

# 基于 BERT-BiGRU 集成学习的情感语义识别

游 兰<sup>1,2</sup>, 曾 晗<sup>1,2</sup>, 韩凡宇<sup>1,3</sup>, 金 红<sup>1,2,3</sup>, 崔海波<sup>1,3,4</sup>, 张家合<sup>1,4</sup>

- (1. 湖北大学 计算机与信息工程学院, 湖北 武汉 430062;
2. 湖北省软件工程技术研究中心, 湖北 武汉 430062;
3. 智慧政务与人工智能应用湖北省工程研究中心, 湖北 武汉 430062;
4. 湖北省教育信息化工程技术研究中心, 湖北 武汉 430062)

**摘 要:**如今,在社交网络上发表评论已成为公众对社会现象表达态度和立场的主要方式。精准识别社交文本的情感倾向性对于舆情管控、社会维稳等有重要价值。针对传统情感识别模型大多集中于评论的表层语义挖掘,存在分类效果不佳、泛化能力有限等问题,提出了一种基于 BERT-BiGRU 多模集成学习的深层情感语义识别方法。首先,通过 BERT 预训练模型获取评论文本的上下文语义特征表示,再结合 BiGRU 提取深层非线性特征向量,实现单模型下的最优效果;接着,为了使模型效果稳定且多方面表现均衡,基于 BERT 系列预训练模型训练出表现优异且具有差异化的多个情感分类器;最后,利用数据扰动和投票策略的集成学习方法,实现各模型深层特征的充分融合。实验结果显示:BERT-BiGRU 模型相较于其他传统模型,在两个公开数据集(COV19 和 ChnSenti)上具有更优的情感识别效果。

**关键词:**情感识别;BERT 预训练模型;双向门控循环单元;集成学习;深层特征

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2023)05-0159-08

doi:10.3969/j.issn.1673-629X.2023.05.024

## Sentiment Semantic Recognition Based on BERT-BiGRU Ensemble Learning

YOU Lan<sup>1,2</sup>, ZENG Han<sup>1,2</sup>, HAN Fan-yu<sup>1,3</sup>, JIN Hong<sup>1,2,3</sup>, CUI Hai-bo<sup>1,3,4</sup>, ZHANG Jia-he<sup>1,4</sup>

- (1. School of Computer Science and Information Engineering, Hubei University, Wuhan 430062, China;
2. Engineering and Technical Research Center of Hubei Province in Software Engineering, Wuhan 430062, China;
3. Engineering Research Center of Hubei Province in Intelligent Government Affairs and Application of Artificial Intelligence, Wuhan 430062, China;
4. Engineering and Technical Research Center of Hubei Province in Educational Information, Wuhan 430062, China)

**Abstract:** Nowadays, posting comments on social networks has become one of the main means for the public to express their attitudes and standpoints on social events. Accurately identifying the sentiment orientation of social texts is of great value for public opinions control, social stability maintenance, etc. Since the traditional sentiment recognition models only focus on mining surface semantic on comments, there are some problems such as poor classification effect and limited generalization ability. Aiming at these problems, we propose a deep sentiment semantic recognition model based on BERT-BiGRU multi-model ensemble learning. Firstly, the contextual semantic feature representation of the comment text is obtained through the BERT pre-training model, and then BiGRU is combined to extract the deep nonlinear feature vectors to achieve the optimal sentiment recognition results under the single model. Next, in order to make the effect of the model stable and achieve balanced performance in many aspects, several differentiated emotion classifiers with excellent performance are trained based on the BERT series pre-training model. Finally, the ensemble learning method of data disturbance and voting strategy is used to achieve the full integration of the deep features of each model. The experimental results on two public datasets (COV19 and ChnSenti) show that the model proposed has better sentiment recognition effects than other traditional models.

**Key words:** sentiment recognition; BERT pre-training model; bidirectional GRU; ensemble learning; deep feature

收稿日期:2022-07-28

修回日期:2022-11-30

基金项目:国家自然科学基金资助项目(61803149);湖北省重点专项(2022BAA044);湖北省教育厅科学技术研究计划重点项目(D20201006)

作者简介:游 兰(1978-),女,教授,博士,CCF 会员(H8967M),研究方向为时空大数据、自然语言处理(NLP)及社会计算等;通信作者:崔海波(1979-),男,副教授,硕士,研究方向为软件工程、系统集成、大数据与人工智能。

## 0 引言

情感识别是从机器角度感知、理解人类通过文本、图像等介质所表达的情感意图,是人工智能的关键技术之一。当前,社交网站是人们关注时事和分享个人观点的重要平台,每天会产生海量的非结构化文本评论,这些评论通常带有发表者的主观情感意图。社交评论的情感识别对于舆情管控<sup>[1]</sup>、商业营销<sup>[2-3]</sup>、社会治理<sup>[4]</sup>等有极为重要的意义,是近年自然语言处理领域的研究热点之一。

