

# 基于多头注意力机制的BiGRU-CNN文本情感分析

郝星跃

东北大学理学院, 辽宁 沈阳

收稿日期: 2021年12月22日; 录用日期: 2022年1月19日; 发布日期: 2022年1月26日

## 摘要

文本情感分析(Sentiment Analysis)当今已是自然语言处理(NLP)的热门研究领域和话题, 该文提出一种基于多头注意力机制(Multi-head-Attention Mechanism)的双向门控循环网络(BiGRU)和卷积神经网络(CNN)的模型(Multi-head-Attention based on BiGRU-CNN model)。首先将文本用词向量表示, 然后依次输入到BiGRU-CNN网络模型中得到本文的BiGRU全局特征和CNN局部特征; 随后将BiGRU-CNN网络得到的特征输入到多头注意力机制层进行特征权重分配以获取文本有区分度且重要的特征信息; 最后利用输出层中的Softmax函数对汽车评论文本进行情感极性分类。在汽车评论数据集中进行实验, 准确率为90.76%, F1值为90.25%, 实验结果与现有模型相比均有所提高。

## 关键词

情感分析, 多头注意力机制, 卷积神经网络

## Text Sentiment Analysis Based on Multi-Head Attention Mechanism with BiGRU-CNN

Xingyue Hao

School of Science, Northeastern University, Shenyang Liaoning

Received: Dec. 22<sup>nd</sup>, 2021; accepted: Jan. 19<sup>th</sup>, 2022; published: Jan. 26<sup>th</sup>, 2022

## Abstract

Text sentiment analysis is now a hot research field and topic in natural language processing (NLP). This paper proposes a new sentiment analysis model, which is based on the multi-head attention mechanism with Bidirectional Gated Recurrent Unit (BiGRU) and Convolutional Neural Network

(CNN). First, the text review is represented by a word vector matrix, and then the BiGRU-CNN model is used to obtain the features input, and introduce the multi-head attention mechanism for feature weight distribution to obtain distinguished and important feature information of the text; finally classify the final sentiment polarity of the text by Softmax function. The experiment was performed on the car review data set, and the experiment achieved the accuracy rate of 90.76% and the F1 value of 90.25%. The experimental results were improved compared with the existing models.

## Keywords

Sentiment Analysis, Multi-Head Attention Mechanism, Convolutional Neural Network

Copyright © 2022 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

随着互联网时代的到来,一种新型生活消费方式出现,即越来越多的消费者选择在传统互联网电商平台,如淘宝、京东、拼多多等和新兴短视频电商平台,如抖音小店、快手小店购买商品,节省了大量的时间和精力。与此同时,消费者在网上进行评论,相关产品的评价以文本的形式指数式增长呈现在大众面前。文本情感分析是自然语言处理(NLP)的一个重要研究领域,情感分析的任务是对文本进行情感分类和判断。利用文本情感分析,一方面能够为消费者提供准确的信息,使其能够挑选和购买到满意心仪的产品;另一方面可以帮助商家通过分析评论数据进行改进以更好地满足消费者需求,实现盈利。

随着文本级情感分析研究领域的出现,文本情感分析最初使用的方法为基于情感词典的方法,但是由于基于情感词典的方法在实践中出现了一些难以解决的问题,人们开始探索新的方法,基于机器学习的方法便在此时出现且被广泛使用和研究了很长一段时间,截至目前,基于深度学习方法已被广泛应用且处于持续改进和成熟的阶段。

基于情感词典的方法,前期需要相关领域专业学者进行情感词典的构建[1],其优势在于实现简单且不需要前期进行数据集情感极性的标注;但是随着互联网快速发展,新词层出不穷,情感词典的构建困难、复杂且需要与时俱进和大量人为干预[2],故仅依靠情感词典进行文本情感分析的方法逐步被淘汰,新的方法即基于机器学习进行文本情感极性分类的方法出现。

基于机器学习的方法进行文本情感分析任务前期需要大量的人工准备工作,如人工对文本评论数据进行特征工程(Feature Engineering)以提取特征,人工对数据集标签进行情感极性的标注等。只有在上述人工准备工作完善且精确的情况下,使用机器学习相关分类方法如:支持向量机、朴素贝叶斯、随机森林等[3]才可实现对文本数据进行情感分类的高准确率和高效率。但是该方法依赖人工进行特征提取,进行复杂且大量的特征工程。

