

ELMo-CNN-BiGRU 双通道文本情感分类模型

吴迪,王梓宇,赵伟超

(河北工程大学 信息与电气工程学院,河北 邯郸 056038)

摘要: 文本情感分类通过对带有情感色彩的主观性文本进行分析和推理,帮助用户更好地做出判断与决策。针对传统情感分类模型难以根据上下文信息调整词向量的问题,提出一种双通道文本情感分类模型。利用ELMo和Glove预训练模型分别生成动态和静态词向量,通过堆叠嵌入2种词向量生成输入向量。采用自注意力机制处理输入向量,计算内部的词依赖关系。构建融合卷积神经网络(CNN)和双向门控递归单元(BiGRU)的双通道神经网络结构,同时获取文本局部特征和全局特征。最终将双通道处理结果进行拼接,经过全连接层处理后输入分类器获得文本情感分类结果。实验结果表明,与同类情感分类模型中性能较优的H-BiGRU模型相比,ELMo-CNN-BiGRU模型在IMDB、yelp和sentiment140数据集上的准确率和F1值分别提升了2.42、1.98、2.52和2.40、1.94、2.43个百分点,具有更好的短文本情感分类效果和稳定性。

关键词: 文本情感分类;双通道;预训练模型;深度学习;自注意力机制

开放科学(资源服务)标志码(OSID):



中文引用格式:吴迪,王梓宇,赵伟超.ELMo-CNN-BiGRU 双通道文本情感分类模型[J].计算机工程,2022,48(8):105-112.

英文引用格式:WU D, WANG Z Y, ZHAO W C.ELMo-CNN-BiGRU dual-channel text sentiment classification model[J]. Computer Engineering, 2022, 48(8): 105-112.

ELMo-CNN-BiGRU Dual-Channel Text Sentiment Classification Model

WU Di, WANG Ziyu, ZHAO Weichao

(College of Information and Electrical Engineering, Hebei University of Engineering, Handan, Hebei 056038, China)

[Abstract] Text sentiment classification helps users make better decisions by analyzing and reasoning subjective texts with emotional colors. Addressing the difficulty in adjusting the word vector according to the context information in traditional sentiment classification models, a dual-channel text sentiment classification method is proposed. To begin, pretrained ELMo and Glove models are used to generate dynamic and static word vectors, respectively. The input vector is generated by stacking and embedding two-word vectors. Second, the self-attention mechanism is used to process the input vector and calculate the internal word dependencies. The dual-channel neural network structure is constructed by a Convolutional Neural Network (CNN) and Bi-directional Gated Recurrent Unit (BiGRU). The local and global features of the text can be obtained simultaneously. Finally, the dual-channel processing results are spliced, processed by the fully connected layer, and sent to the classifier. The classification results can be obtained. The results show that, compared with the H-BiGRU model with the best performance among contrastive sentiment classification models, the accuracy of the proposed ELMo-CNN-BiGRU model on the IMDB, yelp, and sentiment140 datasets improved by 2.42, 1.98, 2.52, respectively, and the F1 value improved by 2.40, 1.94, 2.43 percentage points, respectively. It achieves a better sentiment classification effect and stability for short texts.

[Key words] text sentiment classification; dual channel; pretrained model; deep learning; self-attention mechanism

DOI: 10.19678/j.issn.1000-3428.0062047

0 概述

文本情感分类是自然语言处理领域的重要分支,广泛应用于舆情分析、内容推荐等任务,能帮助用户快

速获取、整理和分析相关信息,并对带有情感色彩的主观性文本进行分析、处理、概括和推理。目前,深度学习技术^[1]和注意力机制得到快速发展并且在文本分类领域取得了一定的研究进展。在利用深度学习技术实

基金项目: 国家重点研发计划“科技冬奥”重点专项子课题“冬奥会公共安全综合风险评估技术”(2018YFF0301004-02);河北省自然科学基金“面向微博短文本的主题模型聚类方法研究”(F2020402003);河北省自然科学基金“面向大数据应用的云计算中心性能分析与预测方法研究”(F2019402428);河北省高等学校科学技术研究重点项目“增量序列模式匹配下网络入侵检测方法研究”(ZD2018087)。

