确地实现文本语义特征挖掘和情感分类.本文主要从以下 3 个方面进行了提升:

(1)在词嵌入层使用 BERT-WWM 代替传统 BERT 模型,利用全词 Mask 方法,使模型在处理中文文本时能更准确地捕捉语义特征信息,提升词级别任务的性能,减少歧义和多义性,在中文文本数据集上能实现更高的分类准确率;

(2)在特征提取层利用双向 LSTM(BiLSTM)代替单向 LSTM 和传统的 RNN 模型,能更好地捕捉长距离的上下文依赖关系,且面对较大的中文文本数据集和计算资源时能更准确地预测情感类别,达到更高的精确率;

(3)提出一种 BERT-WWM 融合双通道语义特征的网络模型,充分结合 CNN 和 BiLSTM 模型各自的优势,分别提取文本局部特征信息和上下文依赖关系,并引入注意力机制,进一步提高文本情感分析的准确率.

## 1 相关技术

### 1.1 BERT-WWM 模型

作为 BERT 的一种改进版本, BERT-WWM 模型能够处理结构复杂和语义丰富的中文文本语言,在预训练阶段引入整词词典,采用全词 Mask 方法,能够更好地保留文本词汇的完整性,提高自然语言理解能力.本文选择 BERT-WWM 作为词向量预训练模型,其网络结构如图 1 所示.

相比原始 BERT 的 Mask 方法,全词 Mask 方法能更准确地理解单词含义和上下文关系,提高模型泛化能力,其具体对比效果如表 1 所示.同时, BERT-WWM 在预训练过程中使用大规模的中文文本数据,能够提供更好的中文语义表示,适用于本文实验所应用的中文公开数据集.

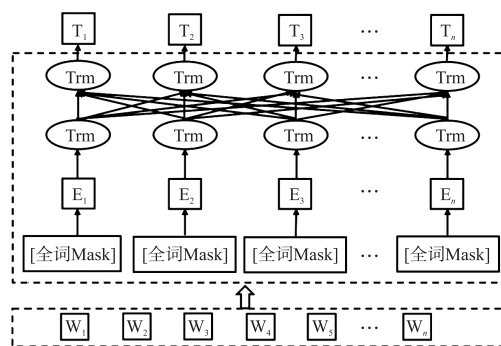


图 1 BERT-WWM 网络结构图

Fig. 1 BERT-WWM network structure

表 1 原始 Mask 和全词 Mask 的对比  
Table 1 Comparison of original Mask and full word Mask

原始句子	今天出去购物了感觉真高兴
需 Mask 的部分	今天出去购物了感觉真高兴
原始 Mask	今[Mask]出去[Mask]物了感觉真高[Mask]
全词 Mask	[Mask][Mask]出去[Mask][Mask]了感觉真[Mask][Mask]

### 1.2 卷积神经网络 CNN

卷积神经网络(CNN)具有局部连接和参数共享等优点,模型训练参数较少,适用于不同长度的文本,近年来被广泛应用于 NLP 任务. 采用 CNN 提取文本的局部情感特征,可以更有效地挖掘中文文本潜在的语义信息. CNN 主要由输入层、卷积层、池化层和全连接层组成.

在处理文本数据时,对于一个给定的句子,其输入层的矩阵表示如式(1)所示:

$$S \in \mathbf{R}^{n \times m}. \tag{1}$$

式中, $n$  为句子中的单词数; $m$  为词向量维度.

卷积层用于提取句子的语义特征,如式(2)所示,每个卷积核对应提取句子中的某一部分特征,并生成一个特征图. 通过卷积操作,卷积核在句子中滑动窗口,并从局部区域提取特征信息:

$$W \in \mathbf{R}^{h \times d}. \tag{2}$$

式中, $h$  表示卷积核的宽度; $d$  表示卷积核的长度.

池化层通常采用最大池化对特征降维后得到一维向量,该一维向量可视为卷积层对文本数据映射后得到的文本主要特征. 最后,全连接层将一维向量投影到类别空间中,得到分类结果. 局部特征  $f_i$  的计算过程如式(3)所示:

$$f_i = f(W_c \otimes S_{i:i+h-1} + b). \tag{3}$$

式中, $W_c$  表示权重矩阵; $S_{i:i+h-1}$  表示从  $i$  到  $i+h-1$  个词向量矩阵; $b$  表示偏置量; $f()$  为 ReLU 激活函数.

### 1.3 双向长短期神经网络 BiLSTM

随着文本序列间隔长度的增加,传统 RNN 在情感分类任务中易产生梯度离散问题,LSTM 的提出解决了这个问题. LSTM 通常由输入门、遗忘门和输出门组成,如图 2 所示. 通过这些门控机制,LSTM 能够有选择性地忽略或存储传入的信息.

