

文章编号: 2095-2163(2022)05-0016-07

中图分类号: TP301.6

文献标志码: A

基于 BERT 的双通道神经网络模型文本情感分析研究

严驰腾, 何利力

(浙江理工大学 信息学院, 杭州 310018)

摘要: 针对当前情感分析任务中使用 Word2Vec、GloVe 等模型生成的文本词向量, 无法有效解决多义词表征、经典神经网络模型无法充分提取文本语义特征等问题, 本文提出基于 BERT 的双通道神经网络模型文本情感分析方法。该方法采用 BERT 模型生成词向量, BERT 模型对下游分类任务进行微调的过程中生成文本词向量的动态表征。然后, 将词向量输入由 CNN 与 BiGRU 构建的双通道模型进行特征提取, 并行获取文本的局部与全局语义特征, 并通过注意力机制为输出特征分配相应的权重分值, 突出文本的情感极性。最后将双通道输出特征融合进行情感分类。在酒店评论数据集上进行实验, 结果表明本文模型与文本情感分析的基线模型相比, 在准确率与 F_1 分值上分别提高了 3.7% 和 5.1%。

关键词: 文本情感分析; BERT 模型; 卷积神经网络(CNN); 双向门控制循环单元(BiGRU); 注意力机制

Research on text sentiment analysis of dual-channel neural network model based on BERT

YAN Chiteng, HE Lili

(School of Informatics Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China)

[Abstract] Aiming at the current sentiment analysis tasks using the text word vector generated by Word2vec, GloVe and other models, it cannot effectively solve the problems of polysemous word representation and that the classic neural network model cannot fully extract the semantic features of the text. This paper proposes a dual-channel neural network model text sentiment analysis based on BERT method. This method uses the BERT model to generate word vectors, and the BERT model generates dynamic representations of text word vectors in the process of fine-tuning downstream classification tasks. Then, the word vector is input to the dual-channel model constructed by CNN and BiGRU for feature extraction, the local and global semantic features of the text are obtained in parallel, and the corresponding weight scores are assigned to the output features through the attention mechanism to highlight the emotional polarity of the text. Finally, the dual-channel output features are fused to perform sentiment classification. Experiments on the hotel review data set show that compared with the baseline model of text sentiment analysis, the accuracy and F_1 score of this model are improved by 3.7% and 5.1%, respectively.

[Key words] text sentiment analysis; BERT model; Convolutional Neural Network; Bidirectional Gate Control Recurrent unit; attention mechanism

0 引言

随着移动互联网的高速发展, 众多互联网用户积极地参与到信息的发布中, 可表现为微博评论、商品评论、直播评论等带有个人情感观点的文本信息。针对这些主观性言论进行情感分析研究可以挖掘出极具价值的信息, 如对社会热点新闻评论进行舆情分析, 把握舆论倾向、对商品评论进行消费倾向分析, 获取消费者的消费意向等。

文本情感分析是对带有情感色彩的主观性文本进行分析、推理和归纳的过程, 其目的是挖掘文本信

息中用户的情感信息。现有的情感分析方法主要分为传统方法、机器学习方法和深度学习方法。其中, 传统的方法主要利用情感词典获取文本中情感词的情感值, 再通过加权计算确定文本的整体情感倾向。这种方法严重依赖于情感词典和评判规则的质量, 其优劣程度取决于人工设计与先验知识, 对新词的扩展性差。基于机器学习的方法是指通过计算机对文本情感数据进行学习, 提取数据特征, 自主进行情感分类。常用的机器学习算法有 K 近邻、朴素贝叶斯、支持向量机和决策树等, 与传统方法相比在分类效果上有一定的提升, 但是泛化能力较差。深度学

基金项目: 国家重点研发计划项目(2018YFB1700702)。

作者简介: 严驰腾(1997-), 男, 硕士研究生, 主要研究方向: 自然语言处理; 何利力(1966-), 男, 博士, 教授, 博士生导师, 主要研究方向: 数据分析、企业智能。

通讯作者: 何利力 Email: llhe@zju.edu.cn

收稿日期: 2021-12-09

习是机器学习的分支,通过深度学习算法可以获得文本数据的连续且稠密的向量表示,再利用卷积神经网络(Convolutional Neural Networks, CNN)、循环神经网络(Recurrent Neural Network, RNN)等模型提取数据特征进行文本分类,这种方法可以更加充分地获取文本中蕴含的情感特征,分类效果更好。

