



河南理工大学学报(自然科学版)

Journal of Henan Polytechnic University(Natural Science)

ISSN 1673-9787,CN 41-1384/N

《河南理工大学学报(自然科学版)》网络首发论文

题目: 基于多头自注意力和并行混合模型的文本情感分析
作者: 李辉, 黄钰杰
收稿日期: 2019-10-10
网络首发日期: 2020-10-21
引用格式: 李辉, 黄钰杰. 基于多头自注意力和并行混合模型的文本情感分析[J/OL]. 河南理工大学学报(自然科学版).
<https://kns.cnki.net/kcms/detail/41.1384.n.20201020.1350.004.html>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

李辉,黄钰杰.基于多头自注意力和并行混合模型的文本情感分析[J].河南理工大学学报(自然科学版),2021,40(1)
doi:10.16186/j.cnki.1673-9787.2019100022

LI H, HUANG Y J. Sentiment analysis based on multi self-attention and parallel hybrid model[J]. Journal of Henan Polytechnic University(Natural Science), 2021, 40(1). doi:10.16186/j.cnki.1673-9787.2019100022

基于多头自注意力和并行混合模型的文本情感分析

李辉¹, 黄钰杰²

(1. 河南理工大学 物理与电子信息学院, 河南 焦作 454000; 2. 河南理工大学 电气工程与自动化学院, 河南 焦作 454000)

摘要:针对以往研究大多使用单一模型进行文本情感分析,导致无法很好地捕捉相关文本的情感特征,从而造成情感分析效果不理想的问题,提出一种基于多头自注意力和并行混合模型的文本情感分析方法。首先,利用 Word2vec 模型捕捉单词的语义特征,训练词向量;其次,借助双层多头自注意力机制(double layer multi-head self-attention, DLMA)学习文本内部的词依赖关系,捕获其内部结构特征;再次,使用并行的双向门限循环神经网络(Bi-directional gated recurrent unit, BiGRU)获取文本的序列特征;最后,借助改进的并行卷积神经网络(convolutional neural network, CNN)提取深层次特征信息。将该模型分别在2个数据集上进行实验验证,其准确率分别达到92.71%和91.08%,结果表明,该方法比其他模型具有更好的学习性能。

关键词:多头自注意力;双向门限循环神经网络;卷积神经网络;情感分析

中图分类号:TP391.1

文献标志码:A

Sentiment analysis based on multi self-attention and parallel hybrid model

LI Hui¹, HUANG Yujie²

(1. School of Physics and Electronic Information, Henan Polytechnic University, Jiaozuo 454000, Henan, China; 2. School of Electrical Engineering and Automation, Henan Polytechnic University, Jiaozuo 454000, Henan, China)

Abstract: Most of the past studies use a single model for text sentiment analysis, which leads to the inability to capture the emotional features of related texts well, which leads to the problem of unsatisfactory sentiment analysis. This paper proposes a model based on self-attention and parallel hybrid model. Text sentiment analysis method. First, use the Word2vec model to capture the semantic features of words and train word vectors. Secondly, the double layer multi-head self-attention (DLMA) is used to learn the word dependence within the text and capture its internal structural features. The sequence characteristics of the text are then acquired using a parallel Bi-directional Gated Recurrent Unit (BiGRU). Finally, the deep hierarchical feature information is extracted by the improved Convolutional Neural Network (CNN). The model was validated on two data sets, and the accuracy rate reached 92.71% and 91.08%. The experimental results show that the method has better learning performance than other models.

Key words: multi-head self-attention; Bi-directional gated recurrent unit; convolutional neural network; senti-

收稿日期:2019-10-10;修回日期:2019-12-26

基金项目:河南省基础与前沿技术研究计划项目(152300410103)

第一作者简介:李辉(1976—),男,河南林州人,博士,教授,主要从事无线通信技术和智能信号处理等方面的教学和科研工作。
Email:li20022004@hpu.edu.cn

通讯作者简介:黄钰杰(1993—),女,河南郑州人,硕士生,主要从事深度学习、自然语言处理和智能信号处理等研究。Email:1577500669@qq.com

ment analysis

0 引言

近年来,随着互联网产业快速发展,出现了众多新兴媒体,它们通过互联网可以获得用户评论、品牌、情感、政治和观点等大量数据。情感分析是一种特殊的文本挖掘工作,是从给定的文本中提取人们的态度或情感。目前,文本情感分析^[1]是自然语言处理(natural language processing, NLP)领域中的研究热点,它是针对非结构化信息,挖掘其蕴涵的深层次情感或倾向。