LSTM 模型中各个门的计算公式如式(4)——(5)所示:

$$\begin{aligned} f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \\ i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \end{aligned} \tag{4}$$

$$\begin{aligned} O_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \\ C_t &= f_t \circ C_{t-1} + i_t \circ \tilde{C}_t, \\ h_t &= O_t \circ \tan(C_t), \end{aligned} \tag{5}$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c).$$

其中, $W_f$  表示遗忘门连接的权重矩阵; $b_f$  表示遗忘门的偏移值, $W_i$  表示输入门连接的权重矩阵; $b_i$  表示输入门的偏移值; $W_o$  表示输出门的权重矩阵; $b_o$  表示输出门的偏移值;“ $\circ$ ”表示逐元素相乘.

在情感分析任务中,往往需考虑上下文信息对文本整体情感极性的影响. 传统 LSTM 只针对单向的语义关系. 本文利用 BiLSTM 代替 LSTM,同时考虑文本的上下文信息. 如图 3 所示,BiLSTM 模型由正反两个 LSTM 构成,通过两个不同顺序来捕捉上下文信息,挖掘到文本中更全面的语义信息.

### 1.4 注意力机制

鉴于不同词语对句子的情感极性贡献度不同,本文引入注意力机制获取文本中贡献度较大的语义信息. 如图 4 所示,注意力机制通常由 Encoder(编码器)和 Decoder(解码器)2 个模块组成. 编码器使用整个原始句子生成语义向量,解码器利用该向量得到一定变换后的输出,能更好地把握整个句子的语义信息.

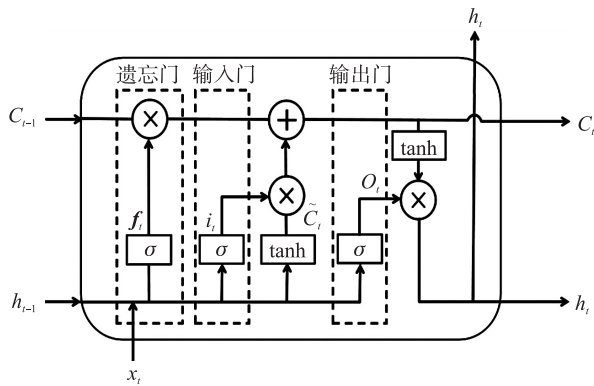


图 2 LSTM 网络结构图

Fig. 2 LSTM network structure diagram

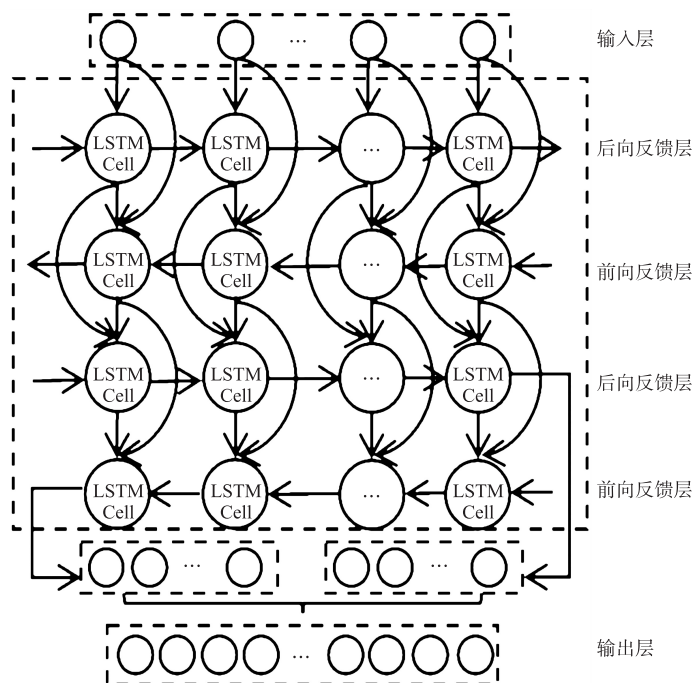


图3 BiLSTM 网络结构图

Fig. 3 BiLSTM network structure diagram

语义特征信息计算公式为:

$$F_c = \sum_{i=0}^L \alpha_i h_i,$$

$$\alpha_i = \frac{\exp(\mathbf{u}_i^T \mathbf{u}_w)}{\sum_{i=0}^L \exp(\mathbf{u}_i^T \mathbf{u}_w)}, \quad (6)$$

$$\mathbf{u}_i = \tanh(\mathbf{W}_i \mathbf{h}_i + b_i).$$

式中,  $\mathbf{u}_w$  为随机初始化的上下文向量, 在训练过程中更新;  $\mathbf{u}_i$  为隐藏层向量  $\mathbf{h}_i$  经过运算得到的结果;  $\mathbf{W}_i$  为权值矩阵;  $b_i$  为偏置项;  $\alpha_i$  为句子中第  $i$  个单词经过归一化处理得到的权重;  $F_c$  为最终经过注意力机制加权后总的语义特征信息.

## 2 模型设计

为了解决现有情感分析模型中文文本特征提取不充分等问题, 将 BERT-WWM、CNN、BiLSTM、注意力机制结合, 可以有效提高模型对中文文本多义词的识别能力, 更好地捕捉中文文本的局部特征和语境特征. 模型结构如图 5 所示, 由输入层、嵌入层、信息提取层、特征融合层和输出层组成.