1 相关工作

在文本情感分析任务中,Kaur 等人^[1]使用 N-Gram 算法对文本信息进行特征提取,KNN 算法对输入特征进行分类计算得到积极、消极以及中性的情感结果。Kim^[2]提出 3 种尺寸大小卷积核的 CNN 模型,在预训练好的词向量上进行特征提取,此后通过最大池化与全连接层得到分类结果,由于 CNN 模型只能提取文本的局部特征,因此长文本序列特征难以捕获。Bahdanau 等人^[3]最早将注意力机制应用在机器翻译领域,接下来 Basiri 等人^[4]、Liu 等人^[5]、Yang 等人^[6]将注意力机制与神经网络相结合,增强了算法模型对文本中关键情感信息的感知能力,进一步提升了文本情感分析的准确率。

文本的词向量表示是文本分类任务的基石,词嵌入方法最早使用 ont-hot、bag-of-words 等模型对文本进行离散表示,后来基于浅层神经网络模型的 Word2Vec 和 GloVe 等方法,可以通过语料数据训练出文本的连续且稠密的词向量表示,与早期方法相比降低了词向量维度,丰富了词向量的语义,同时能够捕捉词与词之间的相似性问题。Rhanoui 等人^[7]使用 Word2Vec 词嵌入技术对文本数据进行向量化表示,然后通过神经网络的模型在 French newspapers 数据集进行情感分析,该模型中 LSTM 用于捕捉数据的时间序列特征,CNN 对数据的局部特征进行提

取,实验结果表明双通道特征提取优于单模型的数据特征提取。基于浅层神经网络模型的词嵌入方法是通过大量数据实现对词的唯一向量表示,无法解决不同语境中词语包含的语义信息,例如“小张考试得了 92 分,可真厉害。”、“小李考试得了 32 分,可真厉害!”,“厉害”在两句话中分别呈现褒义和贬义的含义。2018 年,Devlin 等人^[8]提出了 BERT 模型,该模型采用双向多层 Transformer 编码器结构在大型文本语料库进行预训练,通过预训练的词向量模型在下游 NLP 任务中对 BERT 模型进行微调,实现不同语义环境下词向量的动态表示,从而解决词向量的静态表征问题。湛志群等人^[9]提出基于 BERT 和双向 LSTM 的微博评论情感分析模型,该模型通过 BERT 生成词向量,采用双向 LSTM 模型进行情感分析,实验结果表明,该模型与基于 Word2Vec 的双向 LSTM 相比准确率更高。

基于上述工作,本文提出基于 BERT 的双通道神经网络模型的文本情感分析方法,该方法采用 BERT(Bidirectional Encoder Representations from Transformers, BERT)模型作为词嵌入层获取文本的向量表示,构建 CNN 与 BiGRU 双通道网络模型对词向量进行局部特征和全局特征的提取,利用注意力机制提高特征矩阵中关键的情感信息权重,使得分类效果更加准确。最后本文通过对比实验证明了该方法的有效性。