社交网络文本具有情感语义丰富、文本长度不一等典型特征,如何从长短不一的文本中判断情感极性是当前情感识别系统迫切需要解决的关键问题<sup>[5]</sup>。传统的 Word2Vec<sup>[6]</sup>或 GloVe<sup>[7]</sup>词向量预训练模型在一定程度上能够学习到词语的上下文信息,但存在不同语境下同一词语表达了相同语义的模型语义偏差。例如,“这台车的性价比真高”和“这台车的油耗真高”,其中“真高”在前者表达积极评论,在后者中表达消极评论,该方法难以识别。采用传统卷积神经网络模型(Convolution Neural Network, CNN)<sup>[8]</sup>进行文本分类任务时,能够有效提取词语表示中的局部特征,但忽略了长距离上下文之间的语义关联性。该文使用基于 Transformer 的双向编码表征模型(Bidirectional Encoder Representations from Transformers, BERT)<sup>[9]</sup>,通过动态调整词向量语义特征,并采用双向门控循环单元(Bidirectional Gated Recurrent Unit, BiGRU)<sup>[10]</sup>进行语义编码,以增强文本的语义表达,从而挖掘网络文本中更深层次的长距离上下文情感语义信息。

如今,多数模型采用机器学习或者深度学习<sup>[11-13]</sup>的方法对文本情感进行预测,取得了较大进展。然而多数模型均为单一模型,其具有随机性。针对现有研究对文本的深层情感语义特征的提取不足以及泛化能力有限等问题,该文提出了基于 BERT-BiGRU 多模集成学习的深层情感语义识别模型。主要贡献如下:

(1)采用 BERT 预训练模型代替以往传统模型的词嵌入层,获得评论文本的隐含语义词向量表示,再通过 BiGRU 提取上下文的深层语义特征,提高模型对评论文本情感语义的提取能力;

(2)通过数据扰动和投票策略的集成学习,融合多个优异且具有差异性情感识别模型,得到一个稳定且性能均衡的情感分类器,提高了模型的泛化能力;

(3)在公共数据集上进行的多组对比实验表明,提出的基于 BERT-BiGRU 多模集成学习的深层情感语义识别方法情感分类效果更好。

## 1 相关工作

情感识别相关研究已积累了不少的成果。早期,

大多数情感识别基于语义规则的情感词典或传统机器学习方式,通过计算情感词的权重得出情感倾向<sup>[14]</sup>。但情感识别的质量高度依赖于情感词典和判断规则的构建,且大多数情感词典需要人工构造。随后,传统机器学习方法被使用。Pang 等人<sup>[15]</sup>使用 SVM、朴素贝叶斯和最大熵等算法来分析电影评论情感倾向,实验表明文本特征与 SVM 算法组合的效果更佳。Hu 等人<sup>[16]</sup>提取评论中的情感词,根据句子中占主导的情感极性得到句子的情感倾向。传统机器学习的准确率较情感词典的方法有所提升,但其需要借助高质量的特征构造和专业领域知识。

神经网络方法近年来在自然语言处理领域越来越受学者们的重视。性能表现较好的情感识别模型大多以循环神经网络(Recurrent Neural Networks, RNN)、卷积神经网络或结合多个神经网络<sup>[17,18]</sup>为基础构建。如 Kim 等人<sup>[8]</sup>提出 TextCNN 模型,将训练好的词向量送入 CNN 中实现文本情感分类;陈珂等人<sup>[19]</sup>将不同特征信息结合形成不同的网络通道,作为卷积神经网络的输入,学习到句子更多的隐藏特征信息,但卷积神经网络在长距离文本建模方面能力受限。Socher 等人<sup>[20]</sup>提出利用递归神经网络来构建情感树,在二元化的句法树结构上进行语义合成,提高了短文本情感识别的准确率;Wang 等人<sup>[21]</sup>利用长短期记忆(Long-Short Term Memory, LSTM)网络预测文本情感极性。这些基于 RNN 及其变体 LSTM 模型能够捕捉到较长距离的依赖关系,但仍然存在无法编码从后到前的问题。针对该问题,Zhang 等人<sup>[10]</sup>提出一种基于双向递归神经网络的分层多输入输出模型,该模型采用两个 GRU 来获取词性和句子的词向量表示,加快了多标签情感识别的计算效率。李磊等人<sup>[22]</sup>将对象信息与文本信息进行融合,利用注意力机制强化的 Bi-LSTM 模型得到评论文本的情感分类结果。

上述文献所使用的语言模型均为 Word2Vec<sup>[6]</sup>或 GloVe<sup>[7]</sup>,它们均属于静态词向量,集中于获得词语浅层特征表示,无法进一步解决相同词语在不同的场景下多义性的问题。Devlin 等人<sup>[9]</sup>提出 BERT 预训练模型,使用深度双向 Transformer<sup>[23]</sup>模型,可以随时根据上下文信息动态调整词向量,解决了传统语言模型存在的一词多义问题。曹蕊等人<sup>[24]</sup>设计并研发了基于 BERT-BiGRU 多模集成的食品安全舆情分析系统,该系统将情感分析后的结果在地图可视化、热力图等多个模块进行可视化展示。

