

基于用户性格和语义-结构特征的 文本评论情感分类方法

王友卫¹, 刘瑞¹, 凤丽洲²

(1. 中央财经大学信息学院, 北京 100081; 2. 天津财经大学统计学院, 天津 300222)

摘要: 由于传统文本评论情感分类方法通常忽略用户性格对于情感分类结果的影响, 提出一种基于用户性格和语义-结构特征的文本评论情感分类方法 (User Personality and Semantic-structural Features based Sentiment Classification Method for Text Comments, BF_BiGAC). 依据大五人格模型能够有效表达用户性格的优势, 通过计算不同维度性格得分, 从评论文本中获取用户性格特征. 利用双向门控循环单元 (Bidirectional Gated Recurrent Unit, BiGRU) 和卷积神经网络 (Convolutional Neural Network, CNN) 可以有效提取文本上下文语义特征和局部结构特征的优势, 提出一种基于 BiGRU、CNN 和双层注意力机制的文本语义-结构特征获取方法. 为区分不同类型特征的影响, 引入混合注意力层实现对用户性格特征和文本语义-结构特征的有效融合, 以此获得最终的文本向量表达. 在 IMDB、Yelp-2、Yelp-5 及 Ekman 四个评论数据集上的对比实验结果表明, BF_BiGAC 在分类准确率 (Accuracy) 和加权 macro F_1 值 (F_w) 上均获得较好表现, 相对于拼接 BiGRU、CNN 的情感分类方法 (Sentiment Classification Method Concatenating BiGRU and CNN, BiGRU_CNN) 在 Accuracy 值上分别提升 0.020、0.012、0.017 及 0.011, 相对于拼接 CNN、BiGRU 的情感分类方法 (Sentiment Classification Method Concatenating CNN and BiGRU, ConvBiLSTM) F_w 值上分别提升 0.022、0.013、0.028 及 0.023; 相对于预训练模型 BERT 和 RoBERTa, BF_BiGAC 在保证分类精度的情况下获得了较高的运行效率.

关键词: 情感分类; 大五人格模型; 双向门控循环单元; 卷积神经网络; 注意力机制

基金项目: 国家自然科学基金 (No.61906220); 教育部人文社科资助项目 (No.19YJCZH178); 国家社会科学基金 (No.18CTJ008); 中央财经大学新兴交叉学科建设项目

中图分类号: TP 391 **文献标识码:** A **文章编号:** 0372-2112(2024)05-1657-13

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.12263/DZXB.20220645

A Sentiment Classification Method for Text Comments Based on User Personality and Semantic-Structural Features

WANG You-wei¹, LIU Rui¹, FENG Li-zhou²

(1. School of Information, Central University of Finance and Economics, Beijing 100081, China;

2. School of Statistics, Tianjin University of Finance and Economics, Tianjin 300222, China)

Abstract: Since the traditional sentiment classification methods for text comments usually ignore the influence of user personality on sentiment classification results, a sentiment classification method for text comments based on user personality and semantic-structural features is proposed. According to the advantage of Big Five personality model on effectively expressing the user personality, the user personality feature is obtained from the comment texts by calculating the personality scores from different dimensions. Moreover, the advantages of bidirectional gated recurrent unit (BiGRU) and convolutional neural network (CNN) on effectively extracting the contextual semantic features and the local structural features are taken, and a new text semantic-structural feature acquisition method based on BiGRU, CNN and two-layer attention mechanism is proposed. Finally, in order to distinguish the influence of the features with different types, the hybrid attention layer is introduced to obtain the final text vector representation by integrating the user personality feature and the textural semantic-structural feature effectively. The experimental results on the datasets of IMDB, Yelp-2, Yelp-5 and Ekman show that BF_BiGAC achieves good performance when the measurements of Accuracy and weighted macro F_1 (F_w) are used. Specifi-

cally, it achieves the improvements of 0.020, 0.012, 0.017 and 0.011 compared to sentiment classification method concatenating BiGRU and CNN (BiGRU_CNN) on accuracy, and achieves the improvements of 0.022, 0.013, 0.028 and 0.023 compared to sentiment classification method concatenating CNN and BiGRU (ConvBiLSTM) on F_w . Moreover, when comparing with the pre-trained models of BERT and RoBERTa, BF_BiGAC achieves higher executing efficiency while ensuring the classification accuracy.

Key words: sentiment classification; Big Five personality model; bidirectional gated recurrent unit; convolutional neural network; attention mechanism

Foundation Item(s): National Natural Science Foundation of China (No.61906220); Ministry of Education of Humanities and Social Science Project (No.19YJCZH178); National Social Science Foundation of China (No.18CTJ008); Emerging Interdisciplinary Project of CUFE

1 引言

当前,人们可以借助各类社交点评软件来表达自己对于事物的意见和看法,因此产生了大量的情感倾向信息.基于评论数据的情感分类可以度量出用户对商品的情感倾向,不仅可以实现个性化商品推荐,还可以依据用户情感变化实现对待评价商品的改进,因此具有重要的社会价值和研究意义.

目前,情感分类可以分为基于情感词典的方法、基于传统机器学习的方法和基于深度学习的方法.典型的情感词典包括 WordNet、SentiWordNet等.栗雨晴等人^[1]提出一种基于双语词典的多类别情感分类方法,通过构建双语多类情感词典对博文进行语义倾向性分析,以便更准确有效地捕捉群体意见,及时发现社会舆论倾向.Dashtipour等人^[2]提出了一种用惯用语来扩展波斯语情感词典的方法来进行情感分析,该方法汇总了一千多种习语的表达方式以及极性来对公开的情感词典进行扩展,并提出了一种准确分类波斯语文本的算法.曾雪强等人^[3]基于文本中的情感词蕴含着大量情感信息的特点,提出基于情感轮和情感词典的情感分布标记增强方法.作者基于情绪的心理距离为句子的真实情绪标签和情感词的情绪标签分别生成离散高斯分布,然后通过分布的叠加将两种信息综合为统一的情感分布,最后在7个常用的中英文文本情感数据集上验证了该方法在情绪识别任务上的有效性.

但是,情感词典在不同领域或是针对不同语言时准确性较低,通常需要根据具体的上下文语境去分析判断单词语义,而仅仅依靠情感词典中的定义往往是不准确的,因此限制了此类方法在情感分类领域中的应用.随着机器学习的发展,更多学者将机器学习算法应用于情感分类中.传统基于机器学习的情感分类方法主要划分为无监督、半监督以及有监督3种类别.自隐含狄利克雷分布(Latent Dirichlet Allocation, LDA)^[4]模型于2003年被提出后,不少学者们提出了以JST^[5]、Sentic LDA^[6]和SS-LDA^[7]为典型代表的无监督文本情感分类算法,有效地避免了传统无监督情感分类方法

依赖情感词典的缺点,能达到较好的情感分类效果^[8].半监督方法典型代表包括Self-training^[9]、直推式支持向量机(Transductive Support Vector Machine, TSVM)^[10]等.此类方法通过提取未标记文本的特征,可以使有标注数据文本稀疏的问题得到很好的解决,进而提高情感分析的准确性.有监督方法通过人工标注数据建立训练集,对训练集的数据进行特征提取和学习来构建一个分类模型,最后通过该分类模型对未知标签的测试数据进行分类预测,典型算法包括:朴素贝叶斯(Native Bayesian, NB)^[11]、支持向量机(Support Vector Machines, SVM)^[12]、逻辑回归(Logistic Regression, LR)^[13]、SentiXGboost^[14]等.

但是,传统基于机器学习的文本情感分类方法需要复杂的特征选择,此过程大多依赖人工设计,导致泛化能力较差^[15].为了使情感分类任务可以充分考虑到文本语境中上下文语义之间的关系,近年来很多学者开始转向深度学习领域进行深入研究.Jelodar等人^[16]针对新冠病毒的相关评论,利用长短期记忆(Long Short-Term Memory, LSTM)模型分析其情感极性,得到的实验结果对于疫情防控具有重大意义.Li等人^[17]将Word2 vec、双向门控循环单元(Bidirectional Gated Recurrent Unit, BiGRU)和注意力(Attention)机制结合起来建立在线餐厅评论的情感分类模型.实验结果证明该方法相对于SVM、BiGRU等方法在分类精度上获得显著提升.Gao等人^[18]提出了基于BERT的面向目标情感分类方法.该方法利用BERT获得目标词和句子的嵌入向量,通过全连接层实现目标情感极性判断.叶星鑫等人^[19]提出了一种结合ALBERT和Attention特征分割融合网络的文本情感分析模型ALBERT-AFSFN.该模型首先利用ALBERT对文本进行词向量表示;然后通过注意力特征分割融合网络将特征分割为两组,并对两组不同通道的特征进行提取和融合,以最大程度保留不同通道之间的语义关联信息.除了上述这类单一结构的神经网络模型外,当前很多学者将研究思路放在了模型结构的组合上.Tam等人^[20]提出一种基于卷积神经网络(Convolutional Neural Network, CNN)和双向