作者简介: 吴迪(1984—),女,副教授、博士,主研方向为数据挖掘、文本聚类、自然语言处理;王梓宇、赵伟超,硕士研究生。

收稿日期: 2021-07-12 **修回日期:** 2021-09-06 **E-mail:** wudiwudi@hebeu.edu.cn

现自然语言处理的过程中,常用 word2vec、Glove 等词嵌入方式表示原始文本信息,从而有效捕捉句法和语义相关信息。文献[2]提出一种基于卷积神经网络(Convolutional Neural Network, CNN)的情感分类模型 CNN_Text_word2vec,引入 word2vec 来训练每个单词上的分布式单词嵌入,通过多个不同大小的卷积核来学习文本特征。文献[3]提出卷积神经网络和长短期记忆(Long Short-Term Memory, LSTM)网络模型 CNN-LSTM,采用 word2vec 训练初始单词向量,利用 CNN 提取文本局部特征,通过 LSTM 捕获序列之间的长期依赖关系。文献[4]提出一种基于深度卷积神经网络的文本情感分类模型,堆叠多个卷积层,并利用全局最大池化构建情感分类模块。文献[5]提出 MTL-MSCNN-LSTM 模型,将多尺度 CNN 和 LSTM 相结合,有效利用处理不同尺度文本的局部和全局特征。文献[6]提出 MF-CNN 模型,结合多样化的特征信息,利用句中的情感特征优化情感分类结果。

针对现有深度学习技术在文本情感分类任务中特征提取能力不足的问题,研究人员提出了一系列解决方案。文献[7]提出一种双通道卷积神经网络模型,将扩展文本特征和语义特征分别输入到多通道的 CNN 模型中,增强模型的情感特征提取能力。文献[8]提出一种基于深度学习的词汇集成双通道 CNN-LSTM 情感分析模型,将 CNN 和双向长短期记忆(Bi-directional Long Short-Term Memory, BiLSTM)分支以并行方式组合在一起提取特征。文献[9]提出一种卷积神经网络和长短期记忆结合的模型,串联 CNN 和 LSTM 两个模型提取文本情感特征。文献[10]提出 Multi-Bi-LSTM 模型,对现有的信息生成不同的特征通道,利用 BiLSTM 学习句子中的情感倾向信息。

注意力机制本质上与人类的选择性机制类似,其目标是从众多信息中选择出对当前任务目标更关键的信息,被广泛应用于自然语言处理、图像与语音识别等各种不同类型的深度学习任务中,可有效提高模型分类效果。文献[11]提出一种多通道卷积神经网络模型,采用3种通道进行卷积操作,利用字向量发现文本深层语义特征。文献[12]提出一种基于 RNN 和 CNN 并融入注意力机制的新模型。文

献[13]利用区域 LSTM 来调整不同信息对分类结果的影响程度,并实现了区域划分。文献[14]提出一种基于递归神经网络(Recurrent Neural Network, RNN)的情感分析方法,将预处理后的文本数据输入到一个或者多个 RNN 中,利用注意力机制进行处理。文献[15]对不同注意力模块进行独立训练,用以提取各类文本情感特征。文献[16]提出基于注意力机制的 AT-DPCNN 模型,利用 CNN 进一步提取注意力输入矩阵的特征。文献[17]提出一种基于注意力的双向 CNN-RNN 模型,利用注意力机制给双向特征提取层赋予权重。文献[18]在利用多头注意力机制的同时加入位置编码,采用图卷积网络获取文本的情感信息。

上述模型由于采用传统的静态词向量进行建模,导致上下文相似但情感极性不同的词语映射到向量空间相邻位置,影响情感分类效果,因此本文充分考虑不同上下文中的一词多义现象,并且兼顾提取全局特征与局部特征,提出一种 ELMO^[19]-CNN-双向门控递归单元(Bi-directional Gated Recurrent Unit, BiGRU)双通道文本情感分类模型。将动态词向量与静态词向量相结合,以实现不同上下文语境中的一词多义效果。采用注意力机制处理输入向量,得到内部词之间的依赖关系。构建 CNN-BiGRU 双通道结构,并行提取数据的局部和全局特征,以实现特征的完整提取。

1 相关知识

1.1 ELMo 预训练模型

文献[19]提出的 ELMo 是一个双层双向的 LSTM 模型,分别有一个前向和一个后向语言模型,前向 LSTM 和后向 LSTM 通过 $\prod_{i=1}^n p(c_i | c_1, c_2, \dots, c_{i-1})$ 和 $\prod_{i=1}^n p(c_i | c_{i+1}, c_{i+2}, \dots, c_n)$ 计算获得,其中 (c_1, c_2, \dots, c_n) 表示 n 个文本序列。ELMO 预训练模型结构如图 1 所示,其中 E 表示文本语料的词向量。利用大量语料训练 ELMo 模型结构,获得预训练模型。将 ELMo 预训练模型生成数据的动态词向量与 Glove 预训练模型生成的静态词向量进行堆叠嵌入,作为 ELMo-CNN-BiGRU 模型的输入特征。