目前,有很多使用基于机器学习的情感分析技术进行文本分析,取得了不错的效果^[2-3],但是这些方法不能有效地表达复杂的函数计算,并且需要人工选择数据特征,泛化能力较弱,而深度学习能够从原始数据中自动学习重要的数据特征,并处理各种复杂任务,在建模、解释、表达能力以及优化等方面优势明显。将深度学习应用于文本情感分析领域,极大地推进了文本情感分析的研究与发展^[4]。G. E. Hinton 等^[5]提出了深度学习的概念,即利用深度学习网络构建一个高质量的语言模型处理自然语言问题。在现有的深度学习模型中最热门的网络模型是循环神经网络(recurrent neural network, RNN)模型^[6]和卷积神经网络(convolutional neural network, CNN)模型^[7]。CNN是目前应用最为广泛的一种深度学习结构,具有局部感受野、权值共享和下采样3个特点,使其在捕捉文本局部特征信息上具有很大的优势。它处理的是序列化特征信息,RNN是通过一种链式神经网络对历史信息进行储存和传播,减少网络的训练参数,提升训练速度。N. Kalchbrenner 等^[8]提出一种 DCNN 网络模型,使用一种动态池化的方法,能够处理可变长度的输入;王煜涵等^[9]提出单一的 CNN 结构模型,进行文本情感分析,取得了不错的效果;S. Kai 等^[10]则使用长短期记忆网络(long and short term memory, LSTM)解决情感分析问题,并使分类效果明显提升;I. Augenstein 等^[11]提出的双向 LSTM 网络模型,该模型能同时捕捉词语的前后时序关系,获得词语间的前后关系;K. Cho 等^[12]提出门限循环单元(gated recurrent unit, GRU),GRU 是 LSTM 的改进,将遗忘门和输入门合并成一个更新门,使其结构更加简单,从而减小模型的训练时间;曹宇等^[13]提出双向门限循环神经网络(Bi-directional gated recurrent unit,

BiGRU)进行中文文本情感分析,取得了不错的效果。

考虑到仅使用单一的神经网络使更深层次的特征提取不充分,于是 S. Lai^[14]提出了 RCNN 模型,首先利用双向循环神经网络对序列化的特征进行提取,再经过卷积、池化等操作进行文本情感分类;A. Hassan 等^[15]在 CNN 网络后引入 RNN,构建了混合模型,证明采用多个神经网络模型结合的方式能更好地完成文本情感分析。由于在文本中,每个词或者句子对整个文本的情感分类有着不同的影响,使用注意力机制能够捕捉文本中最具代表性的特征,优化模型结构;王丽亚等^[16]提出 RCNN 并引入注意力机制,有效地提高了文本情感分类的效果。目前,基于注意力的方法已经成功应用于各种任务中,在机器翻译领域为了更准确地提取文本特征,A. Vaswani 等^[17]提出多头自注意力(multi-head self-attention, MA)方法构建了 Transformer 模型;LEI Z 等^[18]采用 MA 对 CNN 处理过的特征矩阵进一步学习情感特定句子表示;关鹏飞等^[19]运用双向 LSTM 捕捉词语间的语义依赖关系,再利用自注意力进一步学习重点词的特征信息。

本文模型结构在文献[14]的基础上进行改进,由于这种传统的 RNN-CNN 模型没有充分考虑序列内部的关键信息对情感分类的影响,于是本文模型结构在此基础上结合文献[17]加入 MA 机制,构建基于 DLMA-BiGRU-CNN 模型的文本情感模型,区别于传统的 RNN-CNN 引入注意力的方法。此外,针对文献[17]单一的 MA 层进行改进,采用双层的 MA 表示句子,学习序列内部的词依赖关系,使文本分类的准确率得到提升。考虑到神经网络存在特征抽取不充分或无法学习到更深层的语义信息的问题,对文献[14]的模型进行改进,使用并行输入的方式,充分提取文本信息,从而使文本分类的效果更好。

1 相关研究工作

1.1 BiGRU

由于标准 RNN 无法学习到句子的长期依赖关系,容易出现梯度爆炸或梯度消失问题,且单向的 GRU^[20]只能获取上文的依赖信息,无法充分考虑到下文的特征信息,造成在文本建模中存在明显缺陷,于是本文使用并行的 BiGRU 对文本进

行特征提取。如图 1 所示,BiGRU 是由一个正向 GRU 和一个反向 GRU 构成,GRU 模型是 LSTM 的一个变体,相对 LSTM 而言,GRU 具有结构简单、计算效率更高、处理速度更快等优点。与具有 3 个门的 LSTM 不同,GRU 结构只有 2 个门,即重置门和更新门。