长短期记忆(Bidirectional Long Short-Term Memory, BiLSTM)的情感分类方法 ConvBiLSTM. 该方法有效结合了 CNN 和 BiLSTM 在提取局部结构特征和上下文语义特征方面的优势,但是不足之处在于无法区分不同特征的重要性;程艳等人^[15]提出一种基于 Attention 机制、多通道卷积神经网络 CNN 和 BiGRU 的神经网络模型. 该模型不仅能够通过注意力机制关注到句子中对情感极性分类重要的词语,还结合了 CNN 和 BiGRU 在捕获不同类型特征方面的优势. 但是,该模型并未细化不同大小卷积核对应的注意力权重,并且只考虑了单词的注意力权重,缺少对于卷积块注意力权重的关注. Yan 等人^[21]提出一种基于 CNN、BiGRU 和注意力机制的情感分类模型(CNN, BiGRU and Attention based Sentiment Classification Method, CNN-BiGRU-AT). 与程艳等人不同的是,该方法首先将文本划分为多个句子,利用 CNN 从每个句子中提取不同粒度的信息构建句子级特征表示;然后,将句子通过 BiGRU 进行顺序整合提取文本的上下文语义信息特征;最后,在 CNN-BiGRU 模型中加入注意力机制对模型施加不同的学习权重. 与上述算法先对文本进行 CNN 操作不同, Sindhu 等人^[22]通过 BiGRU 提取文本序列特征,再通过 CNN 提取文本的局部结构特征,最后使用 Sigmoid 函数获得最终的情感类别. 但是,该方法未区分不同单词和卷积块的影响,限制了情感分类的准确性. 此外,针对现有情感分类方法忽略用户情绪变化影响的问题, Huang 等人^[23]在传统 LSTM 中引入情绪评估模块,结合卷积层、池化层以及注意力机制提出 AEC-LSTM 情感分类方法,有效提升了情感特征表达能力和分类效果; Wen 等人^[24]利用神经网络模型对对话上下文和用户个性进行编码来获得对话系统中用户的情绪变化,在保证个性化的同时显著提升了对话情绪选择的准确性.

研究发现,目前的情感分类方法仍然存在以下问题:(1) 心理学研究表明性格相近的人通常会选择相似的词语或是语言组织方式来表达自身情感,然而目前大多数情感分类方法并没有考虑到性格这种本质化人格特征的影响,性格特征与情感特征缺乏有效的融合方式;(2) 目前基于 CNN 和 BiGRU 的情感分类方法大多基于 CNN-BiGRU 串行的神经网络结构. 该结构虽然能提取卷积块之间的语义关联信息,但丢失了卷积块内部单词的上下文语义信息,影响了情感分类的准确性. 为此,本文引入大五人格理论,提出了一种基于用户性格和语义-结构特征的文本评论情感分类方法. 主要创新点包括:(1) 为完善用户情感特征表达方式,利用大五人格模型蕴含的用户性格知识完善文本特征信息,提出基于大五人格模型的用户性格特征提取方法;(2) 融合评论文本的局部结构特征和上下文序列化特

征,提出基于 BiGRU、CNN 和双层 Attention 的文本语义-结构特征获取方法;(3) 为区分不同类型特征的影响,引入混合 Attention 层实现对用户性格和语义-结构特征的有效融合,提高情感分类的准确性.

2 相关理论

2.1 大五人格模型

大五人格模型是较为权威且应用较为广泛的一种性格模型,不论在心理学领域还是人工智能领域都具有相当高的学术研究价值. 大五人格模型将人类的性格划分为五个维度进行衡量,包含开放性(openness)、尽责性(conscientiousness)、外向性(extraversion)、宜人性(agreeableness)以及神经质(neuroticism). 开放性可以反映出一个人想象力的高低;尽责性更多地体现的是一个人的自律程度以及组织规划的贯彻能力,往往与工作成就等联系在一起;外向性则与一个人的社交沟通能力以及个人魅力值息息相关;宜人性体现的是本质的行为特征,往往与同情、协作和为人处世等词语密切相关;神经质更多的是体现一个人的情绪是否稳定,与心理疾病如焦虑、郁郁寡欢有直接关系.

通过分析大五人格挖掘出文本表达情感特点的首要步骤是得到五大维度的性格得分. 传统的性格分析任务往往采用调查问卷方法,但这通常是费时费力的. 当前利用大五人格模型来预测用户性格的方法主要采用基于闭合性词汇的方法,即在成熟的心理学词库的基础上预测人的性格. Pennebaker 等人^[25]采用语言查询与字数统计(Linguistic Inquiry and Word Count, LIWC)词典分析诸如日记、作业等多种文本中性格层面的词语,研究得出宜人性得分高的人往往喜欢使用冠词,外向性得分低或是尽责性得分低的人往往喜欢使用差距感较大的词汇等. Sumner 等人^[26]也利用了 LIWC 词典在 Facebook 上研究了五百多名用户的文本,发现开放性得分高的人往往喜欢采用引用的方式,而神经质性得分高的人往往喜欢使用简称或是缩略词. 为了便于理解,给出了 LIWC 词典中各个维度对应的部分低分词和高分词,具体如表 1 所示.

2.2 门控循环单元

门控循环单元(Gated Recurrent Unit, GRU)^[27]是一种典型的循环神经网络(Recurrent Neural Network, RNN). 如图 1 所示,GRU 采用了一个更新门取代了 LSTM 输入门和遗忘门,能够有效地获取到状态间的长距离依赖关系.

一个 GRU 网络神经元均含有 1 个记忆单元和 2 个门单元,如式(1)~(4)所示:

$$r_t = \text{Sigmoid}(x_t W_r + h_{t-1} U_r) \quad (1)$$

$$z_t = \text{Sigmoid}(x_t W_z + h_{t-1} U_z) \quad (2)$$

表1 大五人格性格词典示例

性格维度	低分词		高分词	
	词语	得分	词语	得分
Openness	but, without, exclude...	-0.056	buddy, friend, colleague...	0.048
Conscientiousness	cheek, hands, spit...	-0.077	we, us, our...	0.052
Extraversion	maybe, perhaps, guess...	-0.051	love, nice, sweet...	0.108
Agreeableness	no, not, never...	-0.136	love, nice, sweet...	0.183
Neuroticism	very, really, quickly...	-0.119	job, majors, Xerox...	0.117

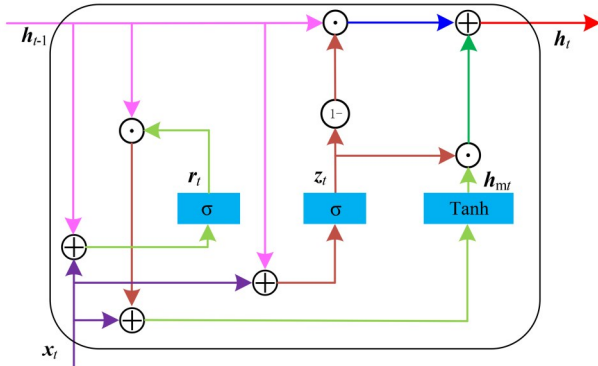


图1 GRU模型结构

$$h_{mt} = \text{Tanh}(x_t W_h + (r_t \cdot h_{t-1}) U_h) \quad (3)$$

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot h_{mt} \quad (4)$$

其中, $x_t \in \mathbb{R}^n$ (n 为向量维度) 代表的是时间为 t 时的输入, $h_t \in \mathbb{R}^n$ 代表的是时间为 t 时的隐藏层状态, $z_t \in \mathbb{R}^n$ 为更新门, $r_t \in \mathbb{R}^n$ 为重置门, $W_h \in \mathbb{R}^{n \times n}$ 、 $W_z \in \mathbb{R}^{n \times n}$ 、 $W_r \in \mathbb{R}^{n \times n}$ 、 $U_h \in \mathbb{R}^{n \times n}$ 及 $U_z \in \mathbb{R}^{n \times n}$ 均为权重矩阵, $h_{mt} \in \mathbb{R}^n$ 为 t 时刻的候选记忆单元, Sigmoid 和 Tanh 均为激活函数, \cdot 为求矩阵的哈达玛积操作. 在基于文本的情感分类中, 单向的 GRU 只能获取到单词的上文信息, 为综合考虑上下文语义信息, 本文使用双向 GRU 模型 (Bidirectional Gated Recurrent Unit, BiGRU)^[15, 17, 21] 将正反向隐含层相结合, 以此获取单词上下文语义关联关系, 提高情感分类的准确性.

3 本文方法

3.1 问题定义

区别于传统情感分类方法只考虑评论文本内容信息的特点, 本文综合考虑了用户性格特征对情感分类结果的影响. 为便于理解, 先给出本文所需的符号定义, 如表 2 所示:

针对向量 a , 其第 i 维分量表示为 a_i ; 针对矩阵 A , 其第 i 行对应的向量表示为 a_i , 第 j 列对应的向量表示为 a_j , 第 i 行、第 j 列位置元素为 a_{ij} . 给定数据集 $TS = [X, Y]$ ($X \in \mathbb{R}^{m \times \text{dim}}$, $Y \in \mathbb{R}^{m \times c}$, m 为训练集中评论数量, dim 为词空间大小, c 为评论类别数量), 其中第 i 条评论 (x_i, y_i) 分别包含此评论的内容信息和类别信息. 基于上面符号定

表2 本文中相关符号定义

符号	表示	举例
标量	小写斜体英文字母	a
向量	小写加粗英文字母	\mathbf{a}
矩阵	大写加粗英文字母	\mathbf{A}
词典	大写希腊字母	Φ
集合	小写希腊字母	θ
函数	首字母大写英文单词	Softmax

义, 本文的目标是学习一个分类器实现对待检测评论 x 的情感类别预测, 定义为函数 $y = \text{Classify}(x, \Phi, \theta)$, 其中 Φ 为大五人格性格词典, θ 表示分类器的全部参数.