此外,训练好的情感分类模型想具有良好的适用性,则需要强大的泛化能力。在以往的模型中,Wan 等人<sup>[25]</sup>基于朴素贝叶斯、SVM、C4.5 决策树和随机森林等多种分类模型集成在一起,涵盖不同的情感特征,

克服了单成员模型处理结果的不足。Deriu 等人<sup>[26]</sup>训练两种不同结构的卷积神经网络作为基础分类器,再将预测结果送入到随机森林算法,其性能优于其他情感分类方法。

在现有研究基础上,针对传统语言模型的一词多义和特征抽取能力较弱的问题,该文提出 BERT-BiGRU 模型。首先,通过 BERT 预训练模型表达文本的隐含语义,再采用比双向 LSTM 结构更简单、运算速度更快的 BiGRU 模型,挖掘文本评论间的深层语义信息。考虑到单模型容易出现泛化能力不足的问题,采用交叉验证的方式训练数据集,然后以投票策略结合多个 BERT 预训练模型组成的基分类器,使集成模型获得最佳的预测结果和泛化性能。

## 2 基于 BERT-BiGRU 多模集成学习的深层情感语义识别

基于 BERT-BiGRU 多模集成学习深层情感语义的算法模型流程如图 1 所示。第一步是对原始数据集进行预处理,去除噪声数据;第二步是构建单个情感识别模型,将 BERT 预训练语言模型与 BiGRU 拼接得到文本分类模型;第三步,为了增强情感识别模型的通用性,构建集成情感识别模型。分别通过数据扰动的方式和基于 BERT 系列预训练模型得到多个情感基分类器,将步骤二的分类结果与集成模块输出的结果进行投票决策。

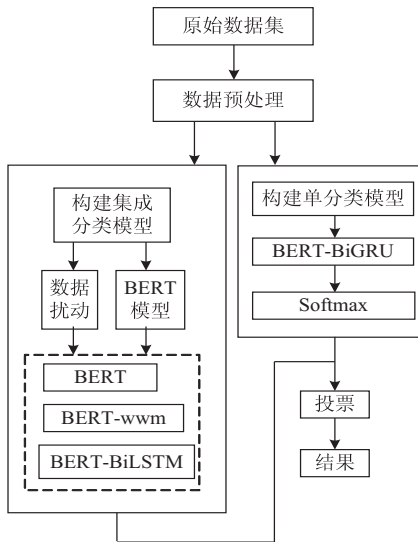


图 1 算法流程

图 1 中的 BERT 是由中文维基百科的数据进行大规模训练得到的最终模型。BERT-wwm-ext 相对于 BERT 使用的是中文维基百科数据和通用(百科、新闻、问答)数据,同时增加了训练步数。BERT-BiLSTM 使用 BERT 得到每条文本的特征向量,将其作为双向 LSTM 的输入,挖掘网络文本中更深层次的长距离上

下文情感语义信息。

### 2.1 基于 BERT-BiGRU 的文本分类模型

该文使用了 BERT 系列的预训练语言模型获取输入文本的语义表示  $\{d_1, d_2, \dots, d_n\}$ , 其中  $n$  表示模型的输入最大长度。如果文本长度小于  $n$ , 则将通过 BERT 模型得到的小于  $n$  的输出文本语义表示进行填充, 填充多个  $m$  维零向量直至输出序列长度为  $n$ ; 如果文本长度大于  $n$ , 则只输出前  $n$  长度的文本语义表示。

为了进一步得到文本间的内在联系, 将经过 BERT 模型得到的语义表示作为 BiGRU 模型的输入。通过正向 GRU 来获取到所有过去时间序列中的完整上下文信息, 再通过反向 GRU 获取所有未来时间序列的完整上下文, 计算公式如下:

$$\vec{h}_t^{(i)} = f(\vec{w}^{(i)} h_t^{(i-1)} + \vec{v}^{(i)} \vec{h}_{t-1}^{(i)} + \vec{b}^{(i)}) \quad (1)$$

$$\overleftarrow{h}_t^{(i)} = f(\overleftarrow{w}^{(i)} h_t^{(i-1)} + \overleftarrow{v}^{(i)} \overleftarrow{h}_{t-1}^{(i)} + \overleftarrow{b}^{(i)}) \quad (2)$$

其中,  $w$  为连接两层的权重,  $b$  为偏置向量,  $f$  为激活函数,  $\vec{h}_t$  和  $\overleftarrow{h}_t$  分别为正向 GRU 和负向 GRU 输出。