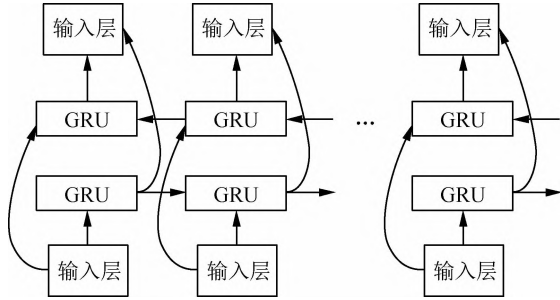


图 1 GRU 结构图

Fig.1 GRU structure diagram

GRU^[21]模型通过门控机制学习文本的序列特征,控制信息的传递与输出,不仅可以保持信息不受时间影响,而且可以删除与预测无关的信息。重置门将新的输入与以前的存储相结合,而更新门则决定了要保留多少以前的存储。所以 GRU 的具体工作原理公式为

$$r_t = \sigma(\mathbf{w}_r \cdot [h_{t-1}, x_t]), \quad (1)$$

$$z_t = \sigma(\mathbf{w}_z \cdot [h_{t-1}, x_t]), \quad (2)$$

$$\hat{h}_t = \tanh(\mathbf{w} \cdot r_t \times h_{t-1}, x_t), \quad (3)$$

$$h_t = (1 - z_t) \times h_{t-1} + z_t \times \hat{h}_t, \quad (4)$$

式中: x_t 为输入数据; h_t 为最终的输出; z_t 为更新门; r_t 为重置门; σ 为激活函数 sigmoid,函数在 $[0,1]$ 内取值; \tanh 为双曲正切函数; \mathbf{w}_r , \mathbf{w}_z 和 \mathbf{w} 分别为隐藏层、重置门和更新门的权重矩阵;下标 t 为当前时刻;下标 $t-1$ 为前一时刻; h_{t-1} 为前一时刻的细胞隐藏状态。

融合前置 GRU 和后置 GRU 的输出,其融合计算公式为

$$h_i = [\vec{h}_i, \overleftarrow{h}_i], \quad (5)$$

式中: \vec{h}_i 和 \overleftarrow{h}_i 为前置 GRU 和后置 GRU 的输出结果; h_i 为前置 GRU 和后置 GRU 输出拼接的结果,即最后的 BiGRU 输出结果。

1.2 CNN

CNN 是基于多层网络结构训练的监督学习算法,是一种前馈人工神经网络^[22]。CNN 采用一系列卷积层、池化层和全连接层构造模型。

卷积层:卷积层的结构是由不同单独层组成的,其核心操作就是利用卷积核对输入特征进行运算。卷积过程的实质是滤波器在输入数据特征

上滑动计算的过程,用来对文本词语间的信息进行捕捉。首先将词向量矩阵输入到 DCNN 网络中,其词向量矩阵公式为

$$\mathbf{x}_i = \mathbf{x}_1 \oplus \mathbf{x}_2 \oplus \mathbf{x}_3 \oplus \cdots \oplus \mathbf{x}_l, \quad (6)$$

式中: \mathbf{x}_i 为一个句子里第 i 个单词对应的词向量;下标 l 为句子的长度; \oplus 为连接运算符。

特征表达式为

$$c_i = f(\mathbf{w} \cdot \mathbf{x}_{i+h-1} + \mathbf{b}), \quad (7)$$

式中: \mathbf{w} 为高度 h 、宽度 d 、窗口 $h \times d$ 的卷积核; \mathbf{b} 为偏置向量; f 为激活函数 Relu。

经过卷积操作后,得到特征图 \mathbf{c} ,

$$\mathbf{c} = [c_1, c_2, c_3, \cdots, c_l]. \quad (8)$$

池化层:在经过卷积层之后,得到了很多特征维度较高的特征图。在卷积操作后进行最大池化操作,降低特征维度,保持特征在空间与结构上不变,减小特征图 \mathbf{c} 的大小,并简化卷积层输出的信息。其池化表达式为

$$\hat{h} = \max(\mathbf{c}). \quad (9)$$

全连接层:将提取到的局部特征信息融合为全局特征,最后使用 softmax 分类函数计算概率大小,得到文本的情感倾向。

2 本文模型

2.1 模型提出

本文提出一种基于 DLMA-BiGRU-CNN 模型的文本情感分析方法,模型构建如图 2 所示。模型包括文本输入层、DLMA 层、并行 BiGRU 层、并行 CNN 层和文本分类层。首先,在输入层利用 word2vec 模型捕捉单词的语义特征并训练词向量;其次,在 DLMA 层借助 DLMA 学习文本内部的词依赖关系;再次,在并行 BiGRU 层使用并行的 BiGRU 模型获取文本的序列特征;最后,并行 CNN 层借助并行 CNN 模型提取深层次特征信息,在分类层完成文本情感分析。