3.2 方法描述

由于现有的情感分类方法缺乏对评论用户性格特征的描述, 本文提出一种基于用户性格和文本语义-结构特征的文本评论情感分类方法. 如图 2 所示, 首先, 利用大五人格模型中蕴含的知识获得用户性格特征; 然后, 鉴于现有基于 BiGRU 和 CNN 的分类模型存在未能有效融合文本局部结构特征和序列化上下文语义特征的问题, 提出基于 BiGRU-CNN 和双层 Attention 的评论文本特征获取方法 (BiGAC) 以获得评论文本的语义-结构特征; 最后, 引入混合 Attention 层对上述不同模型所得特征向量进行融合并通过全连接层获得最终的情感类别. 具体过程描述如下.

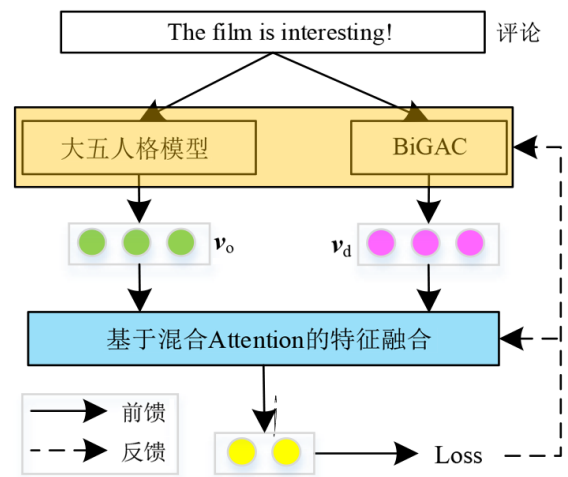


图2 本文方法执行过程

3.2.1 基于大五人格模型的性格特征获取

人们选择什么样的词汇来表达以及选择何种表达方式都与其自身的性格等本质化的人格特质相关. 性格会对人们表达情感的言语方式产生一定的影响, 性

格相近的人通过言语表达情感的方式往往也相近. 如表3所示, 由于一个词可能具有多个LIWC功能类别, Francois等人通过大五人格模型构建了某单词 w 的不同功能类别与大五人格中每一性格维度之间的得分关系^[28].

表3 单词 w 的不同功能类别及在五大性格维度上的得分

w 所属功能类别	功能类别对应各维度得分				
	openness	conscientiousness	extraversion	agreeableness	neuroticism
FC_1	s_{w11}	s_{w12}	s_{w13}	s_{w14}	s_{w15}
FC_2	s_{w21}	s_{w22}	s_{w23}	s_{w24}	s_{w25}
...
FC_n	s_{wn1}	s_{wn2}	s_{wn3}	s_{wn4}	s_{wn5}

表3中, s_{wij} 为词 w 的第 i 个功能类别在第 j 个性格维度上的大五人格得分. 基于文献[28], 本文先根据单词 w 的功能类别获得其在各性格维度上的得分值 s_{wij} , 然后平均这些得分值作为 w 的最终得分 s_{wj} , 具体如式(5)所示:

$$s_{wj} = \frac{\sum_{i=1}^{n_w} s_{wij}}{n_w} \quad (5)$$

其中, n_w 代表词 w 的功能类别数量. 针对LIWC词典 Φ 中所包含的词汇, 利用式(5)求出每个词在各性格维度上的得分值, 之后将评论文档 doc 中所有词在第 j 维性格上的得分相加即可得到 doc 在第 j 维性格上的得分 s_j^{doc} :

$$s_j^{\text{doc}} = \sum_{w \in \text{doc}} s_{wj} \quad (6)$$

然后, 将 doc 的性格特征向量表示为 $v_{\text{doc}} = [s_j^{\text{doc}}] (j=1, 2, 3, 4, 5)$. 为了降低数据差异引起的误差, 进一步对 v_{doc} 进行标准化, 以此得到 doc 对应的用户性格特征向量 $v_o = [o_j^{\text{doc}}] (j=1, 2, 3, 4, 5)$, 其中:

$$o_j^{\text{doc}} = \frac{s_j^{\text{doc}} - \text{Mean}(v_{\text{doc}})}{\text{Std}(v_{\text{doc}})} \quad (7)$$

其中, $\text{Mean}(v_{\text{doc}})$ 函数获得 v_{doc} 各维平均值, $\text{Std}(v_{\text{doc}})$ 函

数获得 v_{doc} 各维标准差. 在此基础上, 给出基于大五人格模型的性格特征获取方法执行过程如算法1所示.

算法1 基于大五人格模型的性格特征获取方法

```

输入: LIWC词典  $\Phi$ ; 文本  $\text{doc}$ .
输出:  $\text{doc}$  对应的用户性格特征向量  $v_o$ .
1: FOR  $j=1$  to 5 STEP 1
2:   FOR word  $w$  IN  $\text{doc}$ 
3:     SET  $s_{wj} \leftarrow 0$ .
4:     IF  $w$  IN  $\Phi$ 
5:       按照式(5)获得  $w$  在第  $j$  个性格维度上的得分  $s_{wj}$ .
6:     END IF
7:   END FOR
8:   按照式(6)获得  $\text{doc}$  在第  $j$  维性格上的得分  $s_j^{\text{doc}}$ .
9: END FOR
10: 按照式(7)对  $v_{\text{doc}}$  进行归一化处理, 获得最终性格特征向量  $v_o$ .
    
```

3.2.2 基于BiGRU和CNN的文本语义-结构特征向量获取

鉴于BiGRU和CNN能有效提取文本上下文语义特征和局部结构特征, 并且注意力机制可以通过设置不同的权重区分模型对不同特征的关注程度, 本文提出了一种基于BiGRU、CNN和双层注意力的文本语义-结构特征向量获取方法 BiGAC, 该方法执行流程如图3所示.

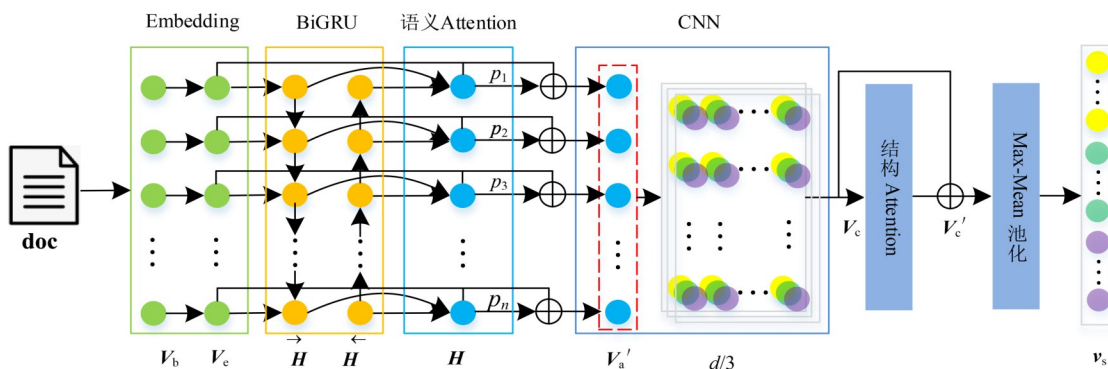


图3 BiGAC方法执行流程

给定输入文档 $\text{doc}=\{w_1, w_2, \dots, w_n\}$ (n 为词个数), Embedding 层输入为 doc 对应的 one-hot 向量 $V_{\text{oh}} \in \mathbf{R}^{n \times \text{dim}}$ (dim 为词空间大小), 输出词向量为 $V_c \in \mathbf{R}^{n \times d}$ (d 为词嵌入维度). BiGRU 层的输入为 V_c , 第 t 个单词对应的正向输出的隐藏状态 $\vec{h}_t \in \mathbf{R}^d$ 、反向输出的隐藏状态 $\overleftarrow{h}_t \in \mathbf{R}^d$ 和文本上下文语义特征向量 $h_t \in \mathbf{R}^d$ 计算过程如下所示:

$$\vec{h}_t = \text{GRU}(\vec{h}_{t-1}, e_t) \quad (8)$$

$$\overleftarrow{h}_t = \text{GRU}(\overleftarrow{h}_{t-1}, e_t) \quad (9)$$

$$h_t = \text{Avg}(\vec{h}_t, \overleftarrow{h}_t) \quad (10)$$

其中, e_t 为 V_c 中第 t 个词对应的词向量, GRU 函数计算过程如式(1)~(4)所示, Avg 函数实现向量对应位置元素求平均运算. 进一步地, 为区分不同单词对输出结果的影响, 引入语义 Attention 层来计算单词序列对应的权重矩阵 $P_{\text{sem}} \in \mathbf{R}^{n \times d}$:

$$P_{\text{sem}} = \text{Softmax} \left(\sigma \left(\left(W_{a1} H + B_{a1} \right) W_{a2} + B_{a2} \right) \right) \quad (11)$$