输入层将中文文本传入到模型中. 鉴于中文文本结构的复杂性, 嵌入层使用 BERT-WWM 来获取文本数据的动态词向量, 利用全词 Mask 方法, 更好地保留中文词汇的完整性, 提高自然语言理解能力. 同时, 利用双向 Transformer 编码器的多头注意力机制<sup>[23]</sup> 可捕获到更准确的语义信息, 有效解决一词多义问题. 对于每个文本数据, 预处理后的单词序列集为  $\{a_1, a_2, \dots, a_n\}$ . 将集合输入到 BERT-WWM 获得向量矩阵, 如式(7)–(8)所示:

$$\mathbf{S}_o \in \mathbf{R}^{n \times e}, \quad (7)$$

$$\mathbf{S}_o = \mathbf{w}_1 \oplus \mathbf{w}_2 \oplus \dots \oplus \mathbf{w}_n. \quad (8)$$

式中,  $n$  表示文本中的单词数;  $\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n\}$  表示经 BERT-WWM 训练生成的词向量;  $e$  为向量维度;  $\oplus$  表示向量连接运算符.

其次, 将得出的向量矩阵  $\mathbf{S}_o$  嵌入到文本词向量  $\mathbf{X}_k$  中, 如式(9)所示:

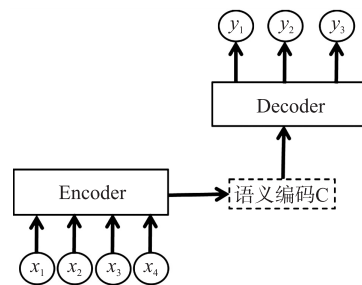


图4 注意力机制网络结构示例图

Fig. 4 Attention mechanism network structure diagram

$$\mathbf{X}_k = \mathbf{S}_o \boldsymbol{\omega}, \quad k \in [1, l]. \quad (9)$$

式中,  $k$  表示数据文本的长度;  $l$  表示文本的填充长度或最大字数;  $\boldsymbol{\omega}$  为权重值。

然后, 将经过处理的文本向量  $\mathbf{X}_k$  传入信息提取层提取特征。考虑到文本数据的多样性及 CNN、BiLSTM 特征提取的优势, 信息提取层由 CNN 通道和 BiLSTM 双通道组成。

CNN 通道用于获取文本的局部特征语义, BiLSTM 用于提取文本的全局语义特征, 并经 Attention 层进行权重调整。在文本情感分析任务中, CNN 主要通过卷积层和池化层来学习和提取重要的特征信息, 且无需进行大量的文本预处理工作。因此, 本文采用 CNN 提取文本情感特征。在 CNN 通道中, 首先执行卷积操作, 词向量  $\mathbf{X}_k$  通过滤波器进行处理。经过卷积运算后, 利用最大池化对特征进行压缩, 并将特征向量输入到全连接层, 得到 CNN 通道的输出, 如式 (10) — (12) 所示:

$$\mathbf{S}_x = \mathbf{X}_1 \oplus \mathbf{X}_2 \oplus \cdots \oplus \mathbf{X}_l, \quad (10)$$

$$\mathbf{f}_i = \mathbf{f}(\mathbf{W} \otimes \mathbf{S}_{x,i:i+m-1} + \mathbf{b}_1), \quad (11)$$

$$\mathbf{F}_{\text{CNN}} = (\mathbf{f}_0, \mathbf{f}_1, \cdots, \mathbf{f}_{l-m}). \quad (12)$$

式中,  $\mathbf{S}_x$  由文本词向量  $\mathbf{X}_k$  组成,  $\mathbf{S}_{x,i:i+m-1}$  表示从  $i$  到  $i+m-1$  个词向量矩阵  $\mathbf{S}_x, i=0, 1, 2, \cdots, l-m$ ;  $\mathbf{F}_{\text{CNN}}$  为词向量经过处理得到的文本局部特征。

由于中文词语间存在关联性, 需要考虑到文本上下文信息对整句情感极性的影响, 而 BiLSTM 由正反两个 LSTM 网络叠加构成, 能较好地学习到文本上下文依赖关系, 并有效缓解梯度消失问题, 从而更好地预测情感极性。在 BiLSTM 通道中, 需对任意长度的序列  $\mathbf{X}_k$  进行处理, 利用 BiLSTM 提取并存储前向和后向的长距离依赖关系, 如式 (13) — (14) 所示:

$$\vec{h}_k = \overrightarrow{\text{LSTM}}(\mathbf{X}_k), \quad k \in [1, l], \quad (13)$$

$$\overleftarrow{h}_k = \overleftarrow{\text{LSTM}}(\mathbf{X}_k), \quad k \in [1, l]. \quad (14)$$

对于每个词语, 将前向输出和后向输出相连接, 生成 BiLSTM 在  $k$  时刻的输出, 如式 (15) 所示:

$$\mathbf{h}_k = \mathbf{w} \vec{h}_k + \mathbf{v} \overleftarrow{h}_k + \mathbf{b}_2, \quad (15)$$