由于以上两种文本情感分析方法缺陷的存在,越来越多的学者使用深度学习相关的方法进行文本情感分类任务。卷积神经网络(CNN)、循环神经网络(RNN) [4]、双向门控循环神经网络(BiGRU)、多头注意力机制(Multi-Head-Attention)等在情感分析中均表现出优越性。

本文提出了一种可应用于文本情感分析领域的深度学习网络模型框架,即:基于多头注意力机制(Multi-Head-Attention Mechanism)的 BiGRU-CNN 的模型。

1) 使用双向门控循环(BiGRU)和卷积神经网络(CNN)相结合的模型框架,即 BiGRU-CNN 模型框架,进行特征提取。该模型整合了考虑上下文语义的 BiGRU 特征和卷积神经网络局部特征提取的优点,提高了模型分类性能。

2) 在 BiGRU-CNN 模型框架的基础上引入多头注意力机制(Multi-Head-Attention)。基于多头注意力机制的原理,引入该机制能够有区分度地捕捉文本评论中的情感信息,即获取到影响文本情感极性分类结果的重要文本特征。通过引入多头注意力,模型更关注于文本序列中对文本评论最终分类结果起关键和重要作用的部分,有效提升了模型的特征表达能力和分类准确率。

## 2. 相关工作

近年来,越来越多的研究人员和学者使用深度学习相关方法进行自然语言处理领域中情感分析的研究。与依赖大量特征工程且泛化能力较差的机器学习方法相比,深度学习方法具有显著的优势,其可自动获取文本情感特征,在各个领域均有很好的表现。

Kim 等[5]首次使用卷积神经网络(CNN)模型方法进行文本情感分析任务。通过卷积操作与最大池化操作获取文本代表性特征,随后输入到全连接层判断句子情感极性。刘龙飞等[6]提出分别将字级别和词级别向量作为 CNN 层的输入,进行文本情感分析任务。但是,CNN 模型存在缺陷,无法挖掘文本全局信息。长短期记忆模型(LSTM)引入了一种基于传统 RNN 的门控机制,可以获取全局信息特征。TANG D 等[7]提出使用 LSTM 模型进行文本分类,但是在如今看来,效果并非理想,原因在于未获取到文本序列前后的特征联系。XIAO Z 等[8]使用双向长短期记忆(BiLSTM)模型进行在本文情感分析任务,准确率有了显著提升,原因在于 BiLSTM 模型通过上下文训练学习到了文本前后文之间的特征和信息关联,弥补了 LSTM 的缺点,但 BiLSTM 网络结构复杂,所需参数较多,且难以进行并行训练。除此之外,当训练集和测试集数据量庞大时,所需计算时长较长。Cho K 等[9]使用门控循环(GRU)网络进行文本情感分析任务。GRU 网络就网络结构而言,比 LSTM 网络结构简单;就模型参数而言,比 LSTM 网络模型参数少,因此在模型训练过程中能有效缩短训练时间,在文本语义特征捕捉中能够捕获文本全局语义特征。缪亚林等[10]提出了一个 CNN-BiGRU 模型,即将卷积神经网络与双向门控循环神经网络相结合的文本情感分析模型,该模型结合卷积神经网络(CNN)和门控循环神经网络(GRU)各自的优势,模型表现优异。

上述神经网络模型在文本情感分析已经取得了很大的成绩,但是很难提取出多层次且全面的文本情感特征信息。注意机制[11]可以实现对重要信息的选择性和针对性的关注。林原等[12]使用双向 LSTM 与多头注意力机制相结合的模型进行文本情感分析。陈欣等[13]使用 GRU 和 CNN 混合网络模型提取文本隐层特征,随后引入注意力机制分配特征权重,最终模型效果优秀。

综上所述,CNN 模型在获取文本局部特征方面表现优异,而 BiGRU 模型网络结构简单从而使得其在模型训练方面表现突出。因此,本文采用 BiGRU-CNN 网络框架学习词嵌入处理后的文本特征,以更全面获取文本序列的局部特征和全局特征;引入多头注意力机制从不同子空间里捕捉文本特征信息,实现对重要信息进行选择。

## 3. 模型

为了实现对汽车评论文本进行情感分析,本文提出了一种基于多头注意力机制的双向 GRU 和 CNN 的网络模型结构(Sentiment analysis based on multi-head attention with BiGRU-CNN),具体如图 1 所示。该网络模型框架包括六层:词嵌入层、BiGRU 层、CNN 层、池化层、多头注意力层、输出层。

假设输入文本语句  $S = \{w_1, w_2, \dots, w_i, \dots, w_n\}$ , 其中,  $w_i$  代表  $S$  中的一个单词,  $n$  表示句子  $S$  中的词个数。