其中, σ 为激活函数, $H = [h_1, h_2, \dots, h_n] \in \mathbf{R}^{n \times d}$, $[\cdot, \cdot]$ 为矩阵行拼接操作, $W_{a1} \in \mathbf{R}^{n \times n}$, $B_{a1} \in \mathbf{R}^{n \times d}$, $W_{a2} \in \mathbf{R}^{d \times d}$, $B_{a2} \in \mathbf{R}^{n \times d}$. 将 P_{sem} 与 H 进行点乘以获得最终的文本上下文语义特征矩阵 $V_a \in \mathbf{R}^{n \times d}$, 如下所示:

$$V_a = P_{\text{sem}} \cdot H \quad (12)$$

为丰富评论向量所含的特征信息, 利用 Max 池化函数对特征矩阵 V_c 、 H 及 V_a 的拼接结果进行按列池化, 结果如式(13)所示:

$$V_a' = \text{Max}([V_c, H, V_a]) \quad (13)$$

进一步地, 将利用 CNN 层来提取 V_a' 中所含的局部结构特征. 给定不同感受野 $k \in \{3, 4, 5\}$, k 对应的卷积核数量 $n_f = d/3$ 和卷积步长 n_s ($n_s = 1$), 可以得到 V_a' 卷积后的矩阵 $V_c^k = [v_{ci}^k] = \left[\left[v_{cji}^k \right] \right] \in \mathbf{R}^{(n-k+1) \times n_f}$ ($1 \leq i \leq n_f, 1 \leq j \leq n-k+1$), 其中

$$v_{cji}^k = \sigma \left(W_{ci}^k * (V_a')_{j:j+k-1} + b^k \right) \quad (14)$$

式(14)中, $*$ 为卷积操作, $W_{ci}^k \in \mathbf{R}^{k \times d}$ 为感受野 k 对应的第 i 个卷积核, $(V_a')_{j:j+k-1} \in \mathbf{R}^{k \times d}$ 表示的是 V_a' 中第 j 到第 $(j+k-1)$ 个单词对应向量所构成的矩阵, $b^k \in \mathbf{R}$ 为偏置项. 进一步地, 对不同感受野 $k \in \{3, 4, 5\}$ 对应矩阵 V_c^k 进行列拼接得到 CNN 特征矩阵 $V_c \in \mathbf{R}^{(n-k+1) \times d}$, 即

$$V_c = [V_c^3 \| V_c^4 \| V_c^5] \quad (15)$$

其中, $[\|]$ 为矩阵列拼接操作. 进一步地, 为区分不同结构特征对输出结果的影响, 引入结构 Attention 层, V_c 对应的权重得分矩阵 $P_{\text{str}} \in \mathbf{R}^{(n-k+1) \times d}$ 计算如下:

$$P_{\text{str}} = \text{Softmax} \left(\left(\sigma \left(W_{a3} V_c + B_{a3} \right) \right) \right) \quad (16)$$

其中, $W_{a3} \in \mathbf{R}^{(n-k+1) \times (n-k+1)}$, $B_{a3} \in \mathbf{R}^{(n-k+1) \times d}$. 将 P_{str} 与 V_c 进行点乘获得矩阵 V_c' , 并采用 Max-Mean 池化方法对特征矩阵 V_c 、 V_c' 的行拼接结果执行式(17)得到矩阵 $V_d \in \mathbf{R}^{2 \times d}$:

$$V_d = \left[\text{Max}([V_c', V_c]), \text{Mean}([V_c', V_c]) \right] \quad (17)$$

式(17)中, Max 函数获得矩阵每列最大值, Mean 函数获得矩阵每列平均值. 最后, 将 V_d 一维化以获得最终的文本语义-结构特征向量 $v_s \in \mathbf{R}^{2d}$. 在此基础上, 给出基于 BiGRU 和 CNN 的文本语义-结构特征向量获取方法 (BiGAC) 执行过程如算法 2 所示.

算法 2 基于 BiGRU 和 CNN 的评论文本语义-结构特征获取方法 (BiGAC)

输入: 文档 $\text{doc}=\{w_1, w_2, \dots, w_n\}$; 激活函数 σ ; 词嵌入维度 d ; CNN 卷积核感受野 $k \in \{3, 4, 5\}$; 卷积核数量 $n_f = d/3$; 卷积步长 n_s .

输出: 文本 doc 的语义-结构特征向量 v_s .

- 1: Embedding 层通过使用 Word2vec 模型获得 doc 的词向量矩阵 V_c .
- 2: 将 V_c 作为 BiGRU 层的输入, 获得输出矩阵 $H=[h_1, h_2, \dots, h_n]$.
- 3: 利用式(11)计算隐藏层状态向量 H 对应的注意力权重矩阵 P_{sem} , 并利用式(12)和式(13)获得特征矩阵 V_a' .
- 4: 将 V_a' 输入至 CNN 层, 按照式(14)~(17)得到 doc 的语义-结构特征矩阵 V_d .
- 5: 将 V_d 一维化获得最终的文本语义-结构特征向量 v_s .

3.2.3 基于混合 Attention 层的特征融合

为避免传统特征融合方法通过简单特征拼接来融合不同类型特征而导致的无法区分不同特征重要性的问题, 本文在 3.2.1 节和 3.2.2 节基础上引入混合 Attention 层来融合用户性格特征和文本语义-结构特征.

具体而言, 如图 4 所示, 首先, 将基于大五人格的性格特征获取方法所得的用户性格特征向量 $v_o \in \mathbf{R}^5$ 映射到与 v_s 维度相同的向量空间, 并将所得结果与 $v_s \in \mathbf{R}^{2d}$ 进行拼接, 如式(18)所示:

$$v = (v_o W_o) \oplus v_s \quad (18)$$

其中, \oplus 为向量拼接操作, $v \in \mathbf{R}^{4d}$, $W_o \in \mathbf{R}^{5 \times 2d}$ 为参数矩阵. 为了区分两类特征中不同维度的重要性, 混合 Attention 层按照式(19)得到 v 对应的注意力权重向量 $p_{\text{hyb}} \in \mathbf{R}^{4d}$:

$$p_{\text{hyb}} = \text{Softmax} \left(\sigma(v W_f + b_f) \right) \quad (19)$$

其中, $W_f \in \mathbf{R}^{4d \times 4d}$, $b_f \in \mathbf{R}^{4d}$. 在此基础上将 p_{hyb} 和向量 v 对应元素相乘, 并将结果输入至全连接层得到评论 doc 的最终预测向量 $\hat{y} \in \mathbf{R}^c$ (c 为类别总数):

$$\hat{y} = \text{Softmax} \left(\sigma \left((v \cdot p_{\text{hyb}}) W_p + b_p \right) \right) \quad (20)$$

其中, $W_p \in \mathbf{R}^{4d \times c}$, $b_p \in \mathbf{R}^c$. 最后, 针对每个训练批次计算评论情感真实类别和交叉熵损失值 Loss, 并通过反向

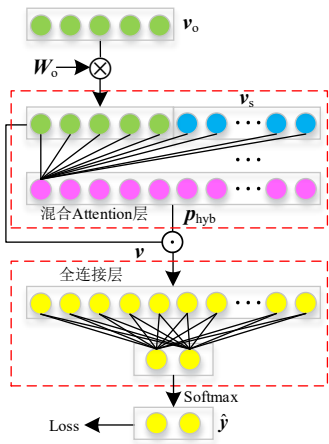


图4 基于混合 Attention 层的特征融合

传播和梯度下降来优化算法中所有训练参数,具体如式(21)所示:

$$\text{Loss} = -\frac{1}{n_b} \sum_{i=1}^{n_b} \sum_{j=1}^c \left(y_{ij} \log \hat{y}_{ij} + (1 - y_{ij}) \log (1 - \hat{y}_{ij}) \right) \quad (21)$$

其中, n_b 为批大小, y_{ij} 为批内第 i 条评论的真实类别向量对应的第 j 维的值, \hat{y}_{ij} 为第 i 条评论的预测向量 \hat{y}_i 对应第 j 维的值. 在此基础上, 给出本文方法的整体流程如算法 3 所示.

4 实验结果与分析

4.1 实验环境和数据集

实验硬件环境: GPU 为 2080 Ti, 内存为 64 GB, 硬盘为 1 TB; 软件环境: 操作系统为 Ubuntu 16.04.7, 深度学习框架为 Tensorflow 2.0, 编程语言为 Python 3.8.5. 为充分验证本文方法的情感分类性能, 分别选择二分类英文评论数据集 IMDB、Yelp-2 以及多分类英文评论数据集 Yelp-5、Ekman^[24, 29] 进行对比实验. 其中, Yelp-2 为在原始数据集 Yelp^[30] 基础上将评分不小于 3 的评论确定为积极, 将评分小于 3 的评论确定为消极而得到的数据集; Yelp-5 则直接使用 Yelp 数据集中 5 个评分等级作为标签类别. 在此基础上, 本文按照 8:1:1 比例将各数据集划分为训练集、验证集和测试集, 并且所有方法使用相同的训练集、验证集和测试集进行实验. 表 4 给出了各数据集中对应的样本数、类别数等信息. 图 5 给出了各数据集中不同长度范围对应的评论数量. 由图 5 知, IMDB 中长度处于 [100, 200] 区间内的评论数最多, Yelp 中长度处于 [0, 100] 区间内的评论数最多, 而 Ekman 中所有评论长度均小于 100. 考虑到算法的计算开销, 本文分别针对 IMDB、Yelp-2、Yelp-5 以及 Ekman 设置其文本最大长度 max_len 为 200、200、200 及 100. 为了避免实验误差, 针对每种方法取 10 次实验平均值作为最终的实验结果.

算法 3 基于大五人格模型和语义-结构特征的文本评论情感分类方法(BF_BiGAC)

输入: 数据集 $TS=[X, Y](X \in \mathbb{R}^{m \times \text{dim}}, Y \in \mathbb{R}^{m \times c}, m$ 为训练集中评论数量, dim 为词空间大小, c 为评论类别数量); 轮次 n_e ; 批大小 n_b ; 待分类评论数据 doc; 训练参数集 θ .