式中,  $\mathbf{w}$  表示正向输出的权重矩阵;  $\mathbf{v}$  表示反向输出的权重矩阵;  $\mathbf{b}_2$  表示偏置参数。

为减少无关噪声字词的干扰, 同时强调关键信息在文本中的作用, 提高模型的鲁棒性, 在 BiLSTM 通道加入 Attention 机制, 根据每个词在文本中的贡献程度分配权重, 并改进特征向量, 最终得到 BiLSTM 输出向量的加权和, 即文本的上下文语境特征, 如式 (16) — (18) 所示:

$$\mathbf{u}_k = \tanh(\mathbf{W}_o \mathbf{h}_k + \mathbf{b}_o), \quad (16)$$

$$\alpha_k = \frac{\exp(\mathbf{u}_k^T \mathbf{u}_o)}{\sum_k \exp(\mathbf{u}_k^T \mathbf{u}_o)}, \quad (17)$$

$$\mathbf{F}_{\text{BiLSTM}} = \sum_k \alpha_k \mathbf{h}_k. \quad (18)$$

其中,  $\mathbf{u}_k$  为隐藏层向量  $\mathbf{h}_k$  经过运算得到的结果;  $\mathbf{u}_o$  为随机初始化的上下文向量, 在训练过程中不断更新。利用  $\mathbf{h}_k$  和  $\mathbf{u}_o$  之间的相似性来计算  $\mathbf{h}_k$  的重要性, 并将其标准化,  $\alpha_k$  为归一化处理后的权重,  $\mathbf{F}_{\text{BiLSTM}}$  为最终经过注意力机制加权后的语义特征。

将两个通道获取的局部情感特征  $\mathbf{F}_{\text{CNN}}$  和全局情感特征  $\mathbf{F}_{\text{BiLSTM}}$  输入到特征融合层进行拼接融合, 得到

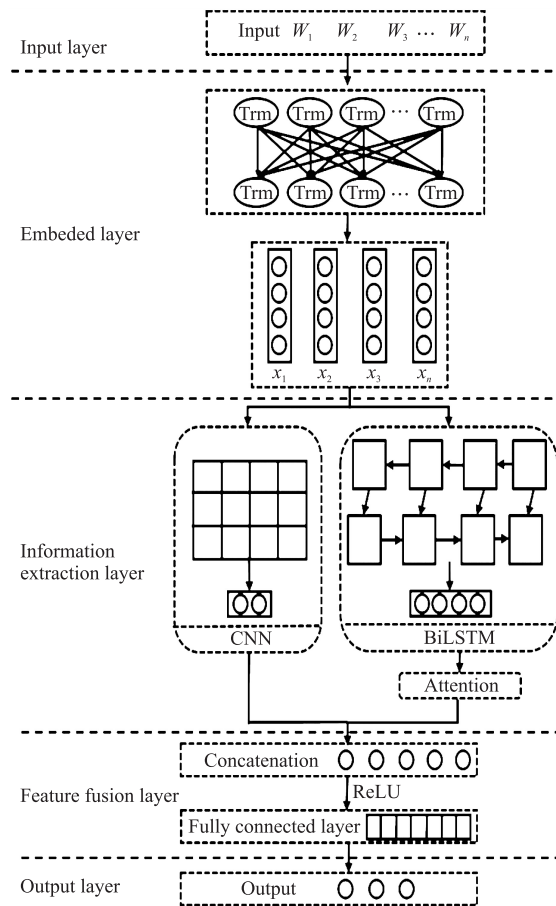


图 5 模型整体结构

Fig. 5 Overall structure of model

最终的情感特征向量  $F_s$  如式(19)所示:

$$F_s = F_{\text{CNN}} + F_{\text{BiLSTM}}. \quad (19)$$

最后,将  $F_s$  通过全连接层输入 Softmax 函数中得到最终分类预测概率,如式(20)所示:

$$P = \text{Softmax}(W_s F_s + b_s). \quad (20)$$

式中,  $W_s$  为全连接层的权值矩阵,  $b_s$  为偏置项.

本模型使用的 Loss 函数是交叉熵损失 (cross-entropy Loss), 用于衡量模型预测的情感类别与真实标签之间的差异, 如式(21)所示:

$$\text{Loss} = -\frac{1}{N} \sum_{i=1}^N (y_i \log(\hat{y}_i) + (1-y_i) \log(1-\hat{y}_i)). \quad (21)$$

式中,  $y_i$  表示第  $i$  个样本的真实标签;  $\hat{y}_i$  表示模型对该样本的预测概率;  $N$  为样本总数.

### 3 实验分析

#### 3.1 实验环境

本文模型代码基于 PyTorch 深度学习框架编写, 实验在服务器端完成, 具体环境为: PyTorch 2.0.0, Python 3.8 (ubuntu20.04), Cuda 11.8, GPU RTX 3080 Ti (12GB) \* 1, CPU 12 vCPU Intel(R) Xeon(R) Silver 4214R CPU@2.40GHz.

#### 3.2 实验数据集

本文实验使用谭松波酒店评论、waimai\_10k、online\_shopping\_10\_cats3 个公开中文数据集, 数据集中的每条评论都已设置情感标签 (“1” 表示积极, “0” 表示消极), 具体信息如表 2 所示.