2.2 文本输入层

首先使用结巴(Jieba)分词对词语进行精准的分词处理,其次利用 Word2vec 模型训练词向量。Word2vec 是由 Google 在 2013 年提出的深度学习模型。Word2vec 模型包括 CBOW 和 Skip-gram 两种创建词向量的模型,CBOW 模型基于周围的上下文词语计算某个词出现的概率,Skip-gram 模型根据目标词计算前后出现几个词的概率。本文选用 Skip-gram 模型对本文词向量进行训练,并将单词转换为高维向量,作为模型的输入数据。将训练文本中句子的最大长度设置为 l ,作

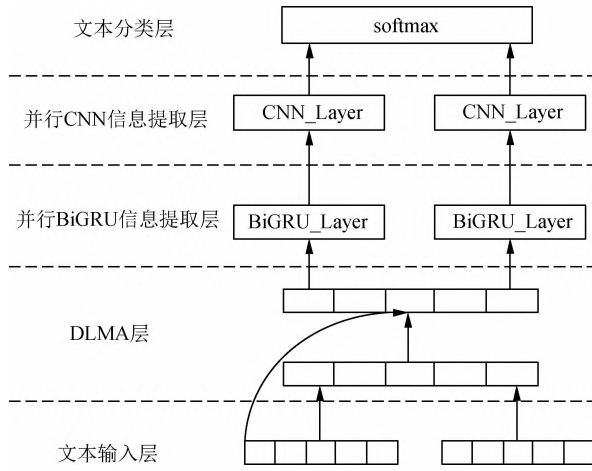


图2 基于 DLMA-BiGRU-CNN 的文本情感分析模型

Fig.2 Text sentiment analysis model based on DLMA-BiGRU-CNN

为模型的输入大小,将每个句子表示成大小为 $l \times d_k$ 的文本词向量矩阵 E ,其表达式为

$$E = [e_1, e_2, e_3, e_4, \dots, e_l]^T, \quad (10)$$

式中: e_i 为句子中第 i ($1 < i < l$) 个词对应的词向量; d_k 为词向量的维度。

2.3 DLMA 特征提取层

多头注意力机制的目的是为了从多方面捕获序列的关键信息,目前 MA 已经在各个领域都有很多应用。本文将双层的 MA 模型作为一个单独的层,有别于传统的 CNN 结合注意力的方法,将双层的 MA 模型置于词向量层后,采用双层 MA 学习序列内部的词依赖关系。工作原理如下。

(1) 第一层 MA 层。将文本的词向量矩阵 E 输送到第一层 MA 层进行句子表示学习,从而捕获句子的长距离依赖性和学习不同位置的不同语义空间的序列信息。首先通过 h 个不同的线性变换,将词向量矩阵 E 分别转换为相同维度的 Q 和 V ,表达式为

$$Q = EW_i^Q, \quad (11)$$

$$K = EW_i^K, \quad (12)$$

$$V = EW_i^V, \quad (13)$$

式中: Q 为查询向量序列; K 为键向量序列; V 为值向量序列; W_i^Q 、 W_i^K 和 W_i^V 为第 i 个线性表示的参数矩阵。

同时进行 h 次自注意力计算,单头自注意力进行逐个向量计算,将 $q_t \in Q$ ($1 \leq t \leq l$) 和各个 $k_s \in K$ ($1 \leq s \leq l$) 进行内积运算,并利用 softmax 获得 $q_t \in Q$ ($1 \leq t \leq l$) 和各个 $v_s \in V$ ($1 \leq s \leq l$) 的相似度,进行加权求和计算,得到一个 d_v 维度的向量,表达式为

$$SDA_i(q_t, K, V) = \sum_{s=1}^l \frac{1}{Z} \exp\left(\frac{(q_t, k_s)}{\sqrt{d_k}}\right) v_s, \quad (14)$$