输出: doc 的情感类别.

Step1 模型训练

1-1 随机初始化 θ .

1-2 FOR $e=1$ TO n_e STEP 1

1-3 计算每轮迭代次数 $t=\text{int}(m/n_b)$.

1-4 FOR $i=1$ TO t STEP 1

1-5 随机选择 n_b 大小的样本集 S .

1-6 SET $\hat{Y} \leftarrow \phi$.

1-7 FOR EACH d in S

1-8 按照算法 1 获得 d 的性格特征向量 v_o .

1-9 按照算法 2 获得 d 的语义-结构特征向量 v_s .

1-10 按照式(19)计算权重向量 p_{hyb} .

1-11 按照式(20)计算 d 对应的预测向量 \hat{y} .

1-12 SET $\hat{Y} \leftarrow [\hat{Y}; \hat{y}]$.

1-13 END FOR

1-14 按照式(21)计算损失值, 反向传播更新参数集 θ .

1-15 END FOR

1-16 END FOR

Step2 预测评论 doc 的情感类别

2-1 按照算法 1 获得 doc 对应的性格特征向量 v_o .

2-2 按照算法 2 获得 doc 对应的语义-结构特征向量 v_s .

2-3 利用式(18)和式(19)获得权重向量 p_{hyb} .

2-4 按照式(20)获得 doc 的情感类别向量 \hat{y} , 继而获得 doc 对应的情感类别 $\text{Class}(\text{doc}) \leftarrow \max_{C_i} (\hat{y}_i)$, 其中 \hat{y}_i 为 \hat{y} 对应第 i 维的值.

表 4 不同数据集的相关信息

数据集	总样本数	类别数	平均长度	最大长度
IMDB	50 000	2	229.397 14	2 353
Yelp-2	50 000	2	101.857	1 027
Yelp-5	50 000	5	101.857	1 027
Ekman	40 898	6	12.959	35

4.2 算法性能评价标准

考虑到不同数据集中包含类别数不同, 使用准确率 (Accuracy) 和加权 macro $F_1 (F_w)$ 值作为情感分类效果的评价指标, 计算方式如下^[31]:

$$\text{Accuracy} = \frac{\sum_{1 \leq i \leq c} \text{TP}_i}{\sum_{1 \leq i \leq c} \text{TP}_i + \sum_{1 \leq i \leq c} \text{FP}_i} \quad (22)$$

$$R_i = \frac{\text{TP}_i}{\text{TP}_i + \text{FN}_i} \quad (23)$$

$$P_i = \frac{\text{TP}_i}{\text{TP}_i + \text{FP}_i} \quad (24)$$

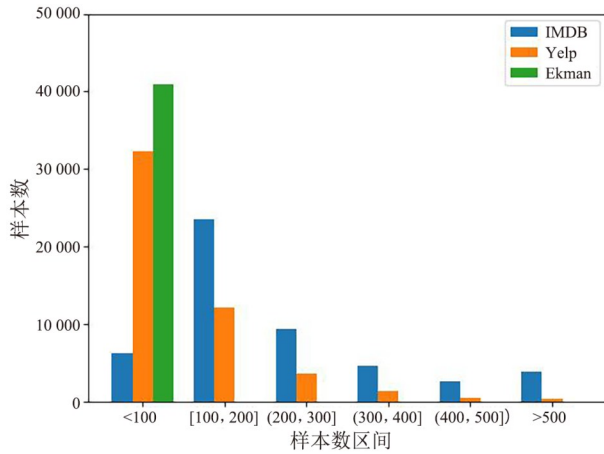


图5 不同数据集中样本长度分布

$$F_w = \sum_{1 \leq i \leq c} \left(\frac{2P_i R_i}{P_i + R_i} \times \frac{N_i}{N} \right) \quad (25)$$

其中, TP_i 表示真实类别标签为 C_i 且被正确分类的评论数量, FN_i 表示真实类别标签为 C_i 但被错误分类的评论数量, FP_i 表示真实类别标签不为 C_i 但被误识为 C_i 的评论数量, N_i 为类别 C_i 中的评论数量, N 为评论总数. 可见, Accuracy 和 F_w 值均属于区间 $[0, 1]$, 取值越接近 1, 说明算法针对评论文本的分类效果越好.

4.3 实验对比方法及参数设置

选取以下典型方法与本文进行对比: BiGRU^[27]、BiGRU_AT (Attention-based BiGRU method)^[17]、MC_AttCNN_AttBiGRU (Multi-Channel Attention-based method concatenating CNN and BiGRU)^[15]、BiGRU_CNN^[22]、CNN_BiGRU_AT (Attention-based method concatenating CNN and BiGRU)^[21]、ConvBiLSTM^[20]、BERT (Bidirectional Encoder Representations from Transformers)^[32]、RoBERTa (Robustly optimized BERT approach)^[33] 及 TextING (Inductive Text classification via Graph Neural network)^[34]. 其中, BiGRU_AT 在 BiGRU 的基础上增加了 Attention 层以区分不同隐状态的贡献; MC_AttCNN_AttBiGRU 为基于注意力机制、多通道 CNN 和 BiGRU 的神经网络模型. 该模型通过注意力机制区分句子中对情感极性分类重要的词语, 并且将多通道 CNN 和 BiGRU 处理后的特征进行拼接以获得最终的评论向量表达; CNN_BiGRU_AT 将词向量依次输入至 CNN 和 BiGRU 中, 并增加 Attention 层以区分不同隐状态的贡献; 区别于 CNN_BiGRU_AT, ConvBiLSTM 直接将词向量依次输入至 CNN 和 BiLSTM 中, 以此获得评论的向量表达. 由于 Skip-gram 结构在捕获语义信息准确性方面要比 CBOW 效果好, 因此本文使用 Skip-gram 获得单词词向量. 经参数调优, 设置所有方法中随机失活率 $r_d=0.5$ 、训练轮次 $n_e=50$; 设置 BERT 和 RoBERTa 对应批大小 $n_b=$

16、学习率 $l_r=10^{-5}$, 设置其他方法中批大小 $n_b=64$ 、词向量维度 $d_e=300$ 、学习率 $l_r=10^{-4}$. 除此之外, 不同方法对应的其他参数如表 5 所示.

表5 不同模型对应的参数

模型	参数	参数值
BiGRU/BiLSTM	隐藏层维度	300
	卷积核数量	100
CNN	卷积步长	1
	卷积核大小	3,4,5
BERT	版本信息	BERT-base-uncased
	嵌入维度	768
RoBERTa	版本信息	roBERTa-base
	嵌入维度	768
TextING	层数	3
	窗口大小	3
	隐藏层维度	300

4.4 大五人格模型有效性验证

为验证大五人格模型在提升评论情感分类性能方面的有效性, 计算了本文方法 BF_BiGAC 和单纯使用语义-结构特征的情感分类方法 (BiGAC_FC) 在不同数据集上对应的 Accuracy 值和 F_w 值, 结果如图 6、图 7 所示. 其中, BiGAC_FC 方法使用 BiGAC 获得文本语义-结构特征向量, 之后输入至全连接层和 Softmax 层完成情感分类. 可见, 综合考虑用户性格特征和文本语义-结构特征的情感分类方法不论是 Accuracy 值还是 F_w 值均高于单纯使用文本语义-结构特征的情感分类方法. 例如, 在 Accuracy 值方面, 本文相对于 BiGAC_FC 在 IMDB、Yelp-2、Yelp-5 及 Ekman 数据集上分类效果分别提升了 0.006、0.010、0.010 及 0.008; 在 F_w 值方面, 本文相对于 BiGAC_FC 在 IMDB、Yelp-2、Yelp-5 及 Ekman 数据集上分类效果分别提升了 0.005、0.007、0.010 及 0.007. 进一步发现, 大五人格模型在数据集 IMDB 和 Ekman 上带来的提升幅度略小于数据集 Yelp-2、Yelp-5 上对应结果, 可能的原因在于前两者数据集中蕴含的用户性格知识较少, 限制了本文方法情感分类精度的提升. 综上, 虽然 BF_BiGAC 方法在不同数据集上的表现稍有不同, 但其表现从整体来看仍普遍优于 BiGAC_FC 方法, 验证了大五人格模型中蕴含的用户性格知识对于丰富评论特征、提升评论情感分类表现方面的有效性.

4.5 BiGAC 方法有效性验证

为验证 BiGAC 方法所得文本上下文语义-结构特征信息在提升情感分类精度方面的有效性, 将上述 BiGAC_FC 方法与以下 3 种典型方法进行对比. 每种方法执行过程如下.