表 2 数据集具体信息

Table 2 Dataset specific information

数据集	训练集		测试集	
	积极/条	消极/条	积极/条	消极/条
谭松波酒店评论	1 800	1 800	1 200	1 200
waimai_10k	2 400	2 400	1 600	1 600
online_shopping_10_cats	18 997	18 666	12 730	12 380

#### 3.3 实验数据处理与模型超参数设置

为提高中文文本情感分析的准确率, 需对文本数据进行预处理. 首先过滤掉所有的标点符号和特殊字符, 使用正则表达式匹配中文字符, 并去除空格. 然后通过 jieba 分词工具进行文本词语的分割, 记录每个句子的长度. 本文实验采用 SGD (stochastic gradient descent)<sup>[24]</sup> 优化算法更新模型参数, 以最小化损失函数, 加快网络收敛速度. 本文模型超参数具体设置如表 3 所示.

表 3 模型超参数设置

Table 3 Model hyperparameter settings

参数	值	参数	值
卷积核大小	400×4	训练周期数	50
BiLSTM 网络层数	4	批处理大小	16
词向量维度	400	文本数据最大长度	400
BERT-WWM 隐藏层大小	768	是否进行填补操作	True
学习率	0.008	随机失活率	0.5

#### 3.4 评价标准

本文采用准确率 (Acc)、F1 值作为评价标准, 如式(22)–(25)所示:

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \quad (22)$$

$$P = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (23)$$

$$R = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (24)$$

$$F1 = \frac{2 \cdot P \cdot R}{P + R} \tag{25}$$

式中, TP 为预测为正的正样本; TN 为预测为负的负样本; FP 为预测为正的负样本; FN 为预测为负的正样本.

### 3.5 单通道模型实验性能对比

为验证本文模型在情感分析任务方面的可行性,在相同的实验环境下使用 3 个中文数据集作为实验数据,并分别与以下 6 种单通道模型方法对比:

(1) Word2Vec-BiLSTM<sup>[25]</sup>: 采用 Word2Vec<sup>[26]</sup> 获取文本词向量表示输入到 BiLSTM 中提取特征,并由 Softmax 进行分类;

(2) Word2Vec-CNN<sup>[25]</sup>: 将 Word2Vec 作为获取词向量的工具,将得到的词向量输入到 CNN 中进行训练和提取特征,最终通过 Softmax 进行分类;

(3) BERT-BiLSTM<sup>[18]</sup>: 通过 BERT 获取文本的词向量表示,输入单一的 BiLSTM 网络中进行训练,并不断微调,利用 Softmax 分类器输出分类结果;

(4) BERT-CNN<sup>[18]</sup>: 通过 BERT 对文本进行编码,并输入 CNN 中进行特征提取,利用 Softmax 分类器输出最终的情感极性;

(5) BERT-WWM-BiLSTM: 结合 BERT-WWM 得到丰富语境信息的动态词向量,输入 BiLSTM 网络中训练和提取特征,运用 Softmax 函数得到分类结果;

(6) BERT-WWM-CNN: 使用 BERT-WWM 得到词向量表示,使用 CNN 进行特征提取,最后通过 Softmax 进行分类.

单通道模型的实验性能对比如表 4 所示. 从表 4 可知,本文模型在中文情感分类任务上相较其他单通道模型,在各项评价指标方面均达到最优.

表 4 单通道模型实验性能对比

Table 4 Comparison of experimental performance of single channel model

模型	谭松波酒店评论		waimai_10k		online_shopping_10_cats	
	Acc	F1	Acc	F1	Acc	F1
Word2Vec-BiLSTM	85.10	85.59	87.12	87.40	89.30	89.90
Word2Vec-CNN	87.14	87.71	88.10	88.61	90.53	90.84
BERT-BiLSTM	88.75	88.54	88.89	89.01	90.98	91.36
BERT-CNN	89.23	89.44	89.67	89.97	91.53	92.33
BERT-WWM-BiLSTM	90.50	90.15	90.85	90.65	92.05	92.73
BERT-WWM-CNN	91.64	91.54	92.01	92.10	93.18	93.38
Ours	<b>92.78</b>	<b>93.02</b>	<b>93.98</b>	<b>94.09</b>	<b>94.44</b>	<b>95.73</b>

本文模型与各通道模型的 F1 值柱状图和 Loss 值折线图如图 6 和图 7 所示. 由图 6 可见,本文模型在 3 个数据集上的 F1 值相比其他模型最佳结果分别提升 1%~3%,在 online\_shopping\_10\_cats 数据集上最高