其中, Z 为归一化因子。

通过计算单头自注意力,得到一个输出,即

$$head_i = SDA_i(Q, K, V) = \sum_{s=1}^l a_{is} v_s, \quad (15)$$

式中, a_{is} 为第 i 次单头自注意力计算得到的 v_s 的权重系数。

拼接 h 次单头自注意力计算结果,并进行线性变换,得到多头自注意力结果。形成第一层 MA 句子表示,即

$$MA_1 = \text{concat}(head_1, head_2, \dots, head_h) w_o, \quad (16)$$

式中, w_o 为输入空间到输出空间的映射参数。

(2) 第二层 MA 层。将第一层得到的 MA_1 和词向量矩阵 E 进行拼接后得到 N_1 ,

$$N_1 = \text{concat}(MA_1, E). \quad (17)$$

将 N_1 输入到第二层 MA 层,提取出文本更高层次的特征信息,得到 DLMA 的输出结果:

$$MA_2 = \text{MultiHead}(N_1, N_1, N_1). \quad (18)$$

2.4 并行 BiGRU 信息提取层

在需要输入上下文的情况下, BiGRU 在获取特征信息方面比 LSTM 更加有效,用 BiGRU 模型文本信息不仅使用 2 种隐藏状态从后向前传递,而且还使用 2 种隐藏状态从前向后传递。由于单一的 BiGRU 模型随着网络层数增加性能将受到影响,不能够充分地提取文本特征,因此,为了解决单通道模型提取信息不充分和受无关信息干扰较大的问题,本文使用并行的 BiGRU 模型获得全局的深层信息,提高网络的收敛性。

将经过 DLMA 的输出结果 MA_2 输入到并行的 BiGRU 模型中,进行式(1)~(5)序列化的特征提取,得到并行的 BiGRU 输出 h_{1t} 和 h_{2t} ,之后 h_{1t} 和 h_{2t} 作为 CNN 层的输入进行深层次的特征提取。

2.5 并行 CNN 特征提取层

由于 BiGRU 主要处理的是时间序列问题,考虑到不同的上下文组成也有可能造成不同的语义表达,为了更大程度地提取不同粒度的层次化文本特征,本文在并行 BiGRU 层后接入并行 CNN 层,处理深层次的特征向量。并行 CNN 特征提取部分由 2 个的卷积神经网络(CNN)组成,且每个 CNN 由卷积层、池化层以及全连接层 3 部分组成。

在卷积层,对 h_{1t} 和 h_{2t} 使用多个卷积核进行卷积处理,每个 CNN 的卷积层是由 3 个 3×3 , $4 \times$

4,5×5 的滤波器构成,且滤波器每次的滑动步长为 1,这样经过一系列的卷积操作后就可得到维度较高的特征图 c_1 和 c_2 。在池化层进行最大池化操作,得到 c'_1 和 c'_2 。在全连接层将 c'_1 和 c'_2 进行融合,得到 $v_{a,i}$ 。

2.6 分类层

在经过上述特征后提取,在分类层使用 softmax 函数计算文本情感分类的概率大小,即

$$p = \text{softmax}(\mathbf{w}_2 v_{a,i} + \mathbf{b}_2), \quad (19)$$

式中, \mathbf{w}_2 和 \mathbf{b}_2 分别为权重矩阵和偏置矩阵。

本文模型训练采用反向传播方式更新参数,目标函数为交叉熵代价函数, y 为文本真实的情感类别值, \hat{y} 为模型所预测的情感类别值,目标函数定义为

$$\text{loss} = -\sum y^{(i)} \ln \hat{y}^{(i)} + (1 - y^{(i)}) \ln (1 - \hat{y}^{(i)}). \quad (20)$$

由式(20)可知,若训练文本标签为 1,目标函数简化为 $\text{loss} = -\ln \hat{y}$,预测文本类别越接近 1,目标函数值越接近于 0。同理,若训练文本标签为 0,目标函数简化为 $\text{loss} = -\ln (1 - \hat{y})$,预测文本类别越接近于 0,目标函数越小。为了防止深度学习网络训练过程中出现过拟合现象,通过放弃部分权重的修改以提高模型的泛化能力,使用 Dropout 策略,使模型学习速度得到提高,同时提升模型的收敛性效果。

3 实验

3.1 实验环境

为了确保实验的准确性,使用 Python 3.5 和 Keras 深度学习开源框架进行实验。本文使用的 2 个数据集分别如下。

SST-2:斯坦福情感数据库二元分类版本,只有 2 种标签,非常积极和积极的评论被标记为正面,消极和非常消极的评论被标记为负面。

自然语言处理及中文计算会议(natural language processing and Chinese computing, NLPCC) 2014 数据集:该语料共包含中文和英文 2 种语言,本文使用其中的中文数据。

如表 1 所示,在文本情感分类时将文本分成

表 1 实验数据设置

Tab. 1 Experimental data settings

数据集	积极/个	消极/个
SST-2	7 531	7 400
NIPCC	5 000	5 000

2 类,即积极和消极。每个数据集都按 8:2 分成

训练集和测试集。

3.2 实验参数

在实验过程中要调试参数,因为参数的不同就会导致结果出现差异,为保证网络输入词向量矩阵的大小统一,将文本的训练样本数定为 500,序列长度定为 250,长度不足的进行补零操作,长度超出的截断。词向量维度设置为 128。对于并行 BiGRU 神经网络,每个隐层神经元个数设置为 128;对于并行 CNN 神经网络,使用不同尺寸的卷积核,设定为(3,4,5),每个尺寸的卷积核个数设置为 128。为了防止过拟合,设定 Dropout 率为 0.5,批处理大小(batch_size)为 64,迭代次数(epoch)为 20。