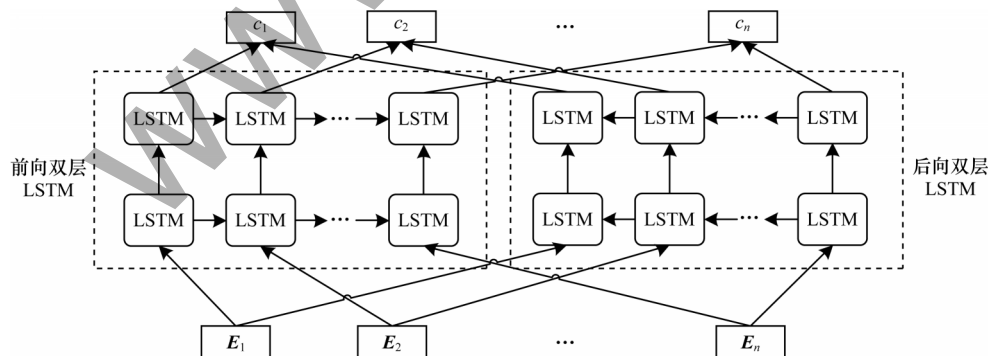


图1 ELMo 预训练模型结构

Fig.1 ELMo pretrained model structure

1.2 注意力机制

注意力机制是一种权重分配机制,通过调整权重系数来修改文本特征的关键程度。权重系数越

大,表示该信息越重要,对文本情感分类结果影响越大。本文采用自注意力(Self-Attention)机制,仅关注自身信息更新训练参数,降低对外部信息的依赖

程度,更好地获取信息的内部关联关系。

1.3 双通道深度学习模型

构建CNN和BiGRU双通道模型,同时获取输入信息的局部特征和全局特征。双通道结构可以有效弥补CNN和BiGRU模型各自缺陷,充分捕捉局部和全局的情感特征,优化文本情感分类结果。

1)局部特征获取。采用CNN获取输入信息的局部特征。通过卷积层抽取信息特征,获取评论文本信息的局部语义特征。利用多尺度卷积核,采用不同尺度的卷积过滤器来帮助模型进行特征学习,获得不同距离单词间的特征信息。经过卷积层提取文本的特征信息后,输入最大池化层中进行特征降维。

2)全局特征获取。采用BiGRU获取输入信息的全局特征,并且充分考虑上下文信息,更好地捕捉文本中的语义信息。BiGRU中的GRU单元结构如图2所示,GRU单元在 t 时刻的更新门 r_t 、重置门 z_t 、当前记忆内容 h_c 、当前时间的最终记忆 h_t 更新过程如下:

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

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

$$h_c = \tanh(W_h x_t + U_h (r_t \odot h_{t-1}) + b_h) \quad (3)$$

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

其中: x_t 为输入向量; σ 、 \tanh 为激活函数; h_{t-1} 是 $t-1$ 时刻隐藏层状态信息; W_r 、 W_z 、 W_h 、 U_r 、 U_z 、 U_h 为权重矩阵; b_r 、 b_z 、 b_h 为偏置值; \odot 符号是向量间点乘运算。

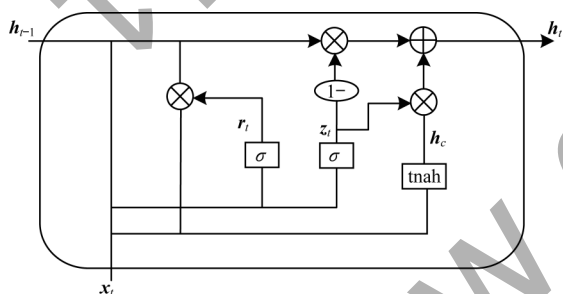


图2 GRU单元结构

Fig.2 GRU unit structure

2 ELMo-CNN-BiGRU模型

本文提出一种ELMo-CNN-BiGRU文本情感分类模型。首先,对原始评论文本数据进行去停用词、无用标签、特殊符号等预处理操作。然后,在词嵌入操作过程中,采用ELMo和Glove模型分别生成动态词向量和静态词向量,将两种词向量进行堆叠嵌入作为模型的输入向量。其次,利用自注意力机制处理输入向量,获取信息的内部关联关系。再次,利用双通道的CNN和BiGRU模型分别提取局部和全局特征。最后,将双通道提取的特征进行拼接,经全连接层处理后,输入分类器进行文本情感分类操作。ELMo-CNN-BiGRU模型框架如图3所示。