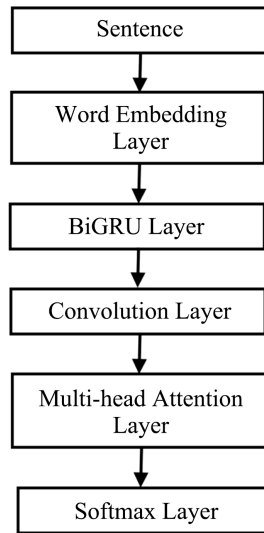


Figure 1. Model structure  
图 1. 模型结构图

### 3.1. 嵌入层

第一层是词嵌入层，该层的主要功能是将文本语句  $S$  表示为一个词向量矩阵  $X$ 。

通俗来讲，词嵌入是指将一个词语(word)转换为一个向量(vector)。常用的词嵌入方法有 Word2Vec [14]、Glove [15]、ELMo [16]和 BERT [17]。本文构建文本词向量采用的是 Google 开源词嵌入工具 Word2Vec 中的 Skip-gram 模型。

Word2vec 有两种模型，分别为跳字模型(Skip-gram)和连续词袋模型(CBOW)，这两种模型预测词的方式恰好相反。两种模型预测词的方式分别为：连续词袋模型通过某个词语的上下文语境来预测当前词语；跳字模型通过当前词语来预测周围的词。Skip-gram 模型结构如下图 2 所示。

在本文汽车文本评论数据集的实验中将句子  $S$  中的每个单词  $w_i$  使用 Word2Vec 中的 Skip-gram 模型转换为一个 100 维的词向量  $x_i$ ，将得到的词向量进行向量拼接计算，得到长度为  $n$  的句子  $S$  对应的词向量矩阵  $X$ ，具体表示如式(1)所示，使用该词向量矩阵  $X$  作为嵌入层的输出、BiGRU 层的输入。

$$X = w_1 \oplus w_2 \oplus \dots \oplus w_n \tag{1}$$

其中： $\oplus$  表示向量拼接操作。

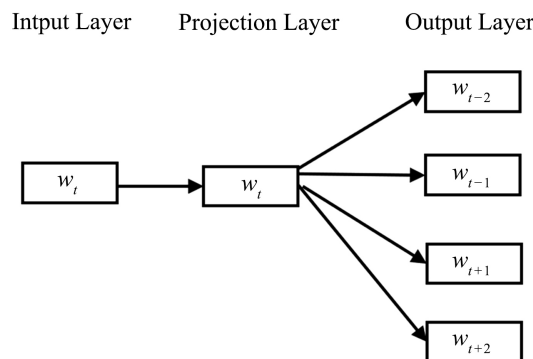


Figure 2. Skip-gram model  
图 2. Skip-gram 模型

### 3.2. BiGRU 层

第二层是 BiGRU 层，主要功能是提取文本序列的上下文特征即全局特征。

GRU 网络模型从原理上解决了递归神经网络模型(RNN)存在的梯度消失的问题，它可以结合之前时刻的历史信息来影响当前的输出，并提取序列数据中的上下文特征。在文本数据中，前面和后面的词都会影响当前词，因此我们使用 BiGRU 模型来提取输入文本的上下文特征。

GRU 模型可理解为 LSTM 模型的变体，常用于处理序列信息。对比变体模型，GRU 模型所做的改变有以下两点。一是将遗忘门和输入门合并成为更新门，故 GRU 模型只有重置门和更新门这两个门；二是把单元状态和隐藏状态合并，故 GRU 模型网络结构更简单，参数更少，更容易实现和计算。对于  $t$  时刻的输入  $x_t$ ，门控的表达式如式(2)~(6)所示，GRU 网络结构如图 3 所示。

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (2)$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad (3)$$

$$\tilde{h}_t = \tanh(W_{\tilde{h}} \cdot [r_t * h_{t-1}, x_t]) \quad (4)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (5)$$

$$y_t = \sigma(W_o \cdot h_t) \quad (6)$$

其中： $r_t$  为重置门，将新输入信息与原始信息合并； $z_t$  为更新门； $W_r$  与  $W_z$  为权值矩阵， $\sigma$  为 sigmoid 函数。