2 基于 BERT 的 CNN-BiGRU-AT 情感分析模型

本文提出结合 BERT 的 CNN-BiGRU-AT 情感分析模型,模型结构如图 1 所示。该模型由以下部分组成。

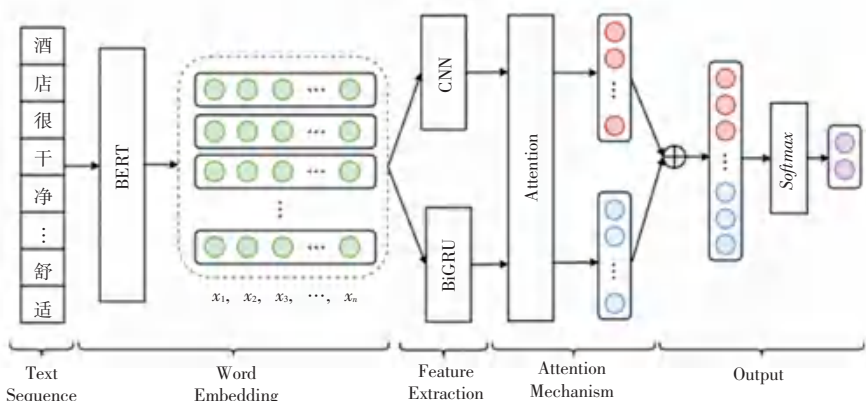


图 1 基于 BERT 的 CNN-BiGRU-AT 模型结构图

Fig. 1 BERT-based CNN-BiGRU-AT model structure diagram

(1)词嵌入层:本文使用 BERT 预训练模型获取文本序列的动态词向量表示。

(2)特征提取层:CNN 能够提取文本的局部特征信息,而 BiGRU 能够根据文本序列的上下文语义提取到全局特征信息,因此本文通过构建 CNN 与 BiGRU 双通道模型进行特征提取。

(3)注意力层:注意力机制能够对特征提取层输出的特征矩阵进行权重分配,突出关键词在文本序列中的重要程度。本文对局部特征矩阵和全局特征矩阵分别进行注意力计算。

(4)输出层:对注意力层输出的特征向量进行融合,使用 *dropout* 正则化防止出现过拟合现象,并通过全连接神经网络计算后,利用 *softmax* 函数进行分类。

2.1 词嵌入层

BERT (Bidirectional Encoder Representations from Transformer)是 Devlin 等人^[8]于 2018 年提出的基于双向 Transformer 编码器结构进行编码的词向量预训练模型。BERT 采用遮挡语言模型 (Masked Language Model, MLM) 和下一句预测 (Next Sentence Prediction, NSP) 来进行词向量训练。MLM 会随机选择一定比例的词元作为预测的遮蔽词元,模型通过全局上下文来学习被遮挡的词元。尽管 MLM 能够编码双向上下文来表示字词,但是不能够显式地表达文本对之间的逻辑关系,而 NSP 可以看作是句子级别的二分类问题,通过判断后一个句子是不是前一个句子合理的下一句来挖掘句子间的逻辑关系,因此 BERT 模型通过结合 MLM 与 NSP 两种方法实现词向量的表示与语义特征的提取。BERT 模型结构如图 2 所示。

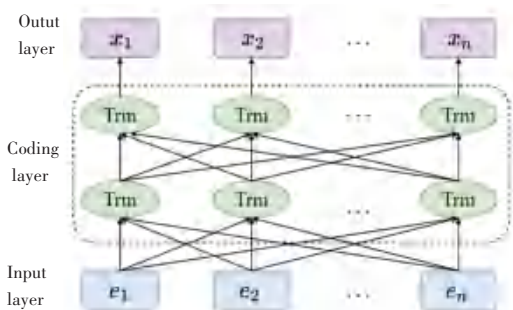


图 2 BERT 模型结构图

Fig. 2 BERT model structure diagram

图 2 中, e_1, e_2, \dots, e_n 为 BERT 模型的输入序列, Trm 为 Transformer 的 Encoder 模型, x_1, x_2, \dots, x_n 为 BERT 模型的输出词向量序列。BERT 的输入序列是词元嵌入 (Token Embeddings)、片段嵌入

(Segment Embeddings) 和位置嵌入 (Positional Embeddings) 信息之和,其中词元嵌入的文本序列需要使用词元 $\langle \text{cls} \rangle$ 作为起始标记。其输入序列模型如图 3 所示。

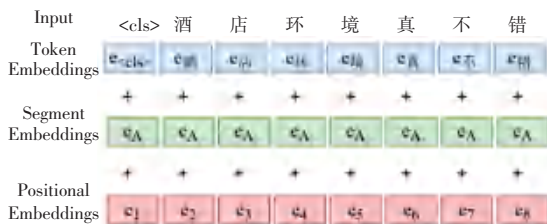


图 3 BERT 模型输入序列图

Fig. 3 BERT model input sequence diagram

2.2 双通道特征提取层

2.2.1 局部语义特征提取

卷积神经网络早期被应用于计算机视觉领域,但近年来也广泛应用到自然语言处理领域中,并取得了不错的成绩。CNN 通过卷积核提取文本的局部语义特征,本文使用多尺度卷积来获取不同距离词语间的特征信息。CNN 通过卷积层对文本进行特征提取后,通常会使用最大池化层对特征矩阵进行压缩,这种方法会破坏文本序列的时序性,造成特征丢失等问题。因此本文对卷积层得到的特征矩阵进行拼接,直接输入注意力层进行权重分配,模型结构如图 4 所示。