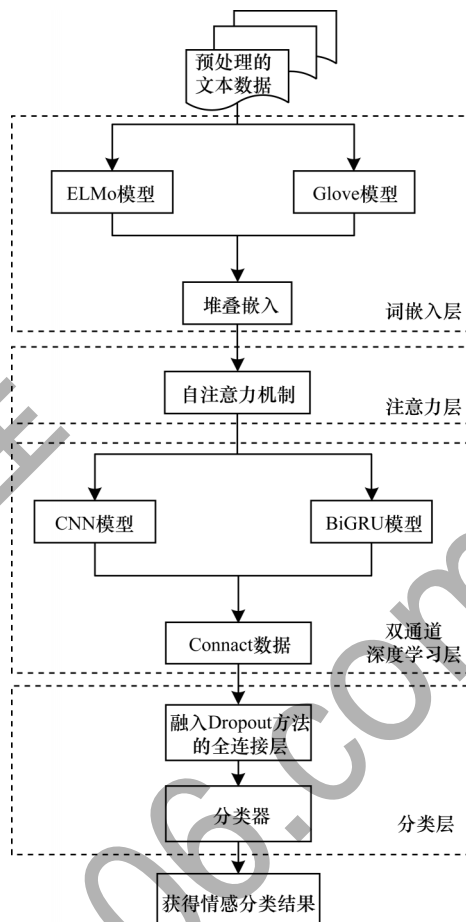


图3 ELMo-CNN-BiGRU模型框架

Fig.3 ELMo-CNN-BiGRU model frame

2.1 词嵌入层

由于静态词向量生成后不会发生改变,因此在应对不同上下文中一词多义情况时存在局限性。动态词向量会随着模型训练在不同上下文中做出调整,其初始化会影响到文本情感分类的收敛速度和最终结果。单独使用静态词向量和动态词向量其中的一种均存在一定的问题。为了在更精准地表示文本词向量的同时加快模型收敛速度,并优化文本情感分类结果,采用Glove模型生成的静态词向量与ELMo预训练模型生成的动态词向量,以堆叠嵌入的方式作为模型的输入向量。利用ELMo预训练模型进行词嵌入操作,应对不同上下文中一词多义的情况,使用堆叠嵌入的方式融合两种词向量,能够有效优化文本情感分类结果。

Glove模型生成词的向量为 w_g ,具体公式如下:

$$w_g = [x_1, x_2, \dots, x_t] \quad (5)$$

其中: t 表示词向量 w_g 的维度。

ELMo模型生成的词向量为 w_e ,具体公式如下:

$$w_e = [x_1, x_2, \dots, x_T] \quad (6)$$

其中: T 表示词向量 w_e 的维度。

将 w_g 与 w_e 堆叠生成的词向量,作为模型隐藏层的输入数据 i_{input} ,具体公式如下:

$$i_{input} = [w_g, w_e] \quad (7)$$

2.2 注意力层

采用Self-Attention机制得到内部的词依赖关

系,具体公式如下:

$$S_{\text{Self-Attention}} = \text{softmax} \left(\frac{W \cdot W^T}{\sqrt{d_k}} \right) W \quad (8)$$

其中: d_k 为输入词嵌入向量的维度; W 为输入信息。

将堆叠嵌入的特征向量,输入到由 Self-Attention 机制构建的注意力层,进一步在双通道深度学习层中获取局部与全局特征。

2.3 双通道深度学习层

ELMo-CNN-BiGRU 模型采用双通道结构,其中 CNN 通道用来捕获评论文本数据的局部特征, BiGRU 通道用来获取评论文本数据的全局特征。该结构可以同步综合考量评论文本数据的局部和全局特征,在提取文本数据全局情感倾向特征的同时,捕获文本的局部情感特征,从而更全面地获取文本的情感特征,有效优化文本情感分类结果。