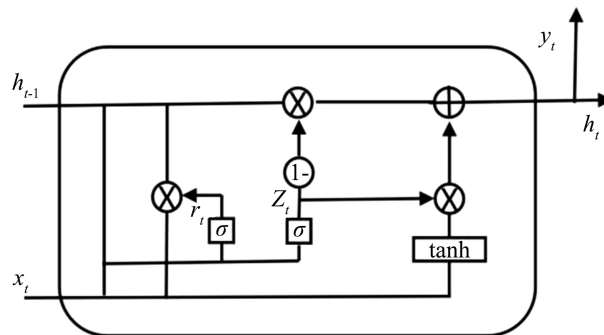


Figure 3. GRU unit structure

图 3. GRU 单元结构

双向 GRU (BiGRU) [18]模型包括一个正向 GRU 和一个反向 GRU，每次同时输入两个相反方向的 GRU，即正向 GRU 和反向 GRU，其中，正向 GRU 用于处理序列正向信息，反向 GRU 用于处理序列反向信息，最终结果由它们共同确定，使得结果更加准确。

本文中的模型利用双向 GRU 网络从词嵌入层输出矩阵  $X$  中学习上下文语义特征，即全局语义特征。在训练过程中，该网络同时使用两个 GRU 对文本序列的前后情感进行建模，对于词向量矩阵  $X$  中的  $x_i$ ，正向 GRU 和反向 GRU 获得的隐藏状态分别为  $h'_i$  和  $h''_i$ ，拼接  $h'_i$  和  $h''_i$  得到隐藏层  $h_i$ ，将提取到的特征组合得到向量  $H$ 。BiGRU 模型结构如图 4 所示，具体的计算过程见公式(7)~(10)：

$$h'_i = \overline{GRU}(X_i) \quad (7)$$

$$h''_i = \overline{GRU}(X_i) \quad (8)$$

$$h_i = [h'_i, h''_i] \tag{9}$$

$$H = (h_1, h_2, \dots, h_n) \tag{10}$$

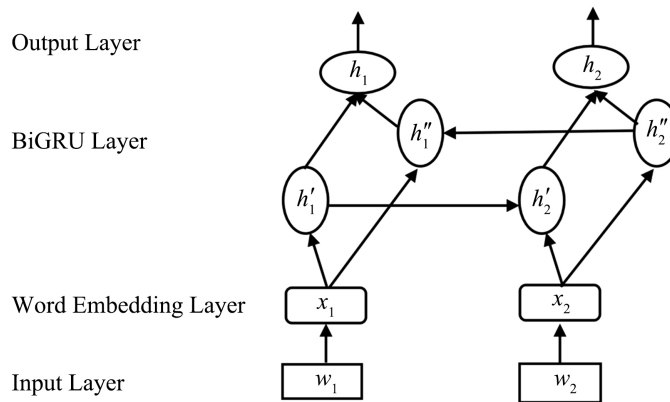


Figure 4. BiGRU network structure  
图 4. BiGRU 网络结构

### 3.3. 卷积层

第三层是卷积层(CNN 层)，主要功能是提取文本序列的 CNN 局部特征。

数据通过 BiGRU 层后获取到文本序列内部的全局依赖特征，但是未捕捉到文本的局部特征，而卷积神经网络 CNN 恰好可以充分捕捉文本局部特征，因此将 BiGRU 层的输出向量  $H$  作为卷积层的输入以获取文本局部特征。

对于输入矩阵  $H$  卷积运算为：

$$v_i = f(W \cdot H_{[i+k-1]} + b) \tag{11}$$

其中： $f$  为  $Relu$  激活函数， $W$  为卷积核权重矩阵， $k$  为卷积核的尺寸大小， $m$  为卷积核宽度， $b$  为偏置项。

$H_{[i+k-1]}$  经过卷积操作提取到的特征为  $v_i$ ，最终卷积层的输出为  $V = [v_1, v_2, \dots, v_{n-k+1}]$ 。

### 3.4. 池化层

第四层是池化层，其主要功能是对模型卷积层得到的文本特征进行特征压缩处理，从而得到模型框架下一层多头注意力机制层的输入。

常见的池化操作有：平均池化和最大池化。在本文实验中，对位于池化层上一层处理得到的特征矩阵进行  $k - \max$  最大池化操作得到向量  $l_i$ ，将得到的向量组合得到池化层的输出向量  $L$ 。具体操作计算如下：