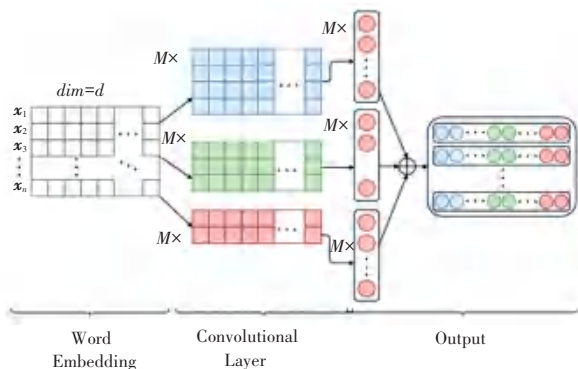


图 4 局部特征提取模型结构图

Fig. 4 Local feature extraction model structure diagram

图 4 中, x_1, x_2, \dots, x_n 为文本的词嵌入向量,通过 3 个不同窗口大小的卷积核进行特征提取,每种卷积核的维度 $dim = 768$ 、通道数 $M = 256$,步长设定为 1。对于窗口大小为 k 的卷积核第 i 次卷积操作得到的局部特征 c_i ,此处需用到的计算公式为:

$$c_i = f(Wx_{[i, i+k-1]} + b) \quad (1)$$

其中, f 为非线性激活函数 $ReLU$; W 为卷积核矩阵中的参数; $x_{[i, i+k-1]}$ 表示词向量矩阵第 i 行到 $(i+k-1)$ 行之间的向量; b 为偏置项。窗口大小

为 k 的卷积核卷积 $(n - k + 1)$ 次, 得到的局部特征向量 $\mathbf{c}_i^k = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_{n-k+1}]$, 各个卷积核卷积输出的 \mathbf{c}_i^k 融合后得到局部特征矩阵 $\mathbf{R} = [\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_m]$ 。

2.2.2 全局语义特征提取

门控循环单元(Gate Recurrent Unit, GRU)是长短期记忆网络(Long Short-Term Memory, LSTM)的改进模型。循环神经网络模型(Recurrent Neural Network, RNN)在处理长序列数据进行学习时会出现梯度消失现象, 无法感知长时间序列的非线性关系。RNN 模型的改进版 LSTM^[10]通过输入门、遗忘门和输出门解决了梯度消失的问题。因为 LSTM 模型有参数较多, 训练时间较长的问题, 因此 Cho 等人^[11]提出了将 LSTM 的单元状态和隐藏状态合并, 并将输入门和遗忘门合为一个更新门的 GRU 模型, 简化了其结构, 因此 GRU 可以使用更少的参数和时间来达到 LSTM 的实验效果。GRU 结构如图 5 所示。

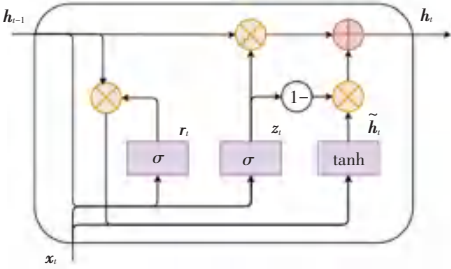


图 5 GRU 结构图

Fig. 5 GRU structure diagram

GRU 模型向前计算公式为:

$$\mathbf{r}_i = \sigma(\mathbf{W}_r \mathbf{x}_i + \mathbf{U}_r \mathbf{h}_{i-1} + \mathbf{b}_r) \quad (2)$$

$$\mathbf{z}_i = \sigma(\mathbf{W}_z \mathbf{x}_i + \mathbf{U}_z \mathbf{h}_{i-1} + \mathbf{b}_z) \quad (3)$$

$$\tilde{\mathbf{h}}_i = \tanh(\mathbf{W} \mathbf{x}_i + \mathbf{U}(\mathbf{r}_i \odot \mathbf{h}_{i-1}) + \mathbf{b}_h) \quad (4)$$

$$\mathbf{h}_i = \mathbf{z}_i \odot \mathbf{h}_{i-1} + (1 - \mathbf{z}_i) \odot \tilde{\mathbf{h}}_i \quad (5)$$