2.3.1 CNN 通道

ELMo-CNN-BiGRU 模型利用 CNN 通道获取文本数据的局部特征,其中 CNN 通道由卷积层和池化层两部分组成。

1) 卷积层

卷积层采用 3 种不同大小的卷积核,获得不同距离词序列之间的特征来提取更全面的局部信息。在 CNN 通道中,利用卷积核进行一维卷积操作,将词嵌入矩阵转化为一维向量,采用 ReLU 函数作为卷积层的激活函数。卷积操作能提取词语序列的特征,从词语序列 $X_{i:i+h-1}$ 中提取特征图,具体公式如下:

$$Z_i = f(W \cdot X_{i:i+h-1} + b) \quad (9)$$

其中: $b \in \mathbb{R}^k$ 为偏置项, k 为向量的维度; f 为非线性函数 ReLU; $W \in \mathbb{R}^{hd}$ 表示卷积核, $h(h < n)$ 为高度, d 为宽度。

2) 池化层

采用最大池化法提取特征图 (Z_{\max}) 中池化区域最大的特征值,可对特征信息进行降维,具体公式如下:

$$Z_{\max} = \max(Z_i) \quad (10)$$

2.3.2 BiGRU 通道

BiGRU 通道用于提取数据的全局特征。在 BiGRU 模型中, t 时刻当前隐藏层 \vec{h}_t 由前向隐藏层 \vec{h}_t 和后向隐藏层 \vec{h}_t 加权求和得到,具体计算如下:

$$\vec{h}_t = \text{GRU}(\vec{x}_t, \vec{h}_{t-1}) \quad (11)$$

$$\vec{h}_t = \text{GRU}(\vec{x}_t, \vec{h}_{t-1}) \quad (12)$$

$$\vec{h}_t = w_t \vec{h}_t + v_t \vec{h}_t + b_t = [\vec{h}_t, \vec{h}_t] \quad (13)$$

其中: \vec{x}_t 表示当前隐藏层的输入向量; \vec{h}_{t-1} 表示 $(t-1)$ 时刻前向隐藏层状态; \vec{h}_{t-1} 表示 $t-1$ 时刻后向隐藏层状态; w_t, v_t 分别表示 t 时刻 BiGRU 所对应的前向隐藏层和后向隐藏层的相关权重值; b_t 表示 t 时刻隐藏层状态的偏置值。

2.4 分类层

将双通道的特征数据进行拼接后,输入到全连接层中,通过在全连接层前融合 Dropout 方法缓解模型过拟合现象。利用 softmax 函数对文本进行情感

分类,具体公式如下:

$$Y = \text{Dropout}(a \cdot y + b) \quad (14)$$

$$\hat{Y} = \text{softmax}(Y) \quad (15)$$

其中: a, b 分别为全连接层的权重矩阵和偏置值; y 为双通道拼接后的向量数据; Y 为利用 Dropout 方法处理后的向量。

2.5 文本情感分类流程

ELMo-CNN-BiGRU 文本情感分类流程具体如下:

1) 将文本 D 同时输入 ELMo 预训练模型和 Glove 预训练模型,生成对应的词向量 $S_c = [w_1, w_2, \dots, w_x], S_g = [w_1, w_2, \dots, w_x]$,其中文本 D 由 x 个句子 $\{d_1, d_2, \dots, d_x\}$ 组成,文本中第 i 个句子由 j 个单词 $\{c_1, c_2, \dots, c_j\}$ 组成, x 和 X 为 Glove 和 ELMo 所生成的词向量的维度。

2) 将生成的词向量进行堆叠式嵌入,得到模型的输入向量 $i_{\text{input}} = [S_c, S_g]$ 。

3) 将词嵌入层生成的向量输入自注意力层进行处理。

4) 将自注意力机制层中的数据分别输入双通道的 CNN 层和 BiGRU 层,分别得到两个通道的隐藏层表示。

5) 拼接双通道的特征向量并输入带有 Dropout 机制的全连接层,经过全连接层处理后使用 softmax 函数进行分类,获得情感分类的最终结果。

3 实验结果与分析

3.1 数据集与实验环境设置

在 Colaboratory 云平台上进行模型搭建,采用 Python 编程语言和 keras 框架进行实验。在 IMDB、yelp、sentiment140 数据集上,将 ELMo-CNN-BiGRU 模型与 6 种模型进行性能对比,此外,针对 ELMo-CNN-BiGRU 模型,将在迭代次数、词向量维度、词向量嵌入方式和通道结构这 4 个方面进行对比实验,以验证 ELMo-CNN-BiGRU 双通道文本情感分类模型的有效性。实验所用数据集中训练集和测试集的数据量比例为 8:2,具体统计信息如表 1 所示。