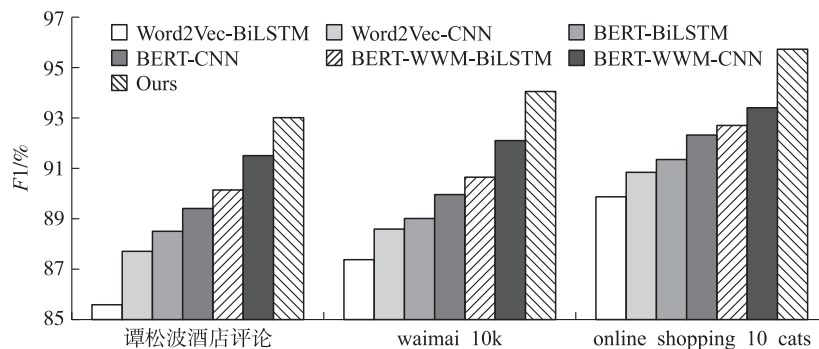


图 6 F1 值柱状图分析

Fig. 6 F1 value histogram analysis

达到 95.73%, 验证了本文模型在文本情感分析任务中的可靠性. 同时, 基于 BERT-WWM 的 CNN 与 BiLSTM 模型相比基于 Word2Vec 和传统 BERT 模型在处理中文数据集上结果均有不错的提升, 证明了 BERT-WWM 模型在处理中文文本数据时的有效性和可行性. 单一 CNN 和 BiLSTM 模型的使用往往会忽略上下文特征和局部特征的重要性. 本文模型通过综合利用 BERT-WWM、BiLSTM 和 CNN 等技术, 实现对文本局部特征和上下文依赖关系的充分考虑, 提高了模型的语义特征提取能力和情感分类效果. 由图 7 可见, 在训练过程中本文模型得到的最终 Loss 收敛值非常小, 且明显低于其他 6 种单通道模型. 虽然在 waimai\_10k 数据集上, BERT-CNN 单通道模型最终 Loss 收敛值优于本文模型, 但并不影响本文模型良好的综合性能效果.

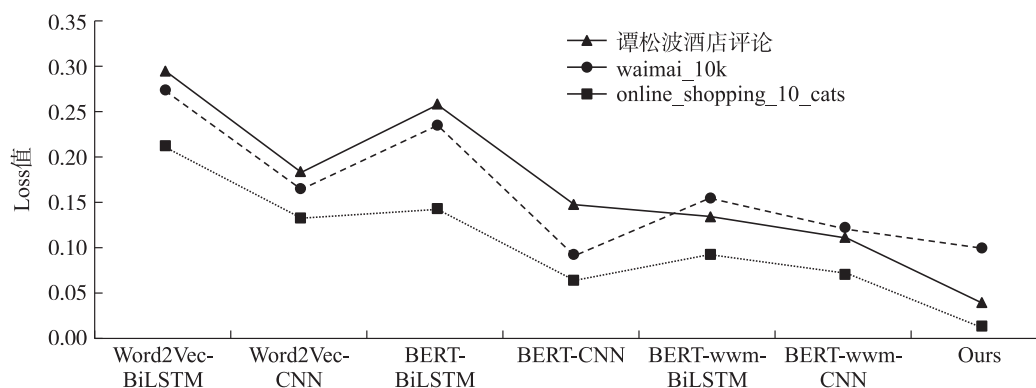


图 7 Loss 值柱状图分析

Fig. 7 Histogram analysis of Loss value

### 3.6 多通道混合模型实验性能对比

在相同的实验环境下, 使用 3 个中文数据集作为实验数据. 为了验证本文设计的双通道模型在文本情感分析任务中的可行性, 分别与 7 个多通道混合模型进行实验对比:

(1) Word2Vec-CNNs-BiLSTM 模型<sup>[27]</sup>: 首先利用传统 Word2Vec 获取文本词向量表示, 输入到多通道 CNNs-BiLSTM 抽取多尺度文本特征, 并结合注意力机制调整特征权重, 最后进行文本情感倾向判断;

(2) BERT-CNN-BiGRU 模型: 利用 BERT 模型获取文本的向量表示, 分别输入 CNN 和 BiGRU 进行特征提取, 利用注意力机制对 BiGRU 输出的全局语义特征进行权重调整, 最后进行特征融合并通过 Softmax 函数得到情感分类极性;

(3) MC-CNN-LSTM 模型<sup>[28]</sup>: 通过多通道 CNN 抽取不同的  $n$ -gram 特征, 输入到 LSTM 中进行处理, 最后通过 Softmax 函数得到分类结果;

(4) MC-CNN-BiGRU 模型<sup>[29]</sup>: 由 3 个 CNN 通道和 BiGRU 通道拼接而成, 在 3 个 CNN 通道都加入注意力机制, 而 BiGRU 通道不加入注意力机制;

(5) MC-AttCNN-AttBiGRU 模型<sup>[29]</sup>: 经过词嵌入得到的词向量分别输入到由 3 个 CNN 通道和 BiGRU 通道组成的结构中进行提取特征, 引入注意力机制, 最后将特征进行融合得到分类结果;

(6) BERT-CNN-BiGRU-attention 模型<sup>[30]</sup>: 基于 BERT 模型的词嵌入生成, 利用改进的 CNN 模型和 BiGRU 构建双通道模型, 并对输出特征赋予相应的注意力权重分数;

(7) BERT-CNN-BiLSTM-Att 模型<sup>[31]</sup>: 基于 BERT 模型生成词向量, 分别输入 CNN 和 BiLSTM 进行特征提取, 最后通过注意力机制实现情感分类.