BiGRU 在正负向的隐藏层上对输入表示  $d_n$  进行计算, 得到  $d_n$  隐藏状态  $h_t$ , 将正负 GRU 采用拼接策略:

$$h_t = \vec{h}_t \oplus \overleftarrow{h}_t \quad (3)$$

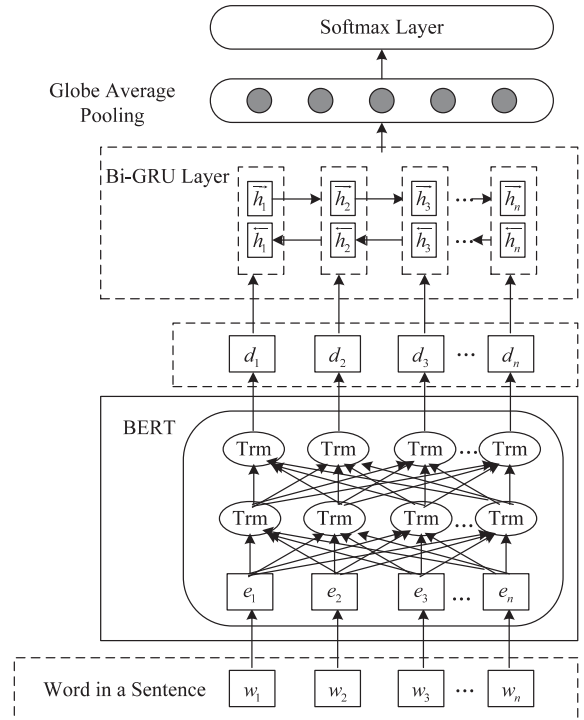


图 2 BERT-BiGRU 模型结构

得到 BiGRU 的隐藏状态, 计算公式如下:

$$H = \{h_1, h_2, \dots, h_d\} \quad (4)$$

最后, 使用全局平均池化来代替全连接层, 得到一维的特征向量, 送入 softmax 函数得到评论文本的情感



类别,如公式(5)所示:

$$\text{TEXT}_c = \text{softmax}(W_l H + b_l) \quad (5)$$

其中,  $W_l$  表示全局平均池化层的权重参数,  $b_l$  表示偏置值,最后得到 BERT-BiGRU 模型的输出  $\text{TEXT}_C$ 。模型结构如图 2 所示。

## 2.2 基于数据扰动的集成方式

由机器学习算法的误差分析<sup>[27]</sup>可知,要使训练后的模型保持良好的泛化能力,则需要降低模型的泛化误差。如公式(6)所示,泛化误差由偏差、方差和噪声组成。

$$E(f; D) = \text{bias}^2(\chi) + \text{var}(\chi) + \varepsilon^2 \quad (6)$$

其中,  $\text{bias}$  表示偏差,即期望输出与真实输出的差别。 $\text{var}$  表示方差,即同同样本下,不同批次训练集产生的方差。 $\varepsilon$  表示噪音,刻画学习问题本身的难度,属于模型的固有误差,该文不予考虑。

随着模型复杂度的上升,偏差逐渐减小,方差增大,如图 3 所示。为了寻求最优平衡点,本节采用数据扰动的方法,改变交叉验证中每一折的数据划分方式,即  $K$  折交叉验证来控制偏差的大小。

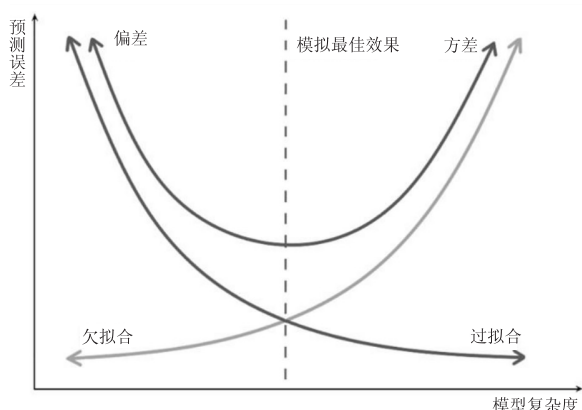


图 3 模型复杂度影响因素分析

首先,对训练集采用  $K$  折交叉验证的方式对原始数据集分组,将训练集划分为互不相交的  $K$  份数据,取其中一个子集作为验证集,剩余  $K-1$  份作为训练集,迭代  $K$  次。然后,取  $K$  组测试集预测结果的平均值作为分类器最终预测标签。

## 2.3 基于多模型投票的集成方式

集成学习<sup>[28]</sup>的思想是构建多个基分类器,将其组合成性能更优的强分类器。该方法的优点在于即使某个基分类器的结果出现错误,其他的基分类器能够及时改正。同时在构建基分类器时,需要注意每个基分类器之间要具有一定差异性。使得在各自方法上的结果达到最优,学习能力越强,关联性越低。

