

周荣, 钱钢. 基于 BERT-BiGRU-CNN 的讽刺文本分类研究[J]. 智能计算机与应用, 2026, 16(3): 94-99. DOI: 10.20169/j.issn.2095-2163.24042805

基于 BERT-BiGRU-CNN 的讽刺文本分类研究

周荣, 钱钢

(南京审计大学 计算机学院, 南京 211815)

摘要: 针对目前讽刺文本分类存在一词多义和准确性不高的问题, 提出了一种基于 BERT-BiGRU-CNN 的文本分类模型。首先, 使用 BERT 预训练模型对短文本进行句子层面的特征向量表示; 其次, 通过双向门控循环单元 (BiGRU) 获取文本全局序列特征; 最后, 采用卷积神经网络 (CNN) 提取局部重点特征, 以此提高模型特征提取性能。实验采用 github 上的 Twitter_sarcastic 数据集, 实验结果证明, 基于 BERT-BiGRU-CNN 的文本分类模型在数据集上的准确率、F1 评分等指标方面取得了良好的效果, 明显优于现有模型。

关键词: 讽刺文本二分类; BERT 预训练模型; 双向门控循环单元; 卷积神经网络

中图分类号: TP391.1 **文献标志码:** A **文章编号:** 2095-2163(2026)03-0094-06

Research on the classification of satirical texts based on BERT-BiGRU-CNN

ZHOU Rong, QIAN Gang

(School of Computer Science, Nanjing Audit University, Nanjing 211815, China)

Abstract: In order to solve the problem of polysemy and low accuracy in the classification of satirical texts, a text classification model based on BERT-BiGRU-CNN is proposed. Firstly, the BERT pre-trained model is used to represent the feature vectors at the sentence level for short texts. Secondly, the global sequence features of the text are obtained through Bidirectional Gated Recurrent Units (BiGRU). Finally, the Convolutional Neural Networks (CNN) is used to extract local key features to improve the feature extraction performance of the model. The experimental results show that the text classification model based on BERT-BiGRU-CNN has achieved good results in terms of accuracy and F1 score on the dataset, which is significantly better than the existing models.

Key words: sarcastic text classification; BERT pre-trained models; Bidirectional Gated Recurrent Units; Convolutional Neural Networks

0 引言

讽刺是一种在日常生活中常见的复杂情绪表达。随着社交媒体在近年来的广泛使用, 社交媒体上也出现了大量讽刺文本, 对于文本分类的要求也越来越高。目前常见的情感分类模型更多的是对直抒胸臆的情感文本进行分类, 面对讽刺这种具有隐喻、反讽意义的文本分类效果不佳。其中, 基于规则的讽刺识别方法精度较高, 但需要耗费大量的时间和人力来编写规则, 另外规则的泛化性也是很大的问题, 即规则无法很好地在不同领域间进行迁移。基于统计机器学习的方法需要人工进行大量的特征工程, 另外还需要大量有标注的讽刺数据, 同样需要

消耗大量的时间和人力。近年来研究的主流是基于有监督的神经网络方法。该方法能减少耗时耗力的手工提取特征的依赖。比如, “今天又下雨了, 没带伞, 好开心啊!” 语义上判别是在反讽, 事实上情感分类模型只能捕捉到关键词“开心”, 从而导致分类结果精确率不高的问题, 提高讽刺文本分类准确率是本文的主要研究内容。

本文提出了一种 BERT-BiGRU-CNN 文本分类模型对目标文本讽刺性进行二分类。针对目标特点改进了 BERT 模型, 首先, 使用 BERT 预训练模型对短文本进行句子层面的特征向量表示; 其次, 通过 BiGRU 获取文本全局序列特征; 最后, 采用 CNN 提取局部重点特征, 再通过 Softmax 函数完成分类, 以

作者简介: 周荣 (1999—), 女, 硕士研究生, 主要研究方向: 自然语言处理, 情感分析模型。

通信作者: 钱钢 (1965—), 男, 博士, 教授, 博士生导师, 主要研究方向: 大数据审计, 现代审计技术与方法等。Email: qgmail@vip.sina.com。