(1) BiGRU_AT_FC. 通过 BiGRU 获得文本语义表示向量 \rightarrow 语义 Attention 层 \rightarrow 最大池化层 \rightarrow 全连接层 \rightarrow

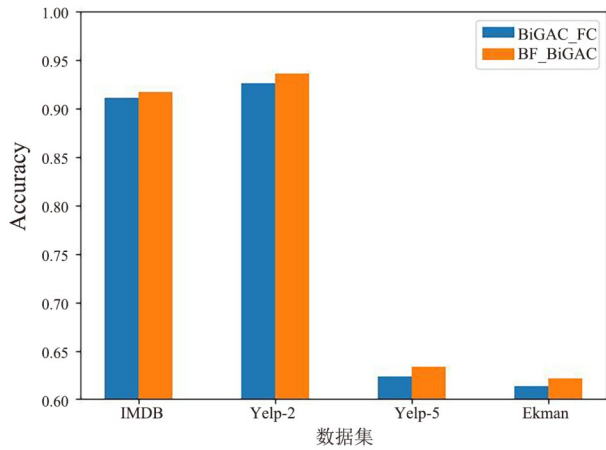


图6 不同情感分类方法的 Accuracy 值比较

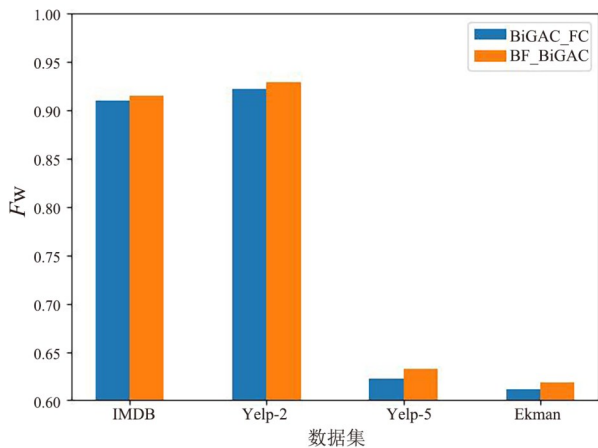


图7 不同情感分类方法的 F_w 值比较

Softmax 层.

(2) CNN_AT_FC. 通过 CNN 获得文本结构表示向量→结构 Attention 层→最大池化层→全连接层→Softmax 层.

(3) CNN_BiGRU_AT_FC. 通过 CNN 获得文本结构表示向量→结构 Attention 层→利用 BiGRU 获得文本语义表示向量→语义 Attention 层→最大池化层→全连接层→Softmax 层.

在此基础上,计算不同算法在不同数据集上对应的 Accuracy 值和 F_w 值,结果如图 8、图 9 所示. 可见, BiGRU_AT_FC 和 CNN_AT_FC 受数据集中类别数量影响较大,在多类别数据集 Yelp-5 和 Ekman 上表现较差,对应 Accuracy 值均小于 0.5,对应的 F_w 值均小于 0.4. 究其原因, BiGRU_AT_FC 直接将所有单词对应的向量进行最大池化操作,因而丢失了较多的局部结构信息,而 CNN_AT_FC 则直接对所有单词向量进行卷积操作,因此忽略了单词的上下文语义信息. 相对于 BiGRU_AT_FC 和 CNN_AT_FC 而言, BiGAC_FC 在 Accuracy 值和 F_w 值

上获得明显提升. 例如,在 Accuracy 值方面, BiGAC_FC 相较于 CNN_AT_FC 在 IMDB、Yelp-2、Yelp-5 及 Ekman 数据集上分别提升了 0.015、0.020、0.204 及 0.313;在 F_w 值方面, BiGAC_FC 相较于 BiGRU_AT_FC 在 IMDB、Yelp-2、Yelp-5 及 Ekman 数据集上分别提升了 0.021、0.032、0.313 及 0.302. 可见, BiGAC 充分发挥了 BiGRU 和 CNN 各自的优势,不仅能够提取到句子中连续单词的局部特征信息,还能够捕获句子上下文语义信息,继而有效提高情感分类的准确性. 进一步比较 BiGAC_FC 和 CNN_BiGRU_AT_FC 得知,在 Accuracy 值方面,前者相较于后者在 IMDB、Yelp-2、Yelp-5 及 Ekman 数据集上分别提升了 0.007、0.002、0.011 及 0.018,在 F_w 值上分别提升了 0.005、0.001、0.015 及 0.022. 可见,除在 Yelp-2 数据集外, BiGAC_FC 对应的结果明显高于后者,尤其在应对多类别数据集 Yelp-5 及 Ekman 时, BiGAC_FC 相对于 CNN_BiGRU_AT_FC 优势较为明显,这主要是因为 CNN_BiGRU_AT_FC 虽然能提取卷积块之间的语义关联信息,但在文本单词嵌入基础上直接进行卷积操作丢失了卷积块内部的上下文语义信息,而 BiGAC_FC 利用 BiGRU 层较好地保留了单词的上下文语义信息,通过结合 CNN 层所得结构信息有效获取了区分不同情感类别的深层特征,进而有效提升了情感分类的准确性.

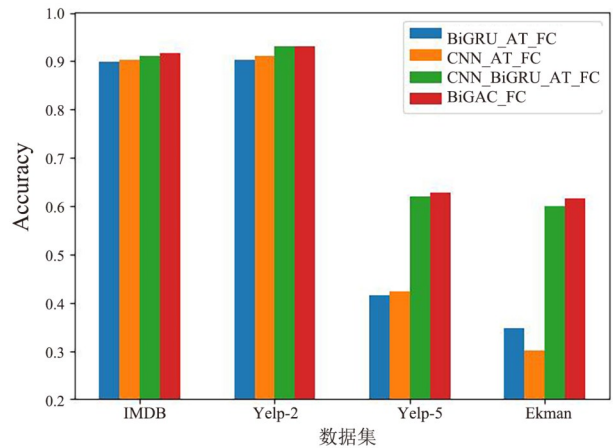
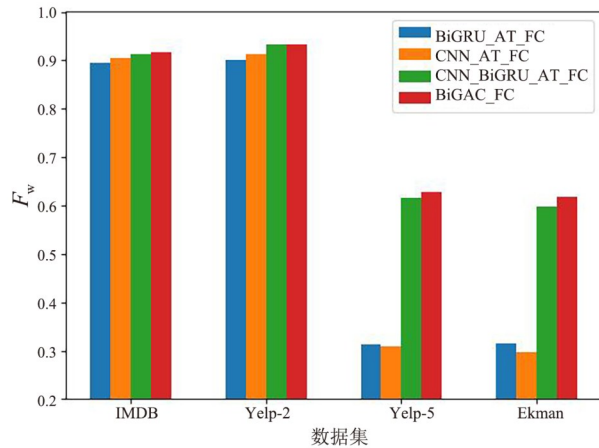


图8 不同方法的 Accuracy 值比较

4.6 注意力机制有效性验证

为了验证本文注意力机制的有效性,通过在本文方法 BF_BiGAC 基础上移除不同的 Attention 层来构造以下几种变体方法并将其与本文进行对比.

- (1) NO_AT: 移除语义 Attention 层、结构 Attention 层和混合 Attention 层;
- (2) GRU_AT: 移除结构 Attention 层和混合 Attention 层;
- (3) CNN_AT: 移除语义 Attention 层和混合 Attention 层;

图9 不同方法的 F_w 值比较

(4) FF_AT: 移除语义 Attention 层和结构 Attention 层;

(5) GRU_CNN_AT: 移除混合 Attention 层;

(6) GRU_FF_AT: 移除结构 Attention 层;

(7) CNN_FF_AT: 移除语义 Attention 层.

表6、表7给出了上述不同算法在不同数据集上对应的 Accuracy 值和 F_w 值. 为便于比较, 每个数据集上的最大值已加粗显示. 由表知, 由于本文在模型中引入了三层 Attention 机制, 其对应结果在4个数据集上普遍大于其他方法. 例如, 在 Yelp-2 数据集上, 本文获得最大 Accuracy 值 0.927、最大 F_w 值 0.922, 在 Ekman 数据集上, 本文获得最大 Accuracy 值 0.625、最大 F_w 值 0.617, 验证了本文引入的三层注意力机制无论在处理两类别情感分类任务还是多类别情感分类任务均显示出明显优势.

进一步将 NO_AT 与 GRU_AT、CNN_AT 和 FF_AT 对比得知, 加入注意力机制可以较好地提高模型分类的准确性, 因为它能让模型关注那些对分类贡献较大的重要信息. 其中, GRU_AT 表现明显优于其他3种方法, 说明在语义 Attention 层中考虑不同单词的权重信息可以有效提升算法的分类精度. 通过将本文与 GRU_CNN_AT、GRU_FF_AT 和 CNN_FF_AT 对比发现, 本文表现显著优于 GRU_FF_AT 和 CNN_FF_AT, 但与 GRU_CNN_AT 相对接近, 说明混合 Attention 层对模型结果的影响最小, 可能原因在于混合 Attention 层针对长度为 $4d$ 的一维向量计算注意力权重, 但其后的全连接层也可以从一定程度区分不同维度的重要性, 进而降低了混合 Attention 层的影响.

4.7 综合对比

4.7.1 不同方法精度比较

为验证本文 BF_BiGAC 方法的有效性, 将其与 4.3 节中给出的方法进行比较. 除在测试集上验证方法表现外, 这里还进一步在验证集上将不同方法对应的最

表6 本文与不同变体方法在 IMDB 和 Yelp-2 数据集上的比较

方法	IMDB		Yelp-2	
	Accuracy	F_w	Accuracy	F_w
NO_AT	0.876	0.876	0.906	0.904
GRU_AT	0.887	0.883	0.919	0.918
CNN_AT	0.882	0.881	0.916	0.914
FF_AT	0.881	0.881	0.908	0.908
GRU_CNN_AT	0.914	0.914	0.920	0.919
GRU_FF_AT	0.901	0.901	0.918	0.916
CNN_FF_AT	0.894	0.894	0.917	0.915
本文	0.912	0.917	0.927	0.922

表7 本文与不同变体方法在 Yelp-5 和 Ekman 数据集上的比较

方法	Yelp-5		Ekman	
	Accuracy	F_w	Accuracy	F_w
NO_AT	0.604	0.596	0.585	0.585
GRU_AT	0.613	0.612	0.604	0.604
CNN_AT	0.616	0.606	0.593	0.589
FF_AT	0.611	0.604	0.589	0.586
GRU_CNN_AT	0.623	0.619	0.617	0.615
GRU_FF_AT	0.620	0.615	0.604	0.603
CNN_FF_AT	0.625	0.615	0.601	0.602
本文	0.632	0.622	0.625	0.617

优 Accuracy 值和 F_w 值进行对比, 结果如表 8~11 所示. 为方便对比, 表中每个数据集上的最大值已加粗显示, 下划线标注结果表示本文方法相对于该结果增幅大于等于 0.005 且小于 0.010, *标注结果表示本文方法相对于该结果增幅小于 0.005.