表 1 数据集统计信息

Table 1 Data set statistics

数据集	训练集数据量	测试集数据量
IMDB	8 000	2 000
yelp	4 000	1 000
sentiment140	8 000	2 000

3.2 超参数设置

将句子最大长度设置为 100。对于小于最大长度的句子,采取补零操作;对于大于最大长度的句子,采取截断操作。采用 Adam 函数作为模型的训练优化函数,Adam 函数的学习率设置为 0.000 1。在 ELMo-CNN-BiGRU 模型的 CNN 通道中,多通道 CNN 选取的卷积核窗口大小分别设置为 2、3、5,采用 Dropout 机制应对模型可能会发生的过拟合现象。

BiGRU模块采用Tanh作为神经元的激活函数,CNN模块采用ReLU函数作为卷积操作的激活函数。超参数设置如表2所示。

Table 2 Hyperparameter setting	
参数名	参数值
学习率	0.000 1
批处理大小	128
迭代次数	30
词向量维度	612
Dropout	0.5

3.3 评价指标

采用准确率(A)和F1值(F)作为评价指标,用来测试ELMo-CNN-BiGRU及其对比模型性能。准确率和F1值的计算公式分别如下:

$$A = \frac{T_{TP} + T_{TN}}{T_{TP} + T_{TN} + F_{FP} + F_{FN}} \tag{16}$$

$$F = \frac{2 \times P \times R}{P + R} \tag{17}$$

$$P = \frac{T_{TP}}{T_{TP} + F_{FP}} \tag{18}$$

$$R = \frac{T_{TP}}{T_{TP} + F_{FN}} \tag{19}$$

其中:P表示精确率;R表示召回率; T_{TP} 表示实际标签是真,模型预测结果为真的文本数量; F_{FN} 表示实际标签为真,但是模型预测结果为假的文本数量; F_{FP} 表示实际标签为假,但是模型预测结果为真的文本数量; T_{TN} 表示实际标签为假,模型预测结果也为假的文本数量。

3.4 结果分析

为验证ELMo-CNN-BiGRU模型性能,首先测试迭代次数对实验结果的影响,其次测试不同词向量维度、词向量嵌入方式、通道结构下的模型准确率和F1值,最后将ELMo-CNN-BiGRU与CNN、BiGRU等6种文本情感分类模型进行对比实验。

3.4.1 迭代次数对比

在IMDB、yelp、sentiment140这3个数据集上,ELMo-CNN-BiGRU模型的准确率和F1值随着迭代次数的变化趋势如图4、图5所示。

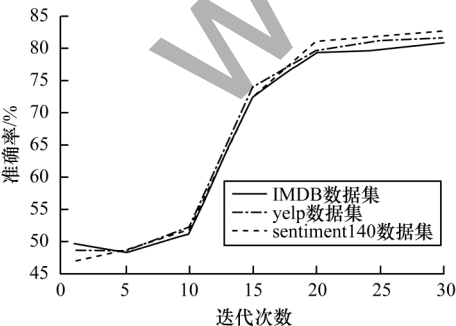


图4 ELMo-CNN-BiGRU模型在3个数据集上的准确率对比
Fig.4 Comparison of the accuracy of the ELMo-CNN-BiGRU model on the three data sets

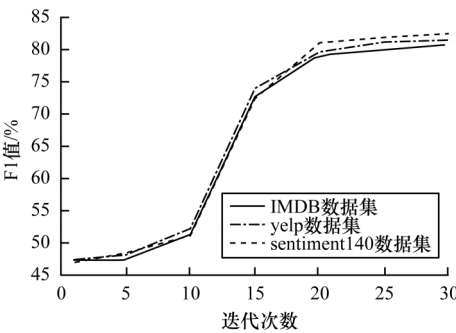


图5 ELMo-CNN-BiGRU模型在3个数据集上的F1值对比
Fig.5 Comparison of F1 value of ELMo-CNN-BiGRU model on three data sets

由图4、图5可知:在第10次迭代前,ELMo-CNN-BiGRU模型的准确率和F1值总体呈现缓慢上升趋势,这是由于模型刚开始训练,权重参数在初始化阶段,需要增加迭代次数才能训练出提升模型分类效果的参数;在第10次迭代后,ELMo-CNN-BiGRU模型的准确率和F1值在IMDB、yelp、sentiment140这3个数据集上普遍呈现较快上升趋势,这是由于经过迭代训练,模型已经训练出部分能提升模型分类效果的有效参数,进而加快了迭代速度;在第20次迭代后,ELMo-CNN-BiGRU模型在准确率和F1值上普遍处于平稳状态,这是由于模型要达到拟合状态,结果变化不明显;在IMDB、yelp、sentiment140这3个数据集上,ELMo-CNN-BiGRU模型的准确率和F1值的变化趋势相近,这表明ELMo-CNN-BiGRU模型面对有差异数据时,分类结果具有一定的稳定性。