3.3 模型分类标准

为了验证模型的可靠性,设置了多个模型进行对比。在本文实验中将准确率(Accuracy)以及 F-Score 作为模型分类的评价标准。其计算式为

$$\text{Accuracy} = \sum_{i=1}^N |y_i - \hat{y}| / N, \quad (21)$$

$$\text{Precision} = \sum_{c_i \in c} \text{True}(c_i) / \sum_{c_i \in c} \text{Doc}(c_i), \quad (22)$$

$$\text{Recall} = \sum_{c_i \in c} \text{True}(c_i) / \sum_{c_i \in c} \text{Response}(c_i), \quad (23)$$

$$F\text{-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\%, \quad (24)$$

式中: y_i 为文本真实的情感类别值; \hat{y} 为模型所预测的情感类别值; $\text{True}(c_i)$ 为分类 c_i 且分类正确的样本数; c_i 为实验中分类 $\text{Doc}(c_i)$ 的所有样本数; $\text{Response}(c_i)$ 为文本中实际分类 c_i 的文本数。

3.4 结果分析

为了验证本文模型的有效性,设置了多个对比模型,将模型参数进行微调,如表 2 所示。经过实验,得出各个模型情感分析结果,模型如下。

(1) BiGRU:由文献[14]提出的单一双向 GRU 网络模型。

(2) CNN:由文献[10]提出的单一卷积神经网络模型。

(3) BiRNN-DCNN:由文献[15]提出的结合双向 LSTM 与 CNN 的模型。

(4) DLMA-BiGRU:该模型是由一个双层多头自注意力模型和一个双向门限循环神经网络组成,对文本进行情感分析。

(5) DLMA-CNN:该模型是由一个双层多头自注意力模型和一个卷积神经网络组成,对文本进行情感分析。

(6) ATT-BiGRU-CNN 模型:由文献[17]提出基于 BiGRU-CNN-ATT 模型的文本情感分析模

型。

(7) DLMA-BiGRU-CNN: 本文提出的基于双层多头自注意力机制和并行 BiGRU-CNN 的文本情感分析模型。

表 2 5 个模型情感分析结果

Tab. 2 Emotional analysis results of five models

数据集	模型	准确率/%	F-Score/%
SST-2	BiGRU	87.39	88.30
	CNN	88.69	89.92
	BiRNN-DCNN	90.34	90.32
	DLMA-BiGRU	88.87	89.45
	DLMA-CNN	89.08	88.76
	ATT-BiGRU-CNN	91.19	90.66
	DLMA-BiGRU-CNN	92.71	91.34
NIPCC	BiGRU	87.24	88.75
	CNN	88.92	89.16
	BiRNN-DCNN	89.55	88.69
	DLMA-BiGRU	90.00	90.30
	DLMA-CNN	89.77	89.01
	ATT-BiGRU-CNN	90.34	89.92
	DLMA-BiGRU-CNN	91.08	91.10

通过对比 2 个数据集可知, 本文所提模型在准确率和 F -Score 上较其他网络模型有显著提升, 模型准确率分别为 92.71% 和 91.08%。

将单一模型 CNN 或 BiGRU 模型与 BiRNN-DCNN 模型对比, 发现将多个模型的优势结合, 混合模型使用了 BiRNN 学习文本序列信息, 并结合 CNN 获取局部信息特征, 更加充分地提取了文本特征, 丰富了特征向量, 比单一模型分类效果有明显提高。

将单一模型 CNN 或者 BiGRU 模型与 DLMA-BiGRU 模型或者 DLMA-CNN 模型进行对比, 发现使用双层多头自注意力模型比单一模型具有更好的文本情感分类效果。

BiRNN-DCNN 混合模型相对于 ATT-BiGRU-CNN 模型可以看出, 其在正确率上有显著提升。这是因为在混合模型中引入注意力模型, 使用注意力层能够提取文本的重要特征, 忽略掉不重要的特征, 使分类结果受无关词语的影响尽可能小, 从而提升了正确率和分类性能。

将 ATT-BiGRU-CNN 混合模型和本文 ATT-BiGRU-DCNN 模型对比, 可以看出, 引入双层多头自注意力机制后, 模型性能得到显著提升。因为双层多头自注意力机制区别于传统的注意力模