$$l_i = \max(v_i, v_{i+1}, \dots, v_{i-k+1}) \tag{12}$$

$$L = (l_1, l_2, \dots, l_i, \dots, l_{m-k+1}) \tag{13}$$

其中， $m$  为向量  $v_i$  的维度。

### 3.5. 多头注意力层

第五层为多头注意力层，其主要功能是获取和学习文本序列内部的依赖关系，充分捕捉文本长距离

特征, 获取全局信息。

传统的注意机制包括注意力机制(Attention)和自注意力机制(Self-Attention), 均是仅从单一层面上获取注意力信息和注意力特征。而本文模型使用的多头注意机制获取信息相较于传统注意力机制模型更全面更有层次, 因为本文模型中加入多头注意力模型, 获得有助于文本情感倾向分类更准确的情感特征和情感信息。本质上, 多头注意机制是多个自注意力(Self-Attention) [19]的组合, 每个自注意力中都有三个矩阵分别为: Query 矩阵( $Q$ )、Key 矩阵( $K$ )和 Value 矩阵( $V$ ), 其网络模型如图 5 所示。

本文中, 池化层输出向量  $L$  为每个自注意力中的 Query 矩阵( $Q$ )、Key 矩阵( $K$ )、Value 矩阵( $V$ )三个矩阵的输入, Attention 的计算采用缩放点积(Scaled Dot-Product)方法, 计算如式(14)~(15)所示。

$$Q = K = V = L \quad (14)$$

$$\text{Attention}(Q, K, V) = \text{Soft max} \left( \frac{QK^T}{\sqrt{d_k}} \right) V \quad (15)$$

其中,  $Q$ 、 $K$ 、 $V$  通过池化层输出矩阵  $L$  与对应的权重矩阵  $W^Q$ 、 $W^K$ 、 $W^V$  相乘得到,  $\sqrt{d_k}$  为缩放因子,  $\sqrt{d_k}$  维度与矩阵向量  $Q$ 、 $K$ 、 $V$  的维度相同。

多头注意力机制的计算过程为:

1)  $Q$ 、 $K$ 、 $V$  向量矩阵进行多种线性变换生成多个头(head), 每个 head [20]被映射到不同子空间进行相对应的自注意力机制, 从而捕捉到不同表示子空间和不同维度的信息。

2) 每个 head 对应一组权重矩阵  $W^Q$ 、 $W^K$ 、 $W^V$  进行自注意力计算得到自注意力矩阵。

3) 拼接 2) 中计算所得的自注意力矩阵, 拼接后的拼接矩阵与附加权重矩阵  $W$  做矩阵乘法计算。

4) 将上述 3) 计算得到的结果作为输出层的输入。

具体计算过程如式(16)~(17)。

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (16)$$

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_k)W \quad (17)$$

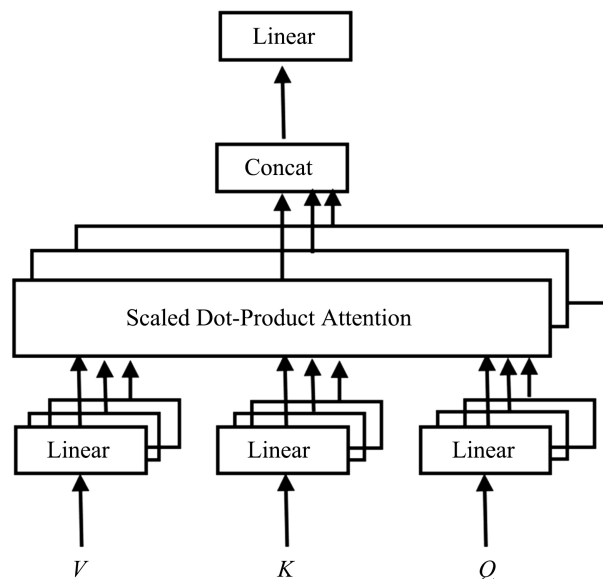


Figure 5. Multi-head attention mechanism network structure  
图 5. 多头注意力机制网络结构

### 3.6. 输出层

第六层为本文提出的模型框架的最后一层输出层，该层的主要功能是对多头注意力层输出的特征矩阵进行 0~1 分类，最终实现本文对文本评论进行情感分析的任务。

将多头注意力层的输出矩阵通过输出层中的 softmax 函数计算，得到输入句子 S 的情感倾向极性分类，具体计算公式如式(18)所示。

$$Y = \text{softmax}(W \cdot Z + b) \quad (18)$$

其中， $W$  为权重矩阵， $b$  为偏置项， $Z$  为多头注意力层的输出矩阵。

## 4. 实验

### 4.1. 实验环境

本文实验环境设置如表 1 所示。

**Table 1.** Experimental environment configuration  
**表 1.** 实验环境配置

实验环境	设置
操作系统	Windows10
硬件平台	NVIDIA Tesla
GPU	P100-16GB
编程语言	Python 3.7
深度学习框架	Pytorch 1.4.0
分词工具	Jieba
词向量训练工具	Word2vec