由表中可知, BiGRU、BiGRU_AT 及 TextING 方法在不同数据集上对应结果均低于其他方法, 说明单纯考虑单词的语义上下文结构信息难以获得较好的情感分类表现. 并且, BiGRU 和 BiGRU_AT 在多类别数据集 Yelp-5 和 Ekman 上的表现明显逊于 TextING, 在测试集和验证集上对应的结果均小于 0.400, 这主要是因为 TextING 除考虑单词上下文语义信息外还利用 GCN 提取了远距离单词对间的关联信息, 因此能更好地实现评论语句的嵌入表达. 对比 MC_AttCNN_AttBiGRU、BiGRU_CNN、CNN_BiGRU_AT 及 ConvBiLSTM 发现, 这4种方法表现普遍优于 BiGRU、BiGRU_AT 及 TextING 方法, 说明综合考虑文本单词上下文语义相关性及文本局部结构特征能较好地提升评论文本的情感分类效果. 进一步发现, MC_AttCNN_AttBiGRU 及 CNN_BiGRU_AT 表现优于 BiGRU_CNN 及 ConvBiLSTM, 说明结合 Attention 机制能细化不同特征的重要性, 进而能有效提升情感分类的准确性. 当将本文与 BiGRU、BiGRU_AT 及 TextING 比较时发现, 本文表现显著优于后3种方法, 对应结果增幅大于等于 0.010 的情况占比为 97.9%; 当将本文与 MC_AttCNN_AttBi-

GRU及CNN_BiGRU_AT比较时发现,本文对应结果普遍偏高,对应结果增幅大于等于0.010的情况占比为50.0%、增幅大于等于0.005的情况占比为87.5%。具体来看,当使用数据集Yelp-2数据集时,本文相对于MC_AttCNN_AttBiGRU和CNN_BiGRU_AT在测试集上的Accuracy值分别提升0.010和0.013, F_w 值分别提升0.005和0.008;当使用数据集Ekman时,本文相对于MC_AttCNN_AttBiGRU和CNN_BiGRU_AT在验证集上的Accuracy值分别提升0.006和0.013, F_w 值分别提升0.013和0.014,说明本文针对以往算法仅考虑评论语义、结构特征的不足引入用户性格特征及三层Attention机制对于提升情感分类效果是有效的。进一步地,当将BF_BiGAC与BERT、RoBERTa比较时发现,RoBERTa对应结果普遍高于BERT,而BF_BiGAC仅在Yelp-2数据集上表现稍逊于BERT和RoBERTa,这可能是因为BERT和RoBERTa虽在超大规模数据集上进行了预训练,但模型所使用的训练参数量较大,因此模型训练结果易受到学习率、批大小等重要参数的影响,难以在所有数据集上均获得最好的学习效果。整体来看,本文相对于表8~11中给出的9种基线方法在Accuracy值和 F_w 值上对应结果增幅大于等于0.010的情况占比分别为77.78%、80.55%,增幅大于等于0.005的情况占比分别为87.50%、94.44%,说明该方法无论从Accuracy值上还是 F_w 值上均具有明显优势,验证了其在提高情感分类精度方面的有效性。

表8 不同方法在测试集上的Accuracy值比较

方法	IMDB	Yelp-2	Yelp-5	Ekman
BiGRU	0.887	0.916	0.351	0.317
BiGRU_AT	0.893	0.919	0.321	0.382
MC_AttCNN_AttBiGRU	0.913*	0.924	0.625	0.618*
BiGRU_CNN	0.896	0.922	0.619	0.610
CNN_BiGRU_AT	<u>0.909</u>	0.921	0.621	<u>0.616</u>
ConvBiLSTM	0.893	0.916	0.617	0.614
BERT	0.898	0.935*	0.626	0.565
RoBERTa	0.906	0.938*	0.633*	0.587
TextING	0.878	0.896	0.571	0.562
BF_BiGAC	0.916	0.934	0.636	0.621

4.7.2 不同方法效率对比

为验证本文方法的执行效率,这里统计了上述不同方法的训练参数量 n_p 以及每种方法在不同训练集上每秒处理的文档数均值 n_q 。为便于比较,对 n_q 执行以10为底取对数操作并将结果 $\lg(n_q)$ 以及 n_p 进行柱状图展示,结果如图10所示。由图知,BiGRU方法执行效率最高,其对应的参数量最小,在单位时间内处理的文档数明显高于其他算法。BiGRU_AT和TextING对应结果相近,这可能是因为两者均包含BiGRU层,并且BiGRU_AT中Attention层所含训练参数量与TextING中GCN层中所含的训练参数

表9 不同方法在验证集上的最优Accuracy值比较

方法	IMDB	Yelp-2	Yelp-5	Ekman
BiGRU	0.891	0.922	0.349	0.324
BiGRU_AT	0.902	<u>0.924</u>	0.323	0.396
MC_AttCNN_AttBiGRU	0.919*	0.931*	0.636	<u>0.626</u>
BiGRU_CNN	0.904	<u>0.926</u>	0.625	0.621
CNN_BiGRU_AT	0.911	<u>0.925</u>	0.627	0.619
ConvBiLSTM	0.906	0.919	0.622	0.616
BERT	0.903	0.936*	0.628	0.572
RoBERTa	0.911	0.941*	0.634	0.603
TextING	0.886	0.906	0.582	0.571
BF_BiGAC	0.923	0.933	0.646	0.632

表10 不同方法在测试集上的 F_w 值比较

方法	IMDB	Yelp-2	Yelp-5	Ekman
BiGRU	0.887	0.914	0.351	0.316
BiGRU_AT	0.893	0.918	0.318	0.379
MC_AttCNN_AttBiGRU	<u>0.906</u>	<u>0.924</u>	<u>0.636</u>	<u>0.622</u>
BiGRU_CNN	0.892	<u>0.922</u>	0.616	0.610
CNN_BiGRU_AT	0.904	<u>0.921</u>	0.628	0.616
ConvBiLSTM	0.893	0.916	0.616	0.605
BERT	0.895	0.929*	0.625	0.565
RoBERTa	0.902	0.933*	0.633	0.585
TextING	0.876	0.897	0.572	0.562
BF_BiGAC	0.915	0.929	0.644	0.628

表11 不同方法在验证集上的最优 F_w 值比较

方法	IMDB	Yelp-2	Yelp-5	Ekman
BiGRU	0.896	0.919	0.355	0.316
BiGRU_AT	0.903	0.921	0.321	0.379
MC_AttCNN_AttBiGRU	<u>0.911</u>	<u>0.928</u>	0.639	0.619
BiGRU_CNN	0.906	<u>0.925</u>	0.622	0.610
CNN_BiGRU_AT	0.908	0.921	<u>0.641</u>	0.618
ConvBiLSTM	0.901	0.919	0.628	0.609
BERT	0.897	0.935*	0.629	0.579
RoBERTa	0.905	0.936*	0.635	0.601
TextING	0.882	0.904	0.587	0.579
BF_BiGAC	0.919	0.933	0.649	0.632

量相当。进一步发现,本文在执行效率上与MC_AttCNN_AttBiGRU、CNN_BiGRU_AT、ConvBiLSTM相近。但是,结合表8~11知,本文在Accuracy和 F_w 值上表现相较于这些方法优势明显,说明其在提升情感分类精度同时并未带来显著的计算开销。与预训练模型BERT和RoBERTa相比,由图10可以看出,本文无论在训练参数量还是每秒处理的评论数量上均具有明显优势,说明该方法能以相对较少的计算开销获得较好的分类表现。综上可知,本文结合了大五人格模型、BiGRU模型和CNN模型各自的优势,通过引入语义Attention层、结构Attention层和混合At-

tention层实现了用户性格特征、上下文语义特征和文本结构特征的有效融合,在增加较小计算开销的前提下,有效提高了评论文本情感分类的准确性.