收稿日期: 2024-04-28

此提高模型特征提取性能。实验结果证明,基于BERT-BiGRU-CNN的文本分类模型在数据集上的准确率、F1评分等指标方面取得了良好的效果,明显优于现有模型。

1 研究现状

Cahyani等学者^[1]研究比较了TF-IDF和Word2Vec模型在情感文本分类中表示特征的性能,使用支持向量机(SVM)和多项式朴素贝叶斯(MNB)方法对通勤线路和雅加达推特数据上的情感文本进行分类,对比得出TF-IDF建模比Word2Vec建模具有更好的性能,并且与之前的研究相比,这项研究提高了分类性能。由于讽刺意味的新闻文本和假新闻文本之间存在高度相似的文本线索,可能会造成一些有意义的讽刺新闻文本被消除,为了解决此类问题,Low等学者^[2]提出了基于DistilBERT架构构建的分类器框架,其性能明显优于现有的机器学习方法。Behzadidoost等学者^[3]提出了颗粒计算中第一个Stacking-BiLSTM-SVM模型,也是第一个在基于深度学习的文本分类背景下使用颗粒计算进行新表示的数据增强的研究,旨在提高基于复杂自然语言问题的分类器精度。Yogatama等学者^[4]对判别式和生成式LSTM模型的文本分类性能进行了经验表征之后发现,生成模型比其判别模型更快地接近其渐近错误率,基于此发现,也证实基于RNN的生成分类模型对数据分布的变化更具鲁棒性。Du等学者^[5]采用一种广义学习系统(BLS),将LSTM与RNN结合而构建了文本分类模型,以此来提高训练效率,同时在文本分类方面也实现了更高的精度。Abas等学者^[6]提出一种新的BERT-CNN模型中检测文本情感,该模型采用BERT训练单词语义表示语言模型。根据单词上下文动态生成语义向量,然后放入CNN中预测输出。文献[7-10]均是采用BERT-CNN模型对文本进行分类处理,周海波等学者^[8]将此模型运用在短文本讽刺文本分类研究上得到了出色的效果。然而,Emanuel等学者^[10]提出CNN在某些应用中并不总是足够准确。在CNN架构中激活函数的选择会影响到CNN的有效性。为了解决LSTM和CNN各自的不足并利用其独特优势,Jang等学者^[11]提出了一个Bi-LSTM-CNN混合模型,该模型使用CNN从句子的不同位置提取特征,从而减少输入特征的数量,再将LSTM用于从卷积层获得的特征中提取上下文信息。文献[12-13]均采用BiLSTM-CNN的方法进

行文本分类处理。基于此,Bai^[14]提出了一种结合长短期记忆单元和注意力机制的改进文本分类方法,该算法优于存在长距离依赖问题的RNN算法和CNN算法。鉴于BERT模型在总结冗长的文档方面存在局限性,Bano等学者^[15]提出了一个新的方法,使用BERT生成句子级嵌入,然后将其输入BiGRU网络,这就能够全面理解完整文档的上下文。同时,在自然语言处理领域存在着隐喻表达、语义多样性和语法特异性等问题,Yu等学者^[16]提出了BERT-BiGRU模型。文献[17-20]是用BERT-BiGRU模型对文本进行分类的研究,结果表明,BERT-BiGRU模型在中文文本分类任务中具有良好的性能。为了进一步提高文本分类的准确性,Huang等学者^[21]提出了一种新的基于BERT嵌入的DCNN-BiGRU(Deep Convolutional Neural Network Bidirection Gated Recurrent)文本分类模型。这样一来,语义向量不仅包含文本的局部特征,还包含文本的上下文特征。Yan等学者^[22]针对大多数情感分析模型难以捕捉文本中复杂的语义和语法信息的问题,在CNN-BiGRU模型中加入注意机制,通过计算注意分数对模型施加不同的学习权值,既节省了时间,又提高了精度。