综上所述,在3个数据集上,ELMo-CNN-BiGRU模型在sentiment140数据集上取得了最优结果,这是由于该模型使用的ELMo模型更适合处理短文本数据,在较短的文本上能获得更好的分类效果。

3.4.2 词向量维度对比

词向量维度对模型性能具有一定的影响。词向量维度越大,所表示的信息量越大,但会淡化词语之间的联系,然而词向量维度过小,则会产生无法有效区分词语的问题,因此需要选择合适的词向量维度。在3个数据集上,对ELMo-CNN-BiGRU模型使用不同维度的词向量进行实验测试。本文使用ELMo模型和Glove模型堆叠的方式生成词向量,其中ELMo分别生成256、512、1024维的向量,Glove分别生成50、100、200维的向量,将两者堆叠后产生不同维度的词向量。在3个数据集上,当词向量维度不同时,ELMo-CNN-BiGRU模型的准确率与F1值对比结果如图6、图7所示。

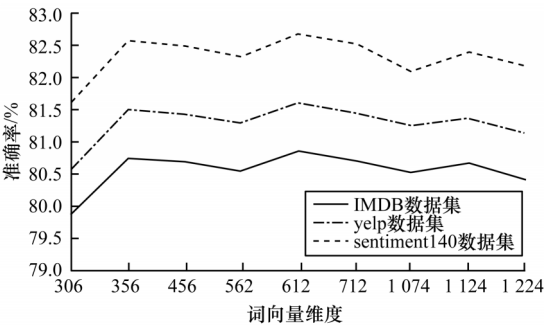


图6 词向量维度对ELMo-CNN-BiGRU模型准确率的影响
Fig.6 Effect of word vector dimension on the accuracy of ELMo-CNN-BiGRU model

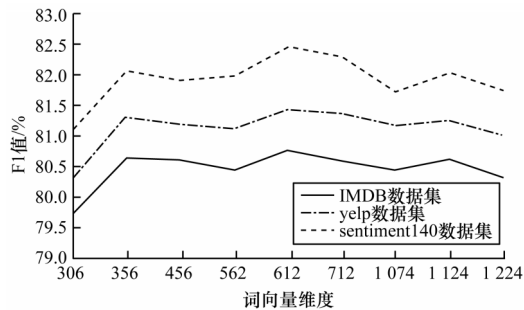


图7 词向量维度对ELMo-CNN-BiGRU模型F1值的影响

Fig.7 Effect of word vector dimension on F1 value of ELMo-CNN-BiGRU model

由图6、图7可知:ELMo-CNN-BiGRU模型在3个数据集上的准确率和F1值都在词向量维度为612时达到最优效果。此时,ELMo模型词向量维度为512、Glove模型词向量维度为100,因此词向量维度并非越大效果越好,过大反而会导致模型的性能下降。在词向量维度到达612之前,模型的性能震荡上升,在612维度后模型的性能震荡下降。可见,ELMo-CNN-BiGRU模型性能变化不稳定,震荡改变是由于本文的词向量采用堆叠嵌入,动态词向量与静态词向量对模型性能的影响程度不同。

3.4.3 词向量嵌入方式对比

采用ELMo和Glove模型分别生成动态词向量与静态词向量,堆叠后作为模型的输入向量,为验证堆叠嵌入方式的优势,将其与word2vec、Glove、ELMo、ELMo+word2vec进行对比,在采用不同的词向量嵌入方式时,预训练模型的准确率与F1值结果如图8、图9所示。由图8、图9可知:ELMo和Glove堆叠嵌入方式明显优于其他4种词嵌入方式;与传统静态词向量嵌入方式相比,动态词向量与静态词向量堆叠嵌入方式能根据不同的上下文调整词向量,使得模型在准确率和F1值上取得更好的效果;与ELMo和word2vec堆叠嵌入方式相比,Glove充分考虑全局信息,在实际应用中效果更好。

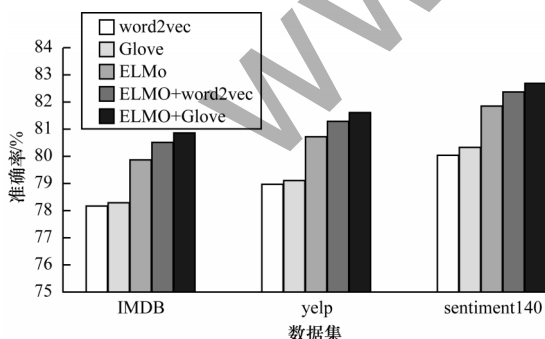


图8 不同词向量嵌入方式对ELMo-CNN-BiGRU模型准确率的影响

Fig.8 Effect of different word vector embedding modes on the accuracy of ELMo-CNN-BiGRU model