### 4.2. 实验数据

本文实验数据集是来自于太平洋汽车网的文本评论，该数据集共有 23,000 条，文本评论的标签有两类，分别为积极(Positive)和消极(Negative)，情感倾向被判定为积极的文本评论标签为 1，被判定为消极的文本评论标签为 0；其中训练集有 20,000 条，测试集有 3000 条；文本平均长度为 35，文本最大长度为 100，本文对句子长度做固定处理，句子长度固定为 100，即将少于该长度的评论用 0 填充。部分数据集如下表 2 所示。

**Table 2.** Data set  
**表 2.** 数据集

评论文本	情感极性
外观与后排直通后备箱小门，整车外观落落大方，极具个性，设计师功底深厚。 后排直通后备箱的小门特方便，这两点最令人叹服。	1
后备箱不敢说很小，但真的不大。要带大量东西出游的同志们可能要失望。	0



### 4.3. 实验设置

本文实验的模型超参数及参数设置如表 3 所示。

**Table 3.** Model hyperparameters and parameter settings

**表 3.** 模型超参数及参数设置

参数	值
词向量维度	100
句子最大长度	100
kernel_size	3
batch_size	128
loss	categorical_crossentropy
optimizer	adam
dropout	0.5
epoches	10
learning_rate	0.001
num_heads	2
hidden_nums	64

### 4.4. 实验评价指标

本文使用的模型评价指标为准确率和 F1 值。

评价指标相关参数如表 4 所示。

**Table 4.** Evaluation index parameters

**表 4.** 评价指标参数

	真实值 Positive	真实值 Negative
预测值 Positive	TP	FN
预测值 Negative	FP	TN

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (19)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (20)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (21)$$

$$F1 = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (22)$$

## 4.5. 实验结果与分析

设置 6 组对比实验与本文提出的基于多头注意力机制的 BiGRU-CNN 模型进行对比, 对比实验分别为:

### 1) TextCNN

该模型为单一的卷积神经网络, 将词向量输入到网络中, 进行最大池化操作, 全连接采用 softmax 进行分类。

### 2) BiGRU

该模型为单一的 BiGRU 网络模型, 将词向量输入到网络中, 输出层采用 softmax 进行文本分类。

### 3) BiGRU-Attention

该模型为 BiGRU 与 Attention 组合网络模型, 首先将词向量输入到 BiGRU 模型中获取文本全局特征, 将 BiGRU 特征输入到 Attention 层有重点地提取情感重要信息, 最终在输出层进行文本分类。

### 4) CNN-BiGRU-Attention

该模型为 CNN、BiGRU 与 Attention 组合网络模型, 首先将词向量输入到 CNN 层获取文本局部特征, 然后通过 BiGRU 层获取文本全局特征, 通过 Attention 层提取情感重要信息, 最终在输出层实现最终的文本情感倾向分类。

### 5) CNN-BiGRU-Multi-Head Attention

该模型为 CNN、BiGRU 与 Multi-Head Attention 的组合网络模型, 将词向量输入到 CNN 层获取文本局部特征, 随后通过 BiGRU 层获取文本全局特征, 最后通过多头注意力层提取情感信息, 在输出层进行文本分类。

### 6) BiGRU-CNN-Attention

该模型为 CNN、BiGRU 与 Attention 的组合网络模型, 将词向量输入到 BiGRU 层获取文本全局特征, 随后通过 CNN 层获取文本局部特征, 最后通过注意力层提取情感信息。

本文比较了基于多头注意力机制的 CNN 和双向 GRU 模型(BiGRU-CNN-Multi-Head Attention)与现有 6 种情感分析模型在汽车评论数据集上的分类效果。比较结果如表 5 和图 6 所示。从 Accuracy、F1 值这两个评价指标来看, BiGRU-CNN-Multi-Head Attention 的准确率达到 90.76%, F1 值达到 90.25%, 均优于对比模型, 比对照模型中最优分类模型准确率提高了 0.8%、F1 值提高了 0.37%。

**Table 5.** Comparison experiment results (unit: %)  
**表 5.** 对比实验结果(单位: %)

	模型	Acc	F1
1	TextCNN	87.26	86.76
2	BiGRU	87.69	87.43
3	BiGRU-Attention	88.42	88.16
4	CNN-BiGRU-Attention	88.91	88.73
5	CNN-BiGRU-Multi-Head Attention	89.76	89.53
6	BiGRU-CNN-Attention	89.96	89.62
7	BiGRU-CNN-Multi-Head Attention	90.76	90.25