该文通过结合策略对基分类器的结果进行融合,使通过集成学习后的模型预测效果达到最好。首先统计所有分类器的输出类别,在生成情感识别的类别概

率分布基础上采用多数投票策略进行决策。给定所有个体分类器的分类结果同样的权重,各个基分类器均只能投一票,采用少数服从多数原则,最后得票数最高的类别将作为该条评论的最终预测结果。投票公式如下所示:

$$V(x) = \text{Max}_{j=1}^n \sum_{i=1}^T c_{i,j} \quad (7)$$

其中,  $n$  和  $T$  分别表示情感分类的类别数和基分类器的个数,  $c_{i,j}$  表示基分类器  $i$  在测试集  $x$  的预测类别为  $j$ ,  $\sum_{i=1}^T c_{i,j}$  表示测试集  $x$  在所有基分类器的预测类别结果为  $j$  的总票数,取票数最多的类别作为该预测样本  $x$  的最后类别结果。

## 3 实验与结果

为了验证模型的通用性和有效性,使用公开的新冠疫情网民情绪识别数据集 COV19,由 CCF 举办的 2020 科学防疫大数据公益挑战赛权威发布,以及中科院谭松波博士收集整理的酒店评论语料 ChnSentiCorp。实验结果对比则采用近年表现较好的情感识别模型,针对各自分类效果进行指标评测。

### 3.1 实验数据集

实验数据集一是疫情期间网民情绪识别数据集,包括新浪微博中与“新冠肺炎”相关的 230 个主题关键词博文,共 10 万条带有积极、中性和消极三种情感极性的微博数据。数据集二是与酒店评论相关的语料,共 1 万多条带有积极和消极情感极性的评论数据。

该文去除了 2 个数据集中空数据、重复数据以及文字繁简替换,清洗得到最后实验数据。实验数据具体情况如表 1 所示。

表 1 实验使用数据统计

$\lambda$	Positive	Neutral	Negative	Total
COV19-train	22 874	51 761	15 285	89 920
COV19-test	5 858	2 518	1 617	9 993
ChnSenti-train	5 178	/	5 168	10 346
ChnSenti-test	608	/	592	1 200

### 3.2 实验设置

选取的最大文本长度标准为能够覆盖 75% 左右的文本内容,不足最大长度的文本内容使用零向量进行填充,大于最大长度的文本内容进行截断。

采用的 BERT 词向量预训练模型为中文版的 BERT-Base 和 BERT-www-ext,均由 12 层 transformers 组成,其中每层 transformer 的隐藏层数为 768,包含 12 个注意力头,接入 GRU 单元数为 128。学习率设置为  $2e-5$ ,采用 dropout 策略,具体比率设置为 0.15。

### 3.3 实验评测指标

实验主要是对文本进行积极、中性、消极三种情感的分类,采用的主要评价指标有精准率(Precision)、召回率(Recall)和 F1 值。精准率(Precision)又称为查准率,是指预测结果属于某一类的个体,实际属于该类的比例。召回率(Recall)又称为查全率,是指被正确预测为某个类别的个体数量与数据集中该类别个体总量的比例。F1 值是综合准确率和召回率的调和平均数,反映整体性能的评价指标。

具体计算公式如下:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (8)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (9)$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

其中,TP(True Positive)是真正例,即预测为正样本,实际也为正样本;FP(False Positive)是假正例,即预测为正样本,实际为负样本;FN(False Negative)是假负例,即预测为负样本,实际为正样本。

表2 多个模型在 COV19 上的实验结果

model	Precision	Recall	F1
TextCNN	0.623	0.652	0.634
BiGRU	0.673	0.651	0.661
BiGRU-ATT	0.662	0.668	0.664
BERT	0.716	0.719	0.717
BERT-BiGRU	0.721	0.737	0.728

表2展示了文中模型和对比模型在 COV19 数据集上的精准率、召回率和 F1 值对比结果。可以看出, BERT 模型的 F1 值达到 0.717 的分类准确率,对比 CNN 模型的 0.634、BiGRU 模型的 0.661,分别提升了 8.3 和 5.6 百分点,验证了 BERT 预训练模型明显优于基于传统的词向量训练模型。其中,基于 Word2Vec 的词向量模型各项指标较低,主要是受中文表达一词多义的影响。而 BERT 模型会根据上下文的含义动态变化词向量,更能准确反映出句子的语义信息。可见 BERT 模型比在 BiGRU 加入单个注意力机制的 BiGRU-ATT 模型提高了 5.3 百分点。在相同的 Word2Vec 词向量模型下, BiGRU 模型比 CNN 模型高了 2.7 百分点,可以看出,使用 BiGRU 比 CNN 提取文本特征的效果更佳。故 BERT 后接入 BiGRU 的分类效果取得了最佳的分类效果,说明在 BERT 的输出中加入 BiGRU 能够提取句子中更深层次的情感特征,从而提升分类准确性。