其中, $\mathbf{W}_r, \mathbf{W}_z, \mathbf{W}, \mathbf{U}_r, \mathbf{U}_z, \mathbf{U}$ 为权重矩阵; $\mathbf{b}_r, \mathbf{b}_z, \mathbf{b}_h$ 为偏置项; σ 为 sigmoid 激活函数; \odot 为 Hardamard 乘积操作; \mathbf{r}_i 为重置门的输出向量, 决定了上一时刻的隐藏状态 \mathbf{h}_{i-1} 与当前时刻输入向量 \mathbf{x}_i 的信息流入程度; \mathbf{z}_i 为更新门的输出向量, 其更新过程与 \mathbf{r}_i 类似。将 \mathbf{r}_i 与 \mathbf{h}_{i-1} 相乘, 同 \mathbf{x}_i 相加经双曲正切激活函数 \tanh 计算后可得候选隐藏状态 $\tilde{\mathbf{h}}_i$, \mathbf{h}_i 为当前时刻隐藏状态, 用于输出和流入下一个时刻的 GRU 单元进行计算。

单向 GRU 的各时刻隐藏状态是从前向后单向传输的, 这种方式忽略了文本的全局语义信息, 无法

推断出后文中的字词对前文的影响, 因此本文采用双向门控循环单元(Bidirectional Gated Recurrent Unit, BiGRU)模型进行全局语义特征提取, 可以提取正向与反向时间序列各时刻的隐藏状态, 相比于单向 GRU 其结果更加准确, BiGRU 模型结构如图 6 所示。

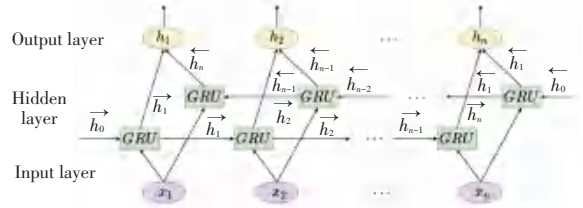


图 6 BiGRU 模型结构图

Fig. 6 BiGRU model structure diagram

图 6 中, $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ 为文本的词嵌入向量, \vec{h}_i 表示 \vec{h}_{i-1} 与 \mathbf{x}_i 经过 GRU 计算得到的正向隐藏状态, \overleftarrow{h}_i 表示 \overleftarrow{h}_{i-1} 与 \mathbf{x}_i 经过 GRU 计算得到的反向隐藏状态, \mathbf{h}_i 表示正向 \vec{h}_i 与反向 \overleftarrow{h}_{i-1} 拼接后的最终隐藏状态。

2.3 注意力层

文本信息通常会包含具有情感倾向的词汇, 这些词汇往往会对文本的情感倾向产生重要影响, 使用注意力机制可以捕获情感词予以分配较大的权重分值, 突出其在文本序列中的重要程度。本文分别对局部特征矩阵 $\mathbf{R}(\mathbf{R} = \mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_m)$ 和全局特征矩阵 $\mathbf{H}(\mathbf{H} = \mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_n)$ 分别使用注意力机制, 加强文本情感极性, 研究后推得的计算公式为:

$$\mathbf{u}_{ri} = \tanh(\mathbf{W}_r \mathbf{r}_i + \mathbf{b}_r) \quad (6)$$

$$\mathbf{u}_{hj} = \tanh(\mathbf{W}_h \mathbf{h}_j + \mathbf{b}_h) \quad (7)$$

其中, \mathbf{W}_r 与 \mathbf{W}_h 为权重参数矩阵; \mathbf{b}_r 与 \mathbf{b}_h 为偏置项; \tanh 为非线性激活函数。

通过对权重向量 \mathbf{u}_{ri} 与 \mathbf{u}_{hj} 进行归一化处理, 可得到关于局部特征 \mathbf{r}_i 和全局特征 \mathbf{h}_i 的注意力分数 α_i 与 α_j , 其计算公式为:

$$\alpha_i = \frac{\exp(\mathbf{u}_{ri})}{\sum_{i=1}^m \exp(\mathbf{u}_{ri})} \quad (8)$$

$$\alpha_j = \frac{\exp(\mathbf{u}_{hj})}{\sum_{j=1}^n \exp(\mathbf{u}_{hj})} \quad (9)$$