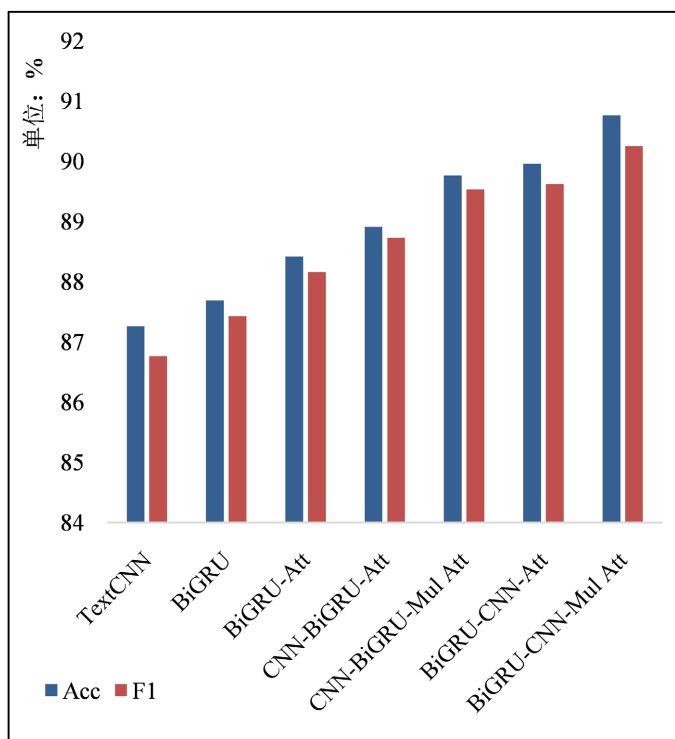


Figure 6. Comparison test results

图 6. 对比试验结果

使用 BiGRU-CNN 组合模型进行文本序列特征获取的模型与其他现有基础模型相比, 表现更优异, 其中 BiGRU-CNN-Attention 模型相对于 CNN-BiGRU-Attention 模型, 准确率(acc 值)提高了 1.05%, F1 值提高了 0.89%; BiGRU-CNN-Multi-Head Attention 模型相对于 CNN-BiGRU-Multi-Head Attention 模型准确率(acc 值)提高了 1%, F1 值提高了 0.72%。而以上模型优异的表现均源于 BiGRU-CNN 组合模型结合了 BiGRU 和 CNN 各自优点, 规避了各自的缺点, 使得模型对于文本特征的获取更深层, 更全面。

引入多头注意力机制层进行文本情感分析的模型与其他现有基础模型相比, 表现更优异, 其中 CNN-BiGRU-Multi-Head Attention 模型相对于 CNN-BiGRU-Attention 模型, 准确率(acc 值)提高了 0.86%, F1 值提高了 0.8%; BiGRU-CNN-Multi-Head Attention 模型相对 BiGRU-CNN-Attention 模型, 准确率(acc 值)提高了 0.8%, F1 值提高了 0.63%。而以上模型优异的表现均源于在模型中引入 Multi-Head Attention, 其能让模型有轻重, 有层次地训练学习与文本情感信息相关的特征, 进而使得模型的分类准确率更高。

## 5. 结论

本文提出一种基于多头注意力机制的 BiGRU-CNN 模型(Multi-Head Attention based on BiGRU-CNN model)用于汽车评论文本数据集的情感分析任务。为了全面地获取文本情感信息以提高模型的准确率和模型效果, 该方法采用 BiGRU-CNN 模型进行文本序列全局与局部特征的获取, 其中 BiGRU 对经过词嵌入处理的词向量矩阵获取文本序列全局特征作为 CNN 层的输入, 通过 CNN 层获取 BiGRU 层输出矩阵的局部特征信息; 引入多头注意力机制进行特征权重分配以有针对性和有层次地捕捉文本情感重要信息。在汽车评论数据上进行对比实验, 评价指标均由于现有情感分析模型。