多通道混合模型实验性能比较如表 5 所示. 由表 5 可知, 本文模型在准确率 Acc 方面均优于其他 7 种多通道混合模型, 在 3 个不同数据集上  $F1$  值分别达到 93.02%、93.98% 和 95.73%, 相比对照模型最佳结果提升近 1%, 证明了本文模型在情感分类任务中的可行性. 其他模型采用多通道的特征提取方式, 结合 CNN 和 RNN 网络的各自优点进行情感分类任务, 但在处理大规模中文文本数据时, 仍存在特征提取能力不足的问题, 同时在对多通道提取的语义特征进行融合时, 易产生部分特征损失且导致向量复杂度过高, 影响模型的性能. 本文模型较好地结合 CNN 和 BiLSTM 网络各自的优点, 巧妙利用双通道的方式, 通过 BERT-WWM 模型对中文文本进行动态词向量表示, 既能考虑到文本的局部特征对于模型性能的重要性, 也能很好地抓住

文本上下文特征的影响,并引入注意力机制调整权重,采用拼接融合的方式将双通道的特征进行合并,最大程度上减少了特征损失和向量复杂度,提升模型的特征提取能力,取得了良好的分类结果.

表 5 多通道混合模型实验性能对比

**Table 5 Experimental performance comparison of multi-channel hybrid model** %

模型	谭松波酒店评论		waimai_10k		online_shopping_10_cats	
	Acc	F1	Acc	F1	Acc	F1
Word2Vec-CNNs-BiLSTM	89.56	89.42	90.12	90.04	91.23	91.63
BERT-CNN-BiGRU	92.06	92.71	93.02	93.14	93.21	93.05
MC-CNN-LSTM	90.78	90.94	91.67	92.00	92.25	92.38
MC-CNN-BiGRU	91.01	91.16	92.00	92.14	93.42	93.37
MC-AttCNN-AttBiGRU	92.28	92.33	93.54	93.01	93.95	93.89
BERT-CNN-BiGRU-attention	92.15	91.98	93.67	93.23	93.72	93.56
BERT-CNN-BiLSTM-Att	92.54	92.41	93.38	93.17	94.15	94.62
Ours	<b>92.78</b>	<b>93.02</b>	<b>93.98</b>	<b>94.09</b>	<b>94.44</b>	<b>95.73</b>

### 3.7 消融实验

本文通过消融实验进一步探索本文所提模型各部分设计的合理性及其优点. 将本文提出的模型进行分解,分别移除 BERT-WWM 模型、CNN 网络、BiLSTM 网络和 Attention 机制,以此验证被移除部分对于总体模型的有效增益. 同时为控制参数对实验的影响,每组消融实验超参数设置相同. 具体实验结果如表 6 所示,其中 w/o 代表 without.

表 6 消融实验结果

**Table 6 Ablation results** %

模型	谭松波酒店评论		waimai_10k		online_shopping_10_cats	
	Acc	F1	Acc	F1	Acc	F1
w/o BERT-WWM	91.83	92.24	92.64	92.92	93.02	93.29
w/o CNN	89.50	90.15	90.25	90.65	89.87	89.93
w/o BiLSTM	90.64	91.04	91.27	91.46	91.77	92.19
w/o Attention	92.35	92.67	93.45	93.92	93.59	94.05
Ours	<b>92.78</b>	<b>93.02</b>	<b>93.98</b>	<b>94.09</b>	<b>94.44</b>	<b>95.73</b>

由表 6 可知,本文模型的每一部分对整体性能均有明显提升. 在相同实验条件下,使用 BERT-WWM 训练出的词向量相比采用传统 BERT 模型在处理中文文本数据集上各项性能均有明显的提升,可见 BERT-WWM 模型对文本词向量训练具有良好的表现. 当模型移除 Attention 机制时,仅通过 CNN 和 BiLSTM 提取特征也能达到不错的效果,但 Attention 的加入能在最终结果中放大关键特征信息的重要性,实现文本特征的进一步深入挖掘. CNN 和 BiLSTM 的结合能很好地提取局部语义特征和上下文信息,有利于总体模型性能的提升. 当单独使用 CNN 或 BiLSTM 网络时,部分文本特征的被忽略导致 Acc 和 F1 值呈现显著降低,整体性能略显不足,这也表明增加 CNN 和 BiLSTM 结构对本文模型有积极作用.

## 4 结论

本文针对中文文本情感分类问题,结合 BERT-WWM、CNN、BiLSTM 网络和 Attention 机制的各自优势,既利用 CNN 有效提取到文本的局部特征,又通过 BiLSTM 捕获句子的全局语义特征,并采用 Attention 机制进行语义特征权重调整. 在 3 个公开中文数据集上取得的实验结果充分证明了本文模型的可行性,同时融合双通道语义特征的模型相比单通道 CNN 和 BiLSTM 网络等提取的特征效果更好,较其他几种多通道混合模型在具体实验中也取得了更好的分类效果.