#### 3.4.2 集成策略对实验的影响

为了验证集成思想在情感识别领域的通用性,实

### 3.4 实验结果及分析

#### 3.4.1 单一情感识别模型的实验对比

该文采用两组对比实验,实验一是 BERT-BiGRU 与传统深度学习模型的对比。

(1)TextCNN<sup>[8]</sup>。使用 Word2Vec 对输入文本编码,送入卷积神经网络,利用多个不同大小的卷积核提取文本特征,最后通过全连接层进行分类。

(2)BiGRU<sup>[23]</sup>。使用 Word2Vec 训练词向量,送入双向 GRU,将前后的特征信息进行关联,挖掘出各个情感特征之间的潜在联系,获得评论文本的情感倾向性。

(3)BiGRU-ATT<sup>[10]</sup>。在使用 BiGRU 抽取文本特征的基础上,引用注意力机制,捕获评论文本中的每个词对情感语义贡献程度大小,进行加权计算。

(4)BERT<sup>[9]</sup>。使用深度双向 Transformer 模型,获取评论文本的上下文语义特征表示。

(5)BERT-BiGRU。首先,通过 BERT 预训练模型表达文本的隐含语义,再采用 BiGRU 模型挖掘文本评论间的深层语义信息,得到最终的情感识别结果。

验二将 BERT 预训练模型接入不同的网络,利用其训练语料和训练步数的不同,使用不同的训练批次和训练方式来达到差异性效果。最后使用投票策略达到集成效果,实验分别在情感二分类和三分类的数据集上进行。模型说明如下:

(1)BERT。使用 BERT-base 预训练模型提取文本特征,训练批次分别设为 32 和 16,最大文本截取长度分别为 128 和 140,经过 3 轮训练分别得到模型 1(M1)和模型 2(M2)。

(2)预训练模型使用 BERT-www-ext,训练批次为 32,最大文本截取长度为 128,训练得到模型 3(M3)。

(3)BERT-BiLSTM。使用 BERT-base 预训练模型得到每条文本的特征,经过双向 LSTM 挖掘网络文本中更深层次的长距离上下文情感语义信息。再进行 5 折交叉验证,训练批次设为 16,最大文本截取长度为 140,训练得到模型 4(M4)。

(4)BERT-BiLSTM。训练方法同(3),批次设为 48,训练得到模型 5(M5)。

(5)BERT-BiGRU。使用 BERT-base 预训练模型

得到每条文本的特征,经过 BiGRU 的输出,进行 5 折交叉验证,训练批次分别设为 16、48 和 64,最大文本截取长度别设为 140、140、128,训练得到模型 6(M6)、模型 7(M7)和模型 8(M8)。

表 3 基分类器在 COV19 上的实验结果

model	超参数集	Precision	Recall	F1
BERT(M1)	batchsize:32,maxlen:128	0.716	0.719	0.717
BERT(M2)	batchsize:16,maxlen:140	0.724	0.731	0.724
BERT-wwm-ext(M3)	batchsize:32,maxlen:128	0.715	0.726	0.720
BERT-BiLSTM(M4)	batchsize:16,maxlen:140	0.716	0.739	0.725
BERT-BiLSTM(M5)	batchsize:48,maxlen:140	0.721	0.734	0.727
BERT-BiGRU(M6)	batchsize:16,maxlen:140	0.721	0.737	0.728
BERT-BiGRU(M7)	batchsize:48,maxlen:140	0.724	0.736	0.729
BERT-BiGRU(M8)	batchsize:64,maxlen:128	0.719	0.733	0.726

由表 3 可见,同样的网络结构,使用不同的超参数集进行训练,所表现的分类结果在不同的评价指标下各有优势。比如 BERT 模型之间的差距达到了 0.01,而 BERT-BiLSTM 和 BERT-BiGRU 使用不同的超参数,差异性在 0.002~0.004 之间。对比于传统的数据集单次划分,该文加入的 5 折交叉验证,多次划分训练数据集使数据集间互不交叉,大幅度降低了随机划分数据集带来的偶然性,增强了模型的稳定性。在多数评价指标下,BERT-BiLSTM、BERT-BiGRU 在批次 48 和最大文本截取长度 140 上分别达到了同类模型最优。

为了更直观地得到各个分类器之间的差异程度,借助协方差来度量模型间的线性相关程度,对模型的预测结果进行归一化处理,得到皮尔逊相关系数作为相关度指标,展示以上 8 种模型的相关度分析。

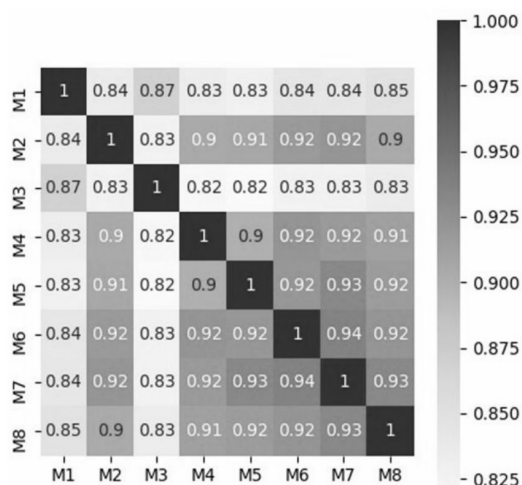


图 4 基于 Pearson 的模型预测结果相关度分析

由图 4 可知,各个模型的皮尔逊相关系数均在 0.825 以上。图中颜色越浅,代表模型之间的相关性越弱。可以得知,M1、M2 和 M3 与其他模型的相关性

实验将各个基分类器在 COV19 数据集上单独预测的分类效果进行比较分析。考虑到相同的分类器对同一个数据集的结果可能存在差别,对各个模型重复运行 3 次,取结果的平均值作为最后的评价指标。