在下一步工作中, 可以考虑讲词嵌入进行改进, 采用多种嵌入方法进行融合词向量特征, 此外还可以将多头注意力机制进行改进进一步提高模型的准确率。

## 参考文献

- [1] Nasukawa, T. and Yi, J. (2003) Sentiment Analysis: Capturing Favorability Using Natural Language Processing. *Proceedings of the 2nd International Conference on Knowledge Capture*, Sanibel Island, 23-25 October 2003, 70-77. <https://doi.org/10.1145/945645.945658>
- [2] 李然, 林政, 林海伦, 等. 文本情绪分析综述[J]. 计算机研究与发展, 2018, 55(1): 30-52.
- [3] 梁军, 柴玉梅, 原慧斌, 等. 基于深度学习的微博情感分析[J]. 中文信息学报, 2014, 28(5): 155-161.
- [4] Irsoy, O. and Cardie, C. (2014) Opinion Mining with Deep Recurrent Neural Networks. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, Doha, 25-29 October 2014, 720-728. <https://doi.org/10.3115/v1/D14-1080>
- [5] Kim, Y. (2014) Convolutional Neural Networks for Sentence Classification. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, Doha, 25-29 October 2014, 1746-1751. <https://doi.org/10.3115/v1/D14-1181>
- [6] 刘龙飞, 杨亮, 张绍武, 等. 基于卷积神经网络的微博情感倾向性分析[J]. 中文信息报, 2015, 29(6): 159-165.
- [7] Tang, D., Qin, B., Feng, X., et al. (2016) Effective LSTMs for Target Dependent Sentiment Classification. *Proceedings of Conference on COLING*, Osaka, 11-16 December 2016, 3298-3307.
- [8] Xiao, Z. and Liang, P.J. (2016) Chinese Sentiment Analysis Using Bidirectional LSTM with Word Embedding. *Proceedings of the 2016 International Conference on Cloud Computing and Security*, Nanjing, 29-31 July 2016, 601-610. [https://doi.org/10.1007/978-3-319-48674-1\\_53](https://doi.org/10.1007/978-3-319-48674-1_53)
- [9] Cho, K., Merriënboer, B.V., Gulcehre, C., et al. (2014) Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation. <https://arxiv.org/abs/1406.1078>
- [10] 缪亚林, 姬怡纯, 张顺, 等. CNN-BiGRU 模型在中文短文情感分析的应用[J]. 情报科学, 2021, 39(4): 85-91.
- [11] Bahdanau, D., Cho, K. and Bengio, Y. (2014) Neural Machine Translation by Jointly Learning to Align and Translate. <https://arxiv.org/abs/1409.0473>
- [12] 林原, 李家平, 许侃, 等. 基于多头注意力的双向 LSTM 情感分析模型研究[J]. 山西大学学报(自然科学版), 2020, 43(1): 1-7.
- [13] 陈欣, 杨小兵, 姚雨虹, 等. 字词融合的双通道混合神经网络情感分析模型[J]. 小型微型计算机系统, 2021, 42(2): 279-284.
- [14] Wei, P., Xu, N. and Mao, W. (2019) Modeling Conversation Structure and Temporal Dynamics for Jointly Predicting Rumor Stance and Veracity. *Proceedings of the Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, Hong Kong, 3-7 November 2019, 4789-4800. <https://doi.org/10.18653/v1/D19-1485>
- [15] Pennington, J., Socher, R. and Manning, C.D. (2014) GloVe: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, Doha, 25-29 October 2014, 1532-1543. <https://doi.org/10.3115/v1/D14-1162>
- [16] Devlin, J., Chang, M.W., Lee, K., et al. (2019) BERT: Pretraining of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, Minneapolis, 2-7 June 2019, 4171-4186.
- [17] Wang, J., Yu, L.C., Lai, K.R., et al. (2016) Dimensional Sentiment Analysis Using a Regional CNN-ISTM Model. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, Berlin, 7-12 August 2016, 225-230. <https://doi.org/10.18653/v1/P16-2037>
- [18] Chen, Q., Zhu, X., Ling, Z.H., et al. (2017) Enhanced LSTM for Natural Language Inference. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Vancouver, 30 July-4 August 2017, 1657-1668. <https://doi.org/10.18653/v1/P17-1152>
- [19] Sun, X., Gao, Y., Sutcliffe, R., et al. (2019) Word Representation Learning Based on Bidirectional GRUs with Drop Loss for Sentiment Classification. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, **99**, 1-11.
- [20] Zhang, H., Jin, W., Zhang, J., et al. (2017) YNU-HPCC at SemEval 2017 Task 4: Using a Multi-Channel CNN-LSTM Model for Sentiment Classification. *Proceeding of the 11th International Workshop on SEMANTIC Evaluation*, Vancouver, 1 January 2017, 796-801. <https://doi.org/10.18653/v1/S17-2134>