从近年来研究现状可见,现有的情感分类模型在文本分类领域已经有了非常高的准确率,但是对于讽刺文本分类识别的研究却仍然少见,讽刺文本这些具有反讽、隐喻含义的文本在情感分析领域的分类效果不佳。因此,为了提高讽刺文本的分类准确率,本文采用BERT生成动态词向量,针对自然语言处理领域存在着隐喻表达、语义多样性和语法特异性等问题,利用BiGRU和CNN模型融合,使语义向量不仅包含文本的局部特征,还包含文本的上下文特征,实验结果证明该模型对于讽刺文本的分类明显优于现有模型。

2 BERT-BiGRU-CNN 讽刺文本分析算法构建

本文提出的BERT-BiGRU-CNN模型实现讽刺文本二分类任务,其结构如图1所示。首先对输入的文本进行文本预处理,去除特殊字符,再经过BERT模型生成词向量,利用BiGRU模型提取全局特征,然后经过CNN模型提取局部特征,最后经由全连接层分类输出文本。

2.1 基于BERT模型的文本表示

BERT模型是2018年Google提出的以Transformer模型的编码部分为基础,由12层

Transformer 叠加而成,构造了一个多层双向的网络架构。其模型结构如图 2 所示。

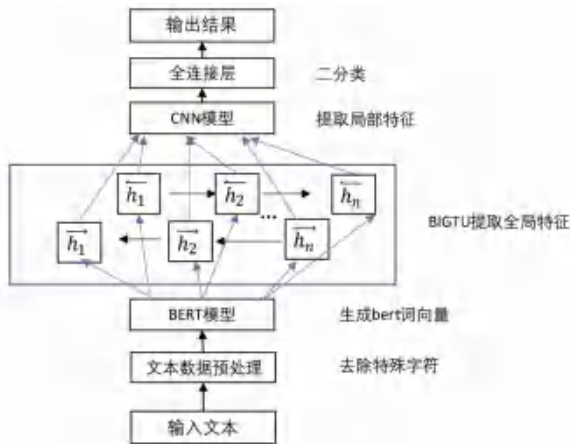


图 1 BERT-BiGRU-CNN 模型
Fig.1 BERT-BiGRU-CNN model

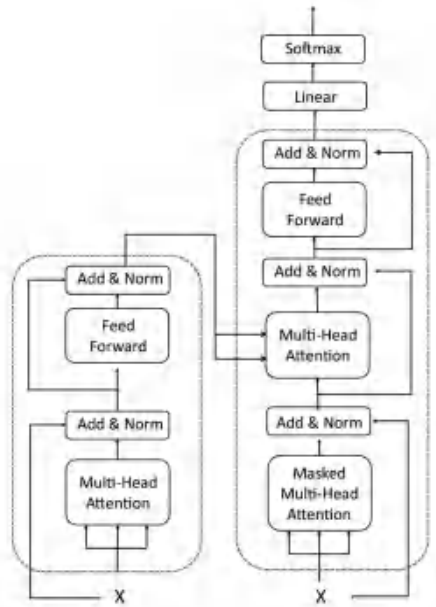


图 3 Transformer 模型
Fig. 3 Transformer model

BERT 模型进行训练有 2 个步骤:预训练和微调。预训练过程主要包含 MLM (Masked Language Model) 掩码任务和 NSP (Next Sentence Prediction) 语句预测任务。预训练的目标是在大型语料库上训练出包含丰富语义信息的文本语义表征,然后将文本语义表征微调运用到具体的 NLP 任务中。MLM 任务中,对 BERT 输入一段文本,将其中部分单词用 Msak 或其它单词替代,让模型根据没有被覆盖的单词预测被掩盖或替换的单词,旨在解决上下文语境缺失的问题。NSP 任务中,给定 BERT 两个句子,并预测这 2 个句子是否连续。作用是为了让模型尽可能全面准确地理解 2 个句子之间的关系。