较弱,这 3 种模型都是基于 BERT 预训练模型进行训练。以上做法是通过改变模型的训练参数生成有差异性的基分类器。而后 5 种模型是在 BERT 预训练模型的基础上融合了其他神经网络模型,主要是改变了模型的网络结构及训练参数,使各个模型内部的分类原理不同及提取特征的能力也各不相同。

综上所述,依据集成学习的两个前提条件,一是基分类器的分类效果要达到一定的优越性,如果基分类器的分类错误率过高将会导致集成学习的准确率降低;二是基分类器之间要具有差异性,即预测结果需要多样性,如果差异性过小,那集成结果基本无变化。因此,由表 3 和图 4 得出,选取 M2、M3、M5、M7 作集成学习的基分类器,对测试集的预测结果使用多数投票法得到最终的分类结果,集成结果如表 4 所示。

表 4 COV19 的最终集成实验结果

	Precision	Recall	F1
集成结果	0.728	0.738	0.732

集成后的 F1 值达到 0.732,相比最优的单模型 BERT-BiGRU 提升了 0.3 百分点。为了进一步验证所提模型的泛化能力,将以上模型在 ChnSenti 语料上进行测试,结果如表 5 所示。

表 5 集成学习在 ChnSenti 上的实验结果

model	Precision	Recall	F1
BERT	0.936	0.942	0.938
BERT-wwm-ext	0.931	0.919	0.924
BERT-BiLSTM	0.953	0.937	0.944
BERT-BiGRU	0.956	0.935	0.945
集成结果	0.964	0.926	0.946

表 5 展示了 4 个单模型和集成后的模型在酒店语料上的分类效果。由表可见,BERT-BiGRU 在 F1 评价



指标上都超越了其他三种模型,分别提高了 1.4、2.8 和 0.8 百分点。证明了 BERT-BiGRU 比其他模型能够挖掘更深层次的语义特征,也验证了提出在 BERT 输出表示后接入 BiGRU 模型的方法具有有效性和优越性。各模型的 F1 值均在 0.92 以上,表明在二分类上各单模型表现均较为优异,导致模型间的差异性不够强,集成后的 F1 值提升不够明显,仅比最佳模型 BERT-BiGRU 提高了 0.1 百分点。由此可见,所使用的集成学习方法的正确性,需要集成表现优异且差距性较大的模型,才能获得明显的提升效果。

#### 4 结束语

围绕社交网络文本的情感识别研究热点,提出了基于 BERT-BiGRU 多模集成学习的深层情感语义识别模型。首先,该模型采用 BERT 预训练模型代替以往传统模型的词嵌入层,获得评论文本的隐含语义词向量表示,再通过 BiGRU 提取上下文的深层语义特征,解决了传统语言模型存在的不同语境下多义词和深层情感语义抽取能力较弱的问题,实现了单模型下最优的情感识别效果。为了提高模型的泛化能力,该模型在综合分析模型的方差和偏差的基础上,将集成学习的思想应用到情感识别上,观测不同模型在不同参数和数据集上的表现,采用交叉验证的方式训练数据集,然后以投票策略结合多个 BERT 预训练模型组成的基分类器,使得它们具有互相纠错的能力,从而获得更优的集成结果。最后,通过在三分类和二分类语料上分别设计实验,展示出 BERT-BiGRU 模型在多个评价指标上优于现有大多数情感识别模型。