虽然本文模型能够提高中文情感分析任务的性能,但也仅考虑了积极和消极二分类情感,对于文本中更多元的情绪识别仍有待进一步提高. 因此下一步工作将考虑把深度学习方法与外部知识结构结合起来提

高模型对于更多元情绪识别的准确率。

### [参考文献] (References)

- [1] WANKHADE M, RAO A C S, KULKARNI C. A survey on sentiment analysis methods, applications, and challenges[J]. *Artificial Intelligence Review*, 2022, 55(7): 5731–5780.
- [2] 尹帮治, 徐健, 唐超尘. 基于图神经网络与表示学习的文本情感分析[J]. *南京师大学报(自然科学版)*, 2024, 47(3): 97–103.
- [3] 任菊香, 刘忠宝. 融合多粒度语义特征的中文情感分析方法[J]. *华东师范大学学报(自然科学版)*, 2023(6): 95–107.
- [4] WANG Q Y, ZHU G L, ZHANG S X, et al. Extending emotional lexicon for improving the classification accuracy of Chinese film reviews[J]. *Connection Science*, 2021, 33(2): 153–172.
- [5] MALVIYA S, TIWARI A K, SRIVASTAVA R, et al. Machine learning techniques for sentiment analysis: A review[J]. *SAMRIDDHI: A Journal of Physical Sciences, Engineering and Technology*, 2020, 12(2): 72–78.
- [6] KIM H, JEONG Y S. Sentiment classification using convolutional neural networks[J]. *Applied Sciences*, 2019, 9(11): 2347.
- [7] 杨璐, 何明祥. 基于门控机制和卷积神经网络的中文文本情感分析模型[J]. *计算机应用*, 2021, 41(10): 2842–2848.
- [8] MADAN A, KUMAR D. CNN-based models for emotion and sentiment analysis using speech data[J]. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 2024, 23(10): 142.
- [9] 陈珂, 梁斌, 柯文德, 等. 基于多通道卷积神经网络的中文微博情感分析[J]. *计算机研究与发展*, 2018, 55(5): 945–957.
- [10] WU P, LI X T, SHEN S, et al. Social media opinion summarization using emotion cognition and convolutional neural networks[J]. *International Journal of Information Management*, 2020, 51: 101978.
- [11] SHARFUDDIN A A, TIHAMI M N, ISLAM M S. A deep recurrent neural network with BiLSTM model for sentiment classification [C]//*Proceedings of the 2018 International Conference on Bangla Speech and Language Processing (ICBSLP)*. Sylhet, Bangladesh: IEEE, 2018.
- [12] WANG J, PENG B, ZHANG X J. Using a stacked residual LSTM model for sentiment intensity prediction[J]. *Neurocomputing*, 2018, 322: 93–101.
- [13] DASHTIPOUR K, GOGATE M, ADEEL A, et al. Sentiment analysis of persian movie reviews using deep learning[J]. *Entropy*, 2021, 23(5): 596.
- [14] BEHERA R K, JENA M, RATH S K, et al. Co-LSTM: Convolutional LSTM model for sentiment analysis in social big data [J]. *Information Processing & Management*, 2021, 58(1): 102435.
- [15] 徐绪堪, 周泽聿. 基于多尺度 BiLSTM-CNN 的微信推文的情感分类模型及应用研究[J]. *情报科学*, 2021, 39(5): 130–137.
- [16] NIU Z Y, ZHONG G Q, YU H. A review on the attention mechanism of deep learning[J]. *Neurocomputing*, 2021, 452: 48–62.
- [17] 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.
- [18] YAN C, LIU J H, LIU W, et al. Research on public opinion sentiment classification based on attention parallel dual-channel deep learning hybrid model[J]. *Engineering Applications of Artificial Intelligence*, 2022, 116: 105448.
- [19] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[EB/OL]. (2018–10–11) [2024–09–07]. <https://doi.org/10.48550/arXiv:1810.04805>.
- [20] 陈丽安, 过弋. 融合个体偏差信息的文本情感分析模型[J]. *计算机应用*, 2024, 44(1): 145–151.
- [21] 赵文辉, 吴晓鸽, 凌捷, 等. 基于 prompt tuning 的中文文本多领域情感分析研究[J]. *计算机工程与科学*, 2024, 46(1): 179–190.
- [22] LIU C, ZHU C, ZHU W J. Chinese named entity recognition based on BERT with whole word masking[C]//*Proceedings of the 2020 6th International Conference on Computing and Artificial Intelligence*. Tianjin, China: ACM, 2020.
- [23] WANG X D, TANG M W, YANG T, et al. A novel network with multiple attention mechanisms for aspect-level sentiment analysis[J]. *Knowledge-Based Systems*, 2021, 227: 107196.
- [24] GOWER R M, LOIZOU N, QIAN X, et al. SGD: General analysis and improved rates [C]//*Proceedings of the 36th International Conference on Machine Learning*. Long Beach, USA: ICML, 2019.
- [25] 赵宏, 王乐, 王伟杰. 基于 BiLSTM-CNN 串行混合模型的文本情感分析[J]. *计算机应用*, 2020, 40(1): 16–22.

(下转第 48 页)