2.2 基于 BiGRU 的文本特征提取

BiGRU 模型是一种递归神经网络 (Recurrent Neural Network, RNN) 模型,用于处理序列数据,如文本、语音等。该模型的基本单位是 GRU 单元,每个 GRU 单元包含一个更新门 (Update Gate) 和一个重置门 (Reset Gate),用于控制信息的流动。GRU 只能处理单向时间序列,无法同时学习上下文两个方向的信息。BiGRU 模型通过引入双向循环结构,将一个 GRU 单元分为 2 部分,一部分从前向后处理序列数据,另一部分从后向前处理序列数据。这样一来,模型可以同时捕捉到过去和未来的上下文信息,从而更好地理解序列数据,提取文本信息的全局语义特征。其中,GRU 的模型结构如图 4 所示,参数说明见表 1。

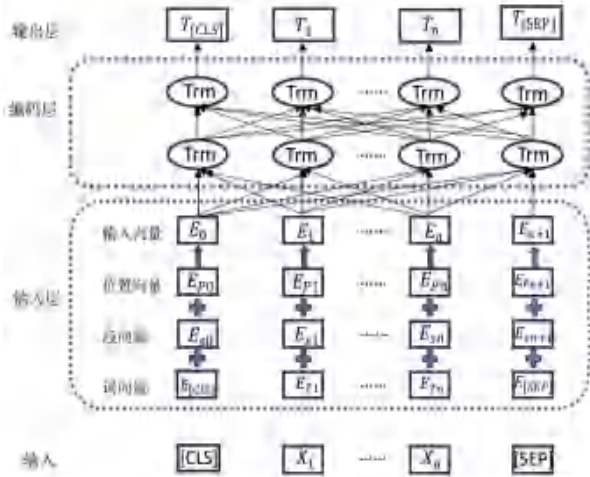


图 2 BERT 模型
Fig. 2 BERT model

(1) 输入层。输入向量 E 由词向量 (Token Embeddings)、段向量 (Segment Embeddings) 和位置向量 (Positions Embeddings) 叠加而成,可表示为:

$$E(X) = EP + ES + ET \quad (1)$$

(2) 编码层。编码层是 BERT 模型的核心,其内部是由多层双向 Transformer 编码器构成。Transformer 的编码端由 6 个 Encoder 构成,解码端由 6 个 Decoder 构成。通过采用多头自注意机制来全方位地获取丰富的语义信息。不仅如此,还采用残差网络架构缓解模型梯度消失和爆炸的问题。Encoder 由输入、多头自注意机制和全连接神经网络组成,如图 3 所示。

(3) 输出层。输入的文本 $X = [x_1, x_2, \dots, x_n]$ 经过输入层和编码层后,得出基于 BERT 模型的动态词向量矩阵 $T = [T_1, T_2, \dots, T_n]$ 。

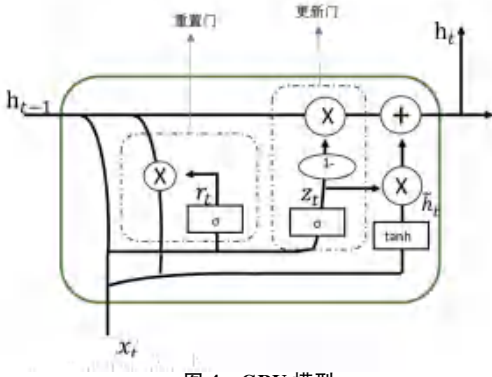


图 4 GRU 模型

Fig. 4 GRU model

表 1 参数说明表

Table 1 Parameter description table

参数	说明
\vec{h}_t	GRU 单元正向输出
\overleftarrow{h}_t	GRU 单元反向输出
GRU	向量非线性变换函数
x_t	文本词向量输入
w_t	正向权重矩阵
v_t	反向权重矩阵
b_t	偏置向量
\vec{h}_{t-1}	正向隐藏状态
\overleftarrow{h}_{t-1}	反向隐藏状态
h_{t-1}'	重置
r	重置门门控
z	更新门门控