将注意力分数 α_i 与局部特征矩阵 \mathbf{R} 对应的子向量计算加权, 可得到经注意力机制优化后的文本局部特征向量 \mathbf{s}_r , 全局特征矩阵 \mathbf{H} 的权重优化类似, 可得到文本全局特征向量 \mathbf{s}_h 。具体的计算公式为:

$$\mathbf{s}_r = \sum_{i=1}^m \alpha_i \mathbf{r}_i \quad (10)$$

$$\mathbf{s}_h = \sum_{j=1}^n \alpha_j \mathbf{h}_j \quad (11)$$

双通道注意力机制层为文本中关键情感词向量分配了相应的注意力权重,进而提高了模型情感分析的准确性。

2.4 输出层

本文模型输出层由全连接神经网络与 *softmax* 函数构成。首先将注意力机制优化后的局部特征向量 \mathbf{s}_r 与全局特征向量 \mathbf{s}_h 融合,得到文本最终的特征表示 \mathbf{s} 。相应的数学公式可表示如下:

$$\mathbf{s} = [\mathbf{s}_r, \mathbf{s}_h] \quad (12)$$

通过在全连接神经网络层前融合 *dropout* 方法,

表1 数据集示例

Tab. 1 Examples of the data set

评论	情感标签
环境安静,离火车站近。房间布置不错,早餐以中餐为主,丰富。下次接着住!	1
房间感觉还可以,但是洗漱用的毛巾浴巾质量不好,感觉没有洗干净,房间隔音效果不好	0
酒店很好,房间挺大,服务态度特别好。地点也很好,很方便。	1
房间小,感觉卫生不太好,整个环境显得很陈旧,服务一般。	0

表1中,情感标签1和0分别表示积极评论和消极评论。实验数据集以4:1比例随机划分为训练集和测试集,每个数据集以同比例划分积极和消极评论。数据集中文本长度与数量关系如图7所示,可以观察到数据集内评论以短文本居多,其中 $\text{textlength} \leq 100$ 的数据占比约50%, $100 < \text{textlength} \leq 200$ 的数据占比约25%。

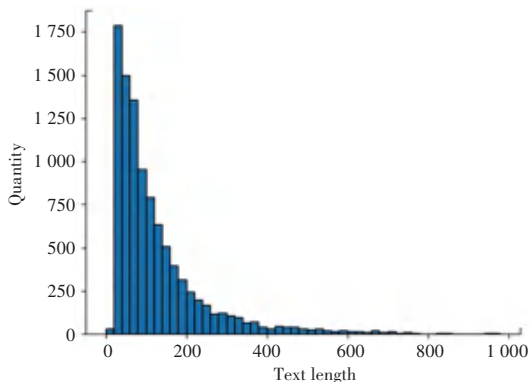


图7 文本长度与数量关系图

Fig. 7 The relationship between text length and quantity

3.2 模型评估指标

为验证模型在文本情感分析任务中的效果,本文采用准确率(*Accuracy*)、查准率(*Precision*)、召回率(*Recall*)和 F_1 -score(F_1 分值)作为评估指标,计算公式如下:

可以缓解模型的过拟合现象,全连接层的输出使用 *softmax* 函数进行分类计算,该值可由如下数学公式计算得到:

$$\mathbf{o} = f(\mathbf{W}\mathbf{s} + \mathbf{b}) \quad (13)$$

$$\mathbf{g} = \text{softmax}(\mathbf{o}) \quad (14)$$

其中, \mathbf{o} 为全连接层的输出向量; \mathbf{W} 为权重矩阵; \mathbf{b} 为偏置项; \mathbf{g} 为模型的最终输出向量。

3 实验方法及结果分析

3.1 实验数据集

本文使用的是谭松波老师收集整理的酒店评论数据集,该数据集包含积极和消极评论各7000条和3000条。使用前对数据集进行预处理,去除掉特殊字符,调整数据集结构,其数据集示例见表1。

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (15)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (16)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (17)$$

$$F_1 = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (18)$$

其中, TP (True Positive) 表示样本正确分类为积极情感的数量; FP (False Positive) 表示样本错误分类为积极情感的数量; FN (False Negative) 表示将样本错误分类为消极情感的数量; TN (True Negative) 表示将样本正确分类为消极情感的数量。

3.3 实验环境与超参数设定

本文模型实验环境见表2。

表2 实验环境

Fig. 2 Laboratory environment

实验环境	环境配置
操作系统	Windows 10
显卡	Nvidia GTX 1650
编程语言	Python 3.7.9
深度学习框架	Pytorch 1.1.0
CUDA 版本	11.0