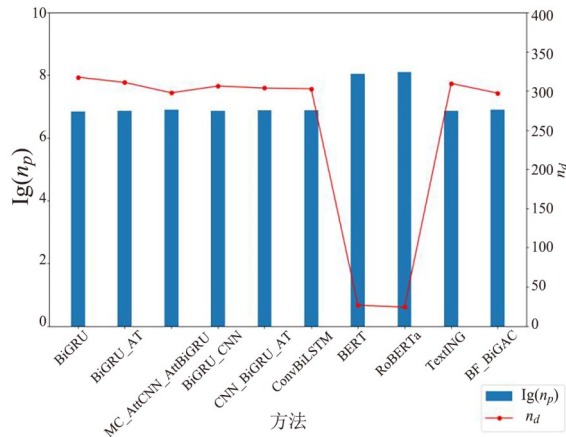


图10 不同方法对应的参数量及每秒处理的文档数比较

5 结论

提出了一种基于用户性格和语义-结构特征的文本评论情感分类方法,主要贡献包括:(1)综合考虑用户的性格因素对其情感倾向性的影响,利用大五人格模型在获取用户性格方面的优势实现用户性格特征表达;(2)针对当前评论特征获取方法难以完整表达文本语义、结构特征的问题,构建了基于BiGRU、CNN和双层注意力的文本语义-结构特征获取方法,该方法引入语义Attention层和结构Attention层将上下文语义特征和文本结构特征进行融合,避免了传统CNN-BiGRU串行网络结构忽略卷积块中单词上下文语义信息的不足;(3)引入混合Attention层,区分了不同类型特征对评论情感分类结果的影响,实现了用户性格和语义-结构特征的有效融合.实验结果表明,本文方法相较于典型现有方法在情感分类准确性和执行效率上均具有一定的优势.未来将考虑利用用户属性、用户社会关系、用户历史评论等信息挖掘用户性格特征,以进一步提高文本评论情感分类的准确性.

参考文献

- [1] 栗雨晴, 礼欣, 韩煦, 等. 基于双语词典的微博多类情感分析方法[J]. 电子学报, 2016, 44(9): 2068-2073.
LI Y Q, LI X, HAN X, et al. A bilingual lexicon-based multi-class semantic orientation analysis for microblogs [J]. Acta Electronica Sinica, 2016, 44(9): 2068-2073. (in Chinese)
- [2] DASHTIPOUR K, GOGATE M, GELBUKH A, et al. Extending persian sentiment lexicon with idiomatic expressions for sentiment analysis[J]. Social Network Analysis and Mining, 2022, 12(1): 1-13.
- [3] 曾雪强, 华鑫, 刘平生, 等. 基于情感轮和情感词典的文本情

感分布标记增强方法[J]. 计算机学报, 2021, 44(6): 1080-1094.

ZENG X Q, HUA X, LIU P S, et al. Emotion wheel and lexicon based text emotion distribution label enhancement method [J]. Chinese Journal of Computers, 2021, 44(6): 1080-1094. (in Chinese)

- [4] BLEI D M, NG A Y, JORDAN M I. Latent dirichlet allocation [J]. Journal of Machine Learning Research, 2003, 3: 993-1022.
- [5] LIN C, HE Y, EVERSON R, et al. Weakly supervised joint sentiment-topic detection from text[J]. IEEE Transactions on Knowledge and Data engineering, 2011, 24(6): 1134-1145.
- [6] PORIA S, CHATURVEDI I, CAMBRIA E, et al. Sentic LDA: Improving on LDA with semantic similarity for aspect-based sentiment analysis[C]//2016 International Joint Conference on Neural Networks (IJCNN). Piscataway: IEEE, 2016: 4465-4473.
- [7] OZYURT B, AKCAYOL M A. A new topic modeling based approach for aspect extraction in aspect based sentiment analysis: SS-LDA[J]. Expert Systems with Applications, 2021, 168: 114231.
- [8] 黄发良, 冯时, 王大玲, 等. 基于多特征融合的微博主题情感挖掘[J]. 计算机学报, 2017, 40(4): 872-888.
HUANG F L, FENG S, WANG D L, et al. Topic sentiment model based on multi-feature fusion[J]. Chinese Journal of Computers, 2017, 40(4): 872-888. (in Chinese)
- [9] MENG Y, ZHANG Y, HUANG J, et al. Text classification using label names only: A language model self-training approach[EB/OL]. (2020)[2022]. <https://arxiv.org/abs/2010.07245>.
- [10] WANG Y, HUANG S T. Training TSVM with the proper number of positive samples[J]. Pattern Recognition Letters, 2005, 26(14): 2187-2194.
- [11] SAMAH K A. Naïve Bayes Twitter sentiment analysis in visualizing the reputation of communication service providers: During Covid-19 pandemic[J]. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 2021, 12(5): 1753-1764.
- [12] XIA H, YANG Y, PAN X, et al. Sentiment analysis for online reviews using conditional random fields and support vector machines[J]. Electronic Commerce Research, 2020, 20(2): 343-360.
- [13] IQBAL M, KARIM A, KAMIRAN F. Balancing prediction errors for robust sentiment classification[J]. ACM Transactions on Knowledge Discovery from Data (TKDD), 2019, 13(3): 1-21.
- [14] HAMA AZIZ R H, DIMILILER N. SentiXGboost: Enhanced sentiment analysis in social media posts with ensemble XGBoost classifier[J]. Journal of the Chinese Institute of Engineers, 2021, 44(6): 562-572.
- [15] 程艳, 叶子铭, 王明文, 等. 基于注意力机制的多通道

- CNN 和 BiGRU 的文本情感倾向性分析[J]. 计算机研究与发展, 2020, 57(12): 2583-2595.
- CHENG Y, YE Z M, WANG M W, et al. Text sentiment orientation analysis of multi-channels CNN and BiGRU based on attention mechanism[J]. Journal of Computer Research and Development, 2020, 57(12): 2583-2595. (in Chinese)
- [16] JELODAR H, WANG Y, ORJI R, et al. Deep sentiment classification and topic discovery on novel coronavirus or COVID-19 online discussions: NLP using LSTM recurrent neural network approach[J]. IEEE Journal of Biomedical and Health Informatics, 2020, 24(10): 2733-2742.
- [17] LI L, YANG L, ZENG Y. Improving sentiment classification of restaurant reviews with attention-based bi-GRU neural network[J]. Symmetry, 2021, 13(8): 1517.
- [18] GAO Z, FENG A, SONG X, et al. Target-dependent sentiment classification with BERT[J]. IEEE Access, 2019, 7: 154290-154299.
- [19] 叶星鑫,徐杨,罗梦诗. 基于 ALBERT-AFSFN 的中文短文本情感分析[J]. 计算机工程与应用, 2021, 3: 1-11.
- YE X X, XU Y, LUO M S. Sentiment analysis of Chinese short text based on ALBERT-AFSFN[J]. Computer Engineering and Applications, 2021, 3: 1-11. (in Chinese)
- [20] TAM S, SAID R B, TANRIOVER Ö. A ConvBiLSTM deep learning model-based approach for Twitter sentiment classification[J]. IEEE Access, 2021, 9: 41283-41293.
- [21] YAN W, ZHOU L, QIAN Z, et al. Sentiment analysis of student texts using the CNN-BiGRU-AT model [J]. Scientific Programming, 2021, 202: 8405623.
- [22] SINDHU C, SOM B, SINGH S P. Aspect-oriented sentiment classification using BiGRU-CNN model[C]//5th International Conference on Computing Methodologies and Communication (ICCMC). Piscataway: IEEE, 2021: 984-989.
- [23] HUANG F, LI X, YUAN C, et al. Attention-emotion-enhanced convolutional LSTM for sentiment analysis[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(9): 4332-4345.
- [24] WEN Z, CAO J, YANG R, et al. Automatically select emotion for response via personality-affected emotion transition[C]//Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Stroudsburg: ACL, 2021: 5010-5020.
- [25] PENNEBAKER J W, KING L A. Linguistic styles: language use as an individual difference[J]. Journal of Personality and Social Psychology, 1999, 77(6): 1296-1312.
- [26] SUMMER C, BYERS A, BOOCHEVER R, et al. Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets[C]//2012 11th International Conference on Machine Learning and Applications. Piscataway: IEEE, 2012, 2: 386-393.
- [27] DUAN Y, LI H, HE M, et al. A BiGRU autoencoder remaining useful life prediction scheme with attention mechanism and skip connection[J]. IEEE Sensors Journal, 2021, 21(9): 10905-10914.
- [28] MAIRESSE F, WALKER M A, MEHL M R, et al. Using linguistic cues for the automatic recognition of personality in conversation and text[J]. Journal of Artificial Intelligence Research, 2007, 30: 457-500.
- [29] DEMSZKY D, MOVSHOVITZ A D, KO J, et al. GoEmotions: A dataset of fine-grained emotions[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL, 2020: 4040-4054.
- [30] ASGHAR N. Yelp dataset challenge: Review rating prediction[EB/OL]. (2016) [2022]. <https://arxiv.org/abs/1605.05362>.
- [31] YANG J, ZOU X, ZHANG W, et al. Microblog sentiment analysis via embedding social contexts into an attentive LSTM[J]. Engineering Applications of Artificial Intelligence, 2021, 97: 104048.
- [32] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of NAACL-HLT. Stroudsburg: ACL, 2019: 4171-4186.
- [33] LIU Y, OTT M, GOYAL N, et al. RoBERTa: A robustly optimized BERT pretraining approach[EB/OL]. (2019) [2022]. <https://arxiv.org/abs/1907.11692>.
- [34] ZHANG Y F, YU X L, CUI Z Y, et al. Every document owns its structure: Inductive text classification via graph neural networks[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL, 2020: 334-339.

作者简介



王友卫 男,1987年出生,山东临沂人.副教授.主要研究方向为情感分析、数据挖掘、深度学习.
E-mail: ywwang15@126.com



刘瑞 女,1998年出生,安徽淮南人.硕士研究生.主要研究方向为情感分析、深度学习.
E-mail: 1056892410@qq.com



凤丽洲 女,1986年出生,吉林长春人.副教授.主要研究方向为情感分析、内容安全、深度学习.
E-mail: lzfeng15@126.com