利用 BiGRU 网络生成 t 时刻正、反向语义信息特征向量 \vec{h}_t 和 \overleftarrow{h}_t , 其隐藏层状态 h 由正反向隐藏状态加权得到, 计算公式为:

$$\vec{h}_t = \text{GRU}(x_t, \vec{h}_{t-1}) \quad (2)$$

$$\overleftarrow{h}_t = \text{GRU}(x_t, \overleftarrow{h}_{t-1}) \quad (3)$$

$$h = w_t \vec{h}_t + v_t \overleftarrow{h}_t + b_t \quad (4)$$

单向 GRU 在接收词向量序列后经重置门“重置”之后得到数据 h_{t-1}' , 经过激活函数映射, 使用更新门进行更新得到特征 h_t , 表达式为:

$$h_{t-1}' = h_{t-1} \odot r \quad (5)$$

$$h' = \tan(w[x_t, h_{t-1}']) \quad (6)$$

$$h_t = (1 - z) \odot h_{t-1}' + z \odot h' \quad (7)$$

2.3 基于 TextCNN 的文本特征提取

利用 BiGRU 获取全局特征后, 再通过卷积神经网络获取文本的局部语义特征。将 CNN 模型划分为 4 层。

第一层为输入层。将输入的文本序列转化为词嵌入向量, 每个单词对应一个向量, 并将这些向量按序列顺序组成一个矩阵。

第二层为卷积层。本模型采用多个不同大小的

卷积核来同时获取文本的不同特征。

第三层为池化层。池化层的作用是对卷积结果进行采样, 缩减特征, 在一定程度上防止过拟合。

第四层为全连接层。将池化后的特征连接, 用于学习特征之间的关系和进行最终的分类, 模型结构如图 5 所示。

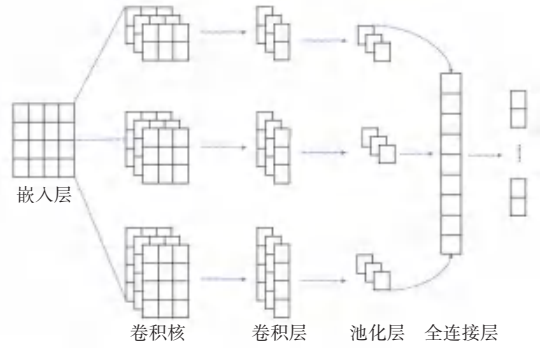


图 5 卷积神经网络

Fig. 5 Convolutional neural networks

3 实验

3.1 数据集简介

为了验证 BERT-BiGRU-CNN 讽刺文本分类模型性能的优劣, 本文采用 GitHub 上的开源数据集 Twitter_sarcastic, 该数据集对文本的讽刺性进行分类, 数据集的文本数据由 13 611 条讽刺评论和 15 008 条非讽刺评论构成。本文将数据集按照 8 : 1 : 1 的比例将数据集随机划分为训练集和测试集, 对文本是否为讽刺文本进行分类。实验数据设置见表 2。

表 2 实验数据设置

Table 2 Experimental data settings

数据集	划分	正向文本	负向文本
Twitter_sarcastic	训练	14 502	13 117
	测试	250	250
	验证	256	244

3.2 实验环境

本实验环境配置见表 3。

表 3 实验环境

Table 3 Experimental environment

实验环境	配置
操作系统	Windows10
GPU	NVIDIA GeForce MX150
CPU	Intel(R) Core(TM) i5-8250U
Python	3.6
深度学习框架	Pytorch
编码格式	UTF-8

3.3 模型参数设置

本文的 BERT 模型采用的是 Google 提供的开源

BERT 预训练模型 BERT-Base, 实验模型参数设置见表 4、表 5。

表 4 BERT 模型参数

Table 4 BERT model parameters

参数	值
学习率	5e-5
词向量维度	128
Dropout	0.5
Transformer 层数	12
BiGRU 隐藏层	128

表 5 CNN 模型参数

Table 5 CNN model parameters

参数	值
卷积核尺寸	2, 3, 4
卷积核数量	256
激活函数	Relu
Epoch	5

3.4 评价指标

本实验使用准确率、精确率、召回率和 $F1$ 值作为讽刺文本分类模型的评估标准。分类结果的混淆矩阵见表 6, 参数说明见表 7。

表 6 分类结果的混淆矩阵

Table 6 Confusion matrix for classification results

真实情况	预测情况	
	正例	反例
正例	真正例 (TP)	假反例 (FN)
反例	假正例 (FP)	真反例 (TN)