型, 不仅对浅层的词向量有更好的关注, 而且更好地获取深层次的句子向量, 提升重要部分对文本情感分类的影响力, 减小无关信息的干扰, 从而在准确率上能够有显著提高。

为了更加直观地显示不同模型准确率和 F -Score 的变化趋势, 将各个模型在 NIPCC 数据集上 ACC 变化和 F -Score 变化情况绘制成如图 3~4 所示的折线图, 将各个模型在 SST-2 数据集上 ACC 变化和 F -Score 变化情况绘制成如图 5~6 所示的折线图。

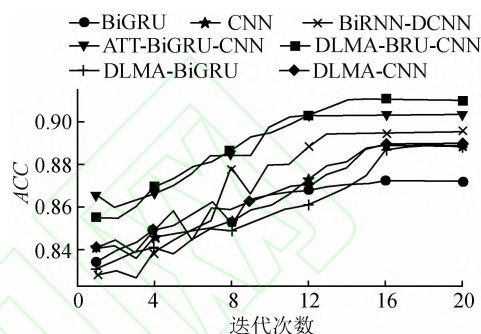


图 3 各个模型在 NIPCC 数据集上的 ACC 折线

Fig. 3 ACC line chart of each model on the NIPCC dataset

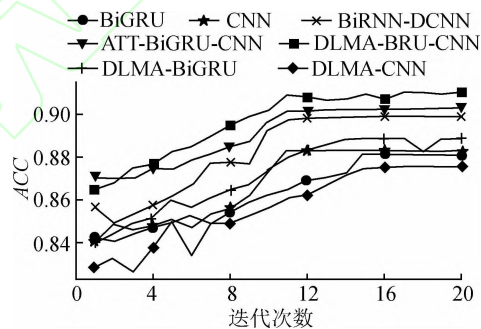


图 4 各个模型在 NIPCC 数据集上的 F -Score 折线

Fig. 4 F -Score line chart of each model on the NIPCC dataset

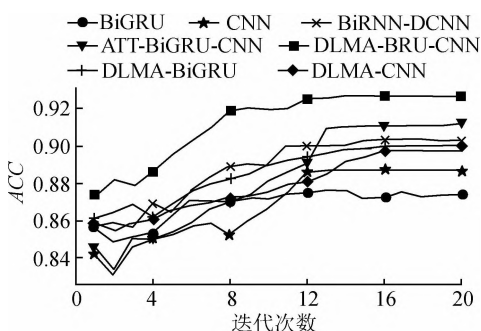


图 5 各个模型在 SST-2 数据集上的 ACC 折线

Fig. 5 ACC line chart of each model on the SST-2 dataset

由图 3~6 可以看出, 本文提出的基于 DLMA-BiGRU-CNN 的文本情感分析模型在准确率和 F -Score 上都有明显提高, 证明了本文模型的有效性和优越性。

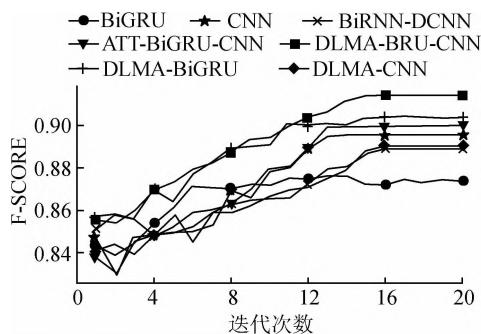


图 6 各个模型在 SST-2 数据集上的 F -Score 折线

Fig. 6 F -Score line chart of each model on the SST-2 dataset

4 结 论

本文提出一种基于多头自注意力和并行混合模型的文本情感分析方法,通过实验对比,该模型在准确率和 F -Score 上都有明显提高,证明了本文模型的优越性。由于神经网络结构的复杂性,下一步工作将继续深入研究模型,并考虑对文本进行多分类或者进行更细粒度的分析,以更加贴合文本的原始情感。

参考文献:

- [1] PANG B, LILLIAN L, VAITHYANATHAN S. Thumbs up? sentiment classification using machine learning techniques [C]//Empirical Methods in Natural Language Processing (EMNLP), philadelphia, july, 2002: 79-86.
- [2] 霍林,陆寅丽.改进粒子群算法应用于 Android 恶意应用检测[J].计算机工程与应用,2020,56(7):69-101.
HUO L, LU Y L. Improved particle swarm optimization algorithm for Android malicious application detection [J]. Computer Engineering and Applications, 2020, 56 (7): 69-101.
- [3] 李婷婷,姬东鸿.基于 SVM 和 CRF 多特征组合的微博情感分析[J].计算机应用研究,2015,32(4): 978-981.
LI T T, JI D H. Spiritual analysis of Weibo based on SVM and CRF Multi-feature combination[J]. Application Research of Computers, 2015, 32(4): 978-981.
- [4] NI X, XUE G R, LING X, et al. Exploring in the weblog space by detecting informative and affective articles[C]// International World Wide Web Conference, 2007: 281-290.
- [5] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313(5786): 504-507.
- [6] 郝志峰,黄浩,蔡瑞初,等.基于多特征融合与双向 RNN 的细粒度意见分析[J].计算机工程,2018,44 (7): 199-204.
- [7] HAO Z F, HUANG H, CAI R C, et al. A fine-grained opinion analysis based on multi-feature fusion and bidirectional RNN [J]. Computer Engineering, 2018, 44 (7): 199-204.
- [8] GICHAG L, JEONG J, SEUNGWAN S, et al. Sentiment classification with word localization based on weakly supervised learning with a convolutional neural network [J]. Knowledge-Based Systems, 2018, 152(7): 70-82.
- [9] KALCHBRENNER N, GREFFENSTETTE E, BLUNSOM P. A convolutional neural network for modeling sentence [C]// Proc of the 52nd Annual Meeting of the Association for Computational, 2014: 655-665.
- [10] 王煜涵,张春云,赵宝林,等.卷积神经网络下的 Twitter 文本情感分析[J].数据采集与处理,2018, 33(5): 921-927.
- [11] WANG Y H, ZHANG C Y, ZHAO B L, et al. Sentiment analysis of twitter text under convolutional neural networks [J]. Data Acquisition and Processing, 2018, 33 (5): 921-927.
- [12] KAI S, RICHARD S, CHRISTOPHE D M. Improved semi-anti representations from tree-structured long short-term memory networks [C]//Association for Computational Linguistics Beijing, China, 2015: 1556-1566.
- [13] AUGENSTEIN I, ROCKTASCH T, MACHOS A, et al. Stance detection with bidirectional conditional encoding [C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, Texas: ACL, 2016: 876-885.
- [14] CHO K, VAN M B, GUL C C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation [C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language, 2014: 1724-1734.
- [15] 曹宇,李天瑞,贾真,等. BGRU: 中文文本情感分析的新方法 [J]. 计算机科学与探索, 2019, 13(6): 973-987.
- [16] CAO Y, LI T R, JIA Z, et al. BGRU: A new method of Chinese text sentiment analysis [J/OL]. Computer Science and Exploration, 2019, 13(6): 973-987.
- [17] LAI S. Recurrent convolutional neural networks for text classification [C]//Proceeding of AAAI, 2015 (333): 2267-2273.
- [18] HASSAN A, MAHMOOD A. Efficient deep learning model for text classification based on recurrent and convolutional layers [C]//IEEE International Conference on Machine Learning and Applications, 2018.
- [19] 王丽亚,刘昌辉,蔡敦波,等.基于联合网络引入注

- 意力机制的中文文本情感分析[J]. 计算机应用, 2019, 39(10): 2841-2846.
- WANG L Y, LIU C H, CAI D B, et al. Chinese text sentiment analysis based on joint network introduction attention mechanism [J/OL]. Computer Application, 2019, 39(10): 2841-2846.
- [17] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. arxiv, 2017.
- [18] LEI Z, YANG Y, YANG M. SAAN: A sentiment-aware attention network for sentiment analysis[C]//The 41th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), 2018.
- [19] 关鹏飞, 李宝安, 吕学强, 等. 注意力增强的双向 LSTM 情感分析[J]. 中文信息学报, 2019, 33(2): 105-111.
- GUAN P F, LI B A, LU X Q, et al. Bidirectional LSTM sentiment analysis with attention enhancement [J]. Journal of Chinese Information Processing, 2019, 33(2): 105-111.
- [20] ZHANG Y, LIAO J, TANG J, et al. Extractive document summarization based on hierarchical GRU [C]//International Conference on Robots & Intelligent System. IEEE, 2018: 342-346.
- [21] 李丽双, 周安桥, 刘阳, 等. 基于动态注意力 GRU 的特定目标情感分类[J]. 中国科学: 信息科学, 2019, 49(8): 1019-1030.
- LI L S, ZHOU A Q, LIU Y, et al. Specific target sentiment classification based on dynamic attention GRU [J]. Chinese Science: Information Science, 2019, 49(8): 1019-1030.
- [22] GICHAG L, JEONG J, SEUNGWAN S, et al. Sentiment classification with word localization based on weakly supervised learning with a convolutional neural network [J]. Knowledge Based Systems, 2018, 152(7): 70-82.

(责任编辑 李文清)