本文词嵌入层使用 Google 发布的“BERT-Base Chinese”中文预训练模型,该模型采用 12 层 Transformer,隐藏层的维度为 768,激活函数为 *Relu*,多头注意力的参数为 12,模型总参数大小为 110 MB。双通道特征提取层的主要参数见表 3。

表 3 特征提取层模型参数设定

Tab. 3 Feature extraction layer model parameters setting

参数	参数值
batch size	6
词向量维度	768
hidden size	256
dropout rate	0.2
epochs	10
学习率	0.001
优化器	Adam
卷积核尺寸	(2, 3, 4)
卷积核通道数	(256, 256, 256)
最大文本长度	150

3.4 对比实验与结果分析

3.4.1 BERT 模型词向量化效果分析

为验证 BERT 预训练模型对文本向量化的表示能力,本文选取 Word2Vec、ELMo 与 BERT 进行实验对比,文本数据分别通过 3 个模型向量化后输入到 BiGRU 中进行训练。实验结果见表 4。结果表明基于 BiGRU 的 BERT 词嵌入模型的准确率高达 0.911 0,相比 Word2Vec 与 ELMo 准确率分别提高了 4.5%与 2.6%,证明了基于双向 Transformer 编码器的 BERT 模型有更出色的词向量表示能力。

表 4 词嵌入层模型试验结果对比

Tab. 4 Comparison of test results of word embedding layer model

词嵌入模型	Accuracy	Precision	Recall	$F_1 - socre$
Word2Vec	0.856 5	0.825 1	0.830 6	0.827 8
LoEM	0.874 0	0.861 9	0.846 4	0.854 1
BERT	0.895 5	0.861 1	0.885 2	0.871 7

3.4.2 本文模型实验对比与结果分析

为验证本文模型的有效性,使用 BERT 模型作为词嵌入层进行文本的向量化表示,同时为验证双通道特征提取模型与注意力机制结合的有效性,分别构建 CNN-BiGRU 与 CNN-BiGRU-AT 模型进行实验对比。在文本情感分析任务中, BiGRU、BiLSTM、TextRNN、TextCNN 是情感分类中的经典模型,本文将选取上述模型进行对比实验,其模型结构如下。

(1) BiGRU:该模型为本文的全局特征提取层,

对比实验中使用初始时刻与最后时刻的融合特征向量作为模型输出。

(2) BiLSTM:该模型采用双向长短期记忆网络进行特征提取,使用初始时刻与最后时刻的融合特征向量作为模型输出。

(3) TextRNN:该模型采用双向长短期记忆网络进行特征提取,与 BiLSTM 不同的是,使用最大池化对 BiLSTM 各时刻的输出特征向量进行特征提取作为模型输出。

(4) TextCNN:该模型采用多尺度卷积对输入向量进行特征提取,不同尺度的卷积核获取到不同长度词语的特征信息,最后使用最大池化对特征向量进行计算作为模型输出。

(5) CNN-BiGRU:该模型为未使用注意力机制的双通道特征提取模型,主要与本文 BERT-CNN-BiGRU-AT 模型进行对比,验证注意力机制的有效性。

所有模型对比实验结果见如表 5。

表 5 不同模型的实验对比结果

Tab. 5 Experimental comparison results of different models

模型	Accuracy	Precision	Recall	$F_1 - socre$
BERT-BiGRU	0.895 5	0.861 1	0.885 2	0.871 7
BERT-BiLSTM	0.898 5	0.860 8	0.892 4	0.874 3
BERT-TextRNN	0.902 5	0.884 2	0.883 8	0.884 0
BERT-TextCNN	0.908 0	0.890 5	0.890 5	0.890 5
BERT-CNN-BiGRU	0.916 5	0.897 5	0.902 6	0.900 0
BERT-CNN-BiGRU-AT	0.929 0	0.921 2	0.912 1	0.916 4