表 7 参数说明表

Table 7 Parameter description table

参数	预测样本个数	实际样本个数
TP	正	正
FP	负	正
FN	正	负
TN	负	负

其计算公式为:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

$$P = \frac{TP}{TP + FP} \quad (9)$$

$$R = \frac{TP}{TP + FN} \quad (10)$$

$$F1 = \frac{2PR}{P + R} \quad (11)$$

其中, $F1$ 值表示精确率和召回率的加权平均。

3.5 实验结果与分析

为验证 BERT-BiGRU-CNN 讽刺文本分类模型的识别分类效果, 以 BERT-CNN 模型、BERT-GRU 模型、BERT-BiGRU 模型为基线进行了对比实验, 实验结果见表 8。各模型依次为:

(1) BERT-CNN 模型。直接利用 BERT 模型获取动态词向量, 再通过 CNN 网络模型提取局部重点特征, 缺点是无法提取全局语义特征, 对于讽刺文本的分类不够细致。

(2) BERT-GRU 模型。同样利用 BERT 模型获取动态词向量, 再将词向量输入 GRU 获取全局语义特征, 缺点是 GRU 只能单向提取, 无法联系上下文语义关系, 存在一词多义等问题, 容错率较低。

(3) BERT-BiGRU 模型。依旧利用 BERT 模型获取动态词向量, 同时双向 GRU 可以联系上下文, 提升训练效果, 缺点是对于关键词无法提取局部重点特征。

(4) BERT-BiGRU-CNN 模型。在 BERT-BiGRU 模型基础上引入 CNN 模型。

表 8 实验结果

Table 8 Experimental results

模型	准确率	精确率	召回率	$F1$ 值	%
BERT-CNN	89.6	88.7	89.5	89.4	
BERT-GRU	92.6	92.8	92.3	91.9	
BERT-BiGRU	95.7	95.3	94.6	94.9	
BERT-BiGRU-CNN	98.6	98.9	97.8	98.5	

由表 8 的结果可以看出, 在推特数据集上, BERT-BiGRU-CNN 讽刺文本分类模型的准确率、精确率、召回率和 $F1$ 值分别达到了 98.6%、98.9%、97.8% 和 98.5%。从实验数据可以看出, BERT-BiGRU-CNN 讽刺文本分类模型在 Twitter_sarcastic 数据集上的分类效果明显优于其它几个模型。该模型利用 BERT 模型获取动态词向量, BiGRU 对长文本序列特征的有效提取和 CNN 对局部重点特征的高效获取, 提升了模型的特征提取能力, 从而取得了较好的分类效果。

4 结束语

针对讽刺文本分析任务中出现的讽刺、反讽等情感特征提取效果不佳的问题, 本文提出了一种 BERT-BiGRU-CNN 讽刺文本分类模型。该模型使用 BERT 将全部的文本转化为动态词向量, 再输入 BiGRU 网络, 提取整体语义特征, 再结合 CNN 网络提取局部语义特征, 从而提升了讽刺文本分类的特

征质量。经过 Twiter_sarcastic 数据集上的测试,其准确率达到了 98.6%,体现了该模型的分​​类效果明显优于其它模型。现有的讽刺文本分类仍旧存在一些问题,比如缺少完整的权威的中文讽刺文本数据集,现有的讽刺文本都是短文本等,在后续的研究中将针对这些问题开展进一步工作。