在未来研究中,还需考虑使用更丰富数据来挖掘情感语义,比如表情符号,图片或视频语义来辅助挖掘深层情感语义。

#### 参考文献:

- [1] 李光敏,张行文,张 磊,等.面向网络舆情的评论文本情感分析研究[J].情报杂志,2014,33(5):157-160.
- [2] 吴 璠,王中卿,周夏冰,等.基于用户和产品表示的情感分析和评论质量检测联合模型[J].软件学报,2020,31(8):2492-2507.
- [3] 彭 敏,席俊杰,代心媛,等.基于情感分析和 LDA 主题模型的协同过滤推荐算法[J].中文信息学报,2017,31(2):194-203.
- [4] 刘春年,肖 迪.情绪性在线评论下舆情演化与焦点事件治理研究[J].现代情报,2020,40(9):133-143.
- [5] 赵传君,王素格,李德玉.跨领域文本情感分类研究进展[J].软件学报,2020,31(6):1723-1746.
- [6] LE Q, MIKOLOV T. Distributed representations of sentences and documents [C]//Proceedings of the 31st international conference on machine learning. Beijing: JMLR. org, 2014: 1188-1196.
- [7] 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). Doha: Association for Computational Linguistics, 2014: 1532-1543.
- [8] KIM Y. Convolutional neural networks for sentence classification [C]//Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). Doha: Association for Computational Linguistics, 2014: 1746-1751.
- [9] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [C]//Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers). Minneapolis: Association for Computational Linguistics, 2019: 4171-4186.
- [10] ZHANG L, ZHOU Y, DUAN X, et al. A hierarchical multi-input and output bi-GRU model for sentiment analysis on customer reviews [J]. IOP Conference Series: Materials Science and Engineering, 2018, 322(6): 062007.
- [11] PORIA S, CAMBRIA E, GELBUKH A. Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis [C]//Proceedings of the 2015 conference on empirical methods in natural language processing. Lisbon: Association for Computational Linguistics, 2015: 2539-2544.
- [12] 黄发良,于 戈,张继连,等.基于社交关系的微博主题情感挖掘[J].软件学报,2017,28(3):694-707.
- [13] 梁 斌,刘 全,徐 进,等.基于多注意力卷积神经网络的特定目标情感分析[J].计算机研究与发展,2017,54(8):1724-1735.
- [14] DING X, LIU B, YU P S. A holistic lexicon-based approach to opinion mining [C]//Proceedings of the 2008 international conference on web search and data mining. New York: Association for Computing Machinery, 2008: 231-240.
- [15] PANG B, LEE L, VAITHYANATHAN S. Thumbs up? Sentiment classification using machine learning techniques [J]. arXiv:cs/0205070, 2002.
- [16] HU M, LIU B. Mining and summarizing customer reviews [C]//Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining. Seattle: Association for Computing Machinery, 2004: 168-177.
- [17] PATTANAYAK S. Natural language processing using recurrent neural networks [M]//Pro deep learning with TensorFlow. Berkeley: Apress, 2017: 223-278.
- [18] IWASAKI R, HASEGAWA T, MORI N, et al. Relaxation method of convolutional neural networks for natural language processing [C]//International symposium on distributed computing and artificial intelligence. [s. l.]: Springer, 2018: 188-

- 195.
- [19] 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.
- [20] SOCHER R, PERELYGIN A, WU J, et al. Recursive deep models for semantic compositionality over a sentiment treebank[C]//Proceedings of the 2013 conference on empirical methods in natural language processing. Seattle: Association for Computational Linguistics, 2013:1631-1642.
- [21] WANG X, LIU Y, SUN C J, et al. Predicting polarities of tweets by composing word embeddings with long short-term memory[C]//Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (volume 1: long papers). Beijing: Association for Computational Linguistics, 2015:1343-1353.
- [22] 李磊,吴旭辉,刘继.融合关键对象识别与深层自注意力的BiLSTM情感分析模型[J].小型微型计算机系统, 2021,42(3):504-509.
- [23] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st international conference on neural information processing systems. New York: Curran Associates Inc., 2017:6000-6010.
- [24] 曹蕊,周毓奇.基于BERT-BiGRU多模集成的食品安全舆情分析系统[J].信息与电脑:理论版,2022,34(7):94-97.
- [25] WAN Y, GAO Q. An ensemble sentiment classification system of twitter data for airline services analysis[C]//2015 IEEE international conference on data mining workshop (ICDMW). Atlantic City: IEEE, 2015:1318-1325.
- [26] DERIU J M, GONZENBACH M, UZDILLI F, et al. Swisschese at semeval-2016 task 4: sentiment classification using an ensemble of convolutional neural networks with distant supervision[C]//Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016). San Diego: Association for Computational Linguistics, 2016:1124-1128.
- [27] 周志华.机器学习[M].北京:清华大学出版社,2016.
- [28] ZHOU Z H. Ensemble methods: foundations and algorithms[M]. UK: CRC Press, 2012.
- .....
- (上接第 158 页)
- natural language processing. Brussels: EMNLP, 2018:1-7.
- [26] TRAN T, KAVULURU R. Neural metric learning for fast end-to-end relation extraction[J]. arXiv:1905.07458, 2019.
- [27] NGUYEN D Q, VERSPOOR K. End-to-end neural relation extraction using deep biaffine attention[C]//European conference on information retrieval. Cologne: ECIR, 2019:729-738.
- [28] BEKOULIS G, DELEU J, DEMEESTER T, et al. Joint entity recognition and relation extraction as a multi-head selection problem[J]. Expert Systems with Applications, 2018, 114:34-45.
- [29] EBERTS M, ULGES A. Span-based joint entity and relation extraction with transformerpre-training[M]//ECAI 2020. [s. l.]: IOS Press, 2020:2006-2013.