实验结果评估指标以 *Accuray* 与 $F_1 - socre$ 为主,根据表 5 可知,本文模型 BERT-CNN-BiGRU-AT 的准确率达到 92.90%, F_1 分数达到了 91.64%,均高于其他模型。其中, BERT-BiGRU 与 BERT-BiLSTM 的模型评分相差不多,因为二者具有相同的模型结构。TextCNN 模型在本文数据集集中的实验效果均好于循环神经网络模型,在酒店评论数据集集中,短文本数据占比约为 75%, TextCNN 通过多尺度卷积可以更好地获取到短文本序列词组之间的特征信息,而循环神经网络擅长捕获较长文本的时间序列特征。通过 BERT-CNN-BiGRU 与 BERT-CNN-BiGRU-AT 对比可知,使用了注意力机制后,准确率提高了约 1.4%, F_1 分数提高了约 1.8%,说明注意力机制可以有效提高模型的准确率。综上所述,本文所提出的基于 BERT 的 CNN-BiGRU-AT 模型在中文文本情感分析任务上具有很好的效果,证明了本文模型的有效性。

4 结束语

本文提出基于 BERT 的双通道神经网络模型的文本情感分类方法,首先采用 BERT 模型进行文本向量化表示,结合 CNN 和 BiGRU 构建双通道模型提取文本中蕴含的局部与全局情感特征,通过建立注意力权重分配机制,融合加权求和后的特征向量,经过输出层计算后,得出文本的情感倾向信息,在酒店评论数据集中表现出了较高的准确率。但是依然有不足之处,表现为模型训练时间过久,其原因是 BERT 模型具有大量参数,且模型深度较深,在对下游任务进行微调的过程中,需要反向传播更新大量参数,消耗较多时间。针对此问题拟在后续研究中加以改进。

参考文献

- [1] KAUR S, SIKKA G, AWASTHI L K. Sentiment analysis approach based on N - gram and KNN classifier [C]//2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC). Jalandhar, India;IEEE, 2018; 1-4.
- [2] KIM Y. Convolutional neural networks for sentence classification [C]//Proceedings of 2014 Conference on Empirical Methods in Natural Language Processing. Doha, Qatar; ACL, 2014; 1746 - 1751.
- [3] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint arXiv: 1409.0473, 2014.
- [4] BASIRI M E, NEMATI S, ABDAR M, et al. ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis[J]. Future Generation Computer Systems, 2021, 115: 279-294.
- [5] LIU Gang, GUO Jiabao. Bidirectional LSTM with attention mechanism and convolutional layer for text classification [J]. Neurocomputing, 2019, 337: 325-338.
- [6] YANG Chao, ZHANG Hefeng, JIANG Bin, et al. Aspect-based sentiment analysis with alternating coattention networks [J]. Information Processing & Management, 2019, 56(3): 463-478.
- [7] RHANOU M, MIKRAM M, YOUSFI S, et al. A CNN-BiLSTM model for document - level sentiment analysis [J]. Machine Learning and Knowledge Extraction, 2019, 1(3): 832-847.
- [8] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding [J]. arXiv preprint arXiv:1810.04805, 2018.
- [9] 谌志群,鞠婷. 基于 BERT 和双向 LSTM 的微博评论倾向性分析研究[J]. 情报理论与实践, 2020, 43(08): 173-177.
- [10] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1780.
- [11] CHO K, Van MERRIËNBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation [J]. arXiv preprint arXiv: 1406. 1078, 2014.
- [17] AHMED O B, BENOIS-PINEAU J, ALLARD M, et al. Classification of Alzheimer's disease subjects from MRI using hippocampal visual features[J]. Multimedia Tools and Applications, 2015, 74(4): 1249-1266.
- [18] DAI Dai, HE Huiguang, VOGELSTEIN J T, et al. Accurate prediction of AD patients using cortical thickness networks[J]. Machine Vision and Applications, 2013, 24(7): 1445-1457.
- [19] CHEN G, WARD B D, XIE C, et al. Classification of Alzheimer disease, mild cognitive impairment, and normal cognitive status with large - scale network analysis based on resting - state functional MR imaging [J]. Radiology, 2011, 259(1): 213-221.
- [20] LIU Sheng, YADAV C, FERNANDEZ-GRANDA C, et al. On the design of convolutional neural networks for automatic detection of Alzheimer's disease [C]//Machine Learning for Health Workshop. PMLR, 2020; 184-201.
- [21] HUO Yuankai, XU Zhoubing, XIONG Yunxi, et al. 3D whole brain segmentation using spatially localized atlas network tiles[J]. NeuroImage, 2019, 194: 105-119.

(上接第 15 页)