参考文献

- [1] CAHYANI D E, PATASIK I. Performance comparison of tf-idf and word2vec models for emotion text classification[J]. Bulletin of Electrical Engineering and Informatics, 2021, 10(5): 2780-2788.
- [2] LOW J F, FUNG B C M, IQBAL F, et al. Distinguishing between fake news and satire with transformers [J]. Expert Systems with Applications, 2022, 187: 115824.
- [3] BEHZADIDOOST R, MAHAN F, IZADKHAH H. Granular computing - based deep learning for text classification [J]. Information Sciences, 2024, 652: 119746.
- [4] YOGATAMA D, DYER C, WANG Ling, et al. Generative and discriminative text classification with recurrent neural networks[J]. arXiv preprint arXiv,1703.01898, 2017.
- [5] DU Jie, VONG C M, CHEN C L P. Novel efficient RNN and LSTM-like architectures; Recurrent and gated broad learning systems and their applications for text classification [J]. IEEE Transactions on Cybernetics, 2020, 51(3): 1586-1597.
- [6] ABAS A R, ELHENAWY I, ZIDAN M, et al. BERT-CNN: A deep learning model for detecting emotions from text [J]. Computers, Materials & Continua, 2022, 71(2): 2943-2961.
- [7] KAUR K, KAUR P. BERT - CNN; Improving BERT for requirements classification using CNN [J]. Procedia Computer Science, 2023, 218: 2604-2611.
- [8] 周海波,李天. 基于 BERT-CNN 中间任务转移模型的短文本讽刺文本分类研究 [J]. 智能计算机与应用, 2023, 13(5): 156-160.
- [9] HAMED K A, SALEH F A, AKHIL K, et al. A novel multi-layer feature fusion-based BERT-CNN for sentence representation learning and classification [J]. Robotic Intelligence and Automation, 2023, 43(6): 704-715.
- [10] EMANUEL R H K, DOCHERTY P D, LUNT H, et al. The effect of activation functions on accuracy, convergence speed, and misclassification confidence in CNN text classification; A comprehensive exploration [J]. The Journal of Supercomputing, 2024, 80(1): 292-312.
- [11] JANG B, KIM M, HARERIMANA G, et al. Bi-LSTM model to increase accuracy in text classification; Combining Word2vec CNN and attention mechanism [J]. Applied Sciences, 2020, 10(17): 5841.
- [12] LIU Gang, GUO Jiabao. Bidirectional LSTM with attention mechanism and convolutional layer for text classification [J]. Neurocomputing, 2019, 337: 325-338.
- [13] ZHOU Peng, QI Zhenyu, ZHENG Suncong, et al. Text classification improved by integrating bidirectional LSTM with two-dimensional max pooling [J]. arXiv preprint arXiv, 1611.06639, 2016.
- [14] BAI Xue. Text classification based on LSTM and attention [C]// Proceedings of 2018 Thirteenth International Conference on Digital Information Management (ICDIM). Piscataway, NJ: IEEE, 2018: 29-32.
- [15] BANO S, KHALID S, TAIRAN N M, et al. Summarization of scholarly articles using BERT and BiGRU; Deep learning-based extractive approach [J]. Journal of King Saud University - Computer and Information Sciences, 2023, 35(9): 101739.
- [16] YU Qing, WANG Ziyin, JIANG Kaiwen. Research on text classification based on Bert - BiGRU model [J]. Journal of Physics; Conference Series, 2021, 1746: 012019.
- [17] SHENG Deming, YUAN Jingling. An efficient long Chinese text sentiment analysis method using BERT-based models with BiGRU [C]// Proceedings of 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD). Piscataway, NJ: IEEE, 2021: 192-197.
- [18] 张鑫玉,才智杰. 基于 Bert-BiGRU-CNN 的文本情感分析 [J]. 计算机仿真, 2023, 40(7): 519-523.
- [19] ZHU Q, JIANG X, YE R. Sentiment analysis of review text based on BiGRU-attention and hybrid CNN [J]. IEEE Access, 2021, 9: 149077-149088.
- [20] 李芸,潘雅丽,肖冬. 基于改进 BERT-BiGRU 模型的文本情感分类研究 [J]. 电子技术应用, 2023, 49(2): 9-14.
- [21] HUANG H, JING X Y, WU F, et al. DCNN-BiGRU text classification model based on BERT embedding [C]// Proceedings of 2019 IEEE International Conferences on Ubiquitous Computing & Communications (IUCC) and Data Science and Computational Intelligence (DSCI) and Smart Computing, Networking and Services (SmartCNS). Piscataway, NJ: IEEE, 2019: 632-637.
- [22] YAN Wei, ZHOU Lifan, QIAN Zhengjiang, et al. Sentiment analysis of student texts using the CNN-BiGRU-AT model [J]. Scientific Programming, 2021, 2021: 8405263.