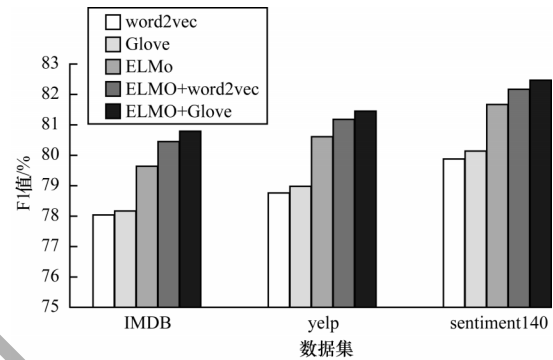


图9 不同词向量嵌入方式对ELMo-CNN-BiGRU模型F1值的影响

Fig.9 Effect of different word vector embedding modes on the F1 value of ELMo-CNN-BiGRU model

3.4.4 通道结构对比

ELMo-CNN-BiGRU模型采用双通道结构,将数据送入双通道结构,能更好地捕捉文本数据的全局与局部特征。在3个数据集上分别进行双通道结构与CNN-BiGRU串行结构的准确率和F1值对比,实验结果如图10、图11所示。

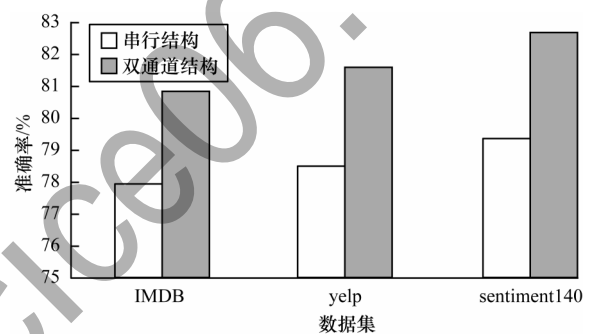


图10 双通道结构对模型准确率的影响

Fig.10 Effect of dual-channel structure on model accuracy

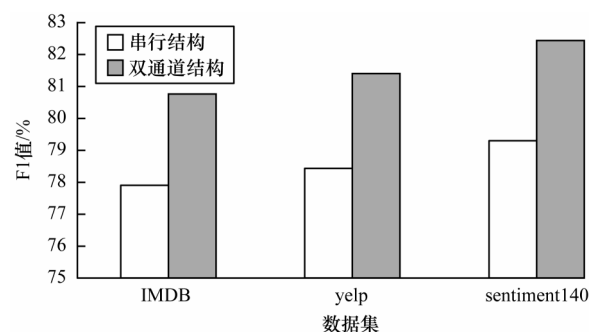


图11 双通道结构对模型F1值的影响

Fig.11 Effect of dual-channel structure on model F1 value

由图10、图11可知,双通道结构较串行结构可有效提升模型性能,在IMDB、yelp、sentiment140这3个数据集上,准确率分别提升了2.91、3.08、3.28个百分点,F1值分别提升了2.87、2.98、3.13个百分点。

3.4.5 文本情感分类模型对比

将ELMo-CNN-BiGRU模型与如下6种模型进

行对比实验:

- 1)CNN,采用卷积层提取特征,池化层降低特征维度。
- 2)BiGRU,在门控递归单元的基础上,采用双向结构进行情感分析。
- 3)BiLSTM+EMB_ATT^[20],使用注意力机制学习句子的情感倾向权重分布,利用LSTM获取文本的语义信息。
- 4)BiGRU-CNN^[21],引入注意力机制,将BiGRU和CNN串行叠加构建情感分类模型。
- 5)H-BiGRU^[22],利用BiGRU编码表示词向量和句向量,采用注意力机制加权求和获得数据的最终表示形式。
- 6)PF-CNN^[23],分段提取句子特征,并融合词性特征与词向量区分同义词。

文本情感分类模型的准确率和F1值对比结果如表3、表4所示。

表3 准确率对比结果

Table 3 Accuracy comparison results			%
模型	IMDB数据集	yelp数据集	sentiment140数据集
CNN	72.75	73.87	76.80
BiGRU	74.85	76.95	77.54
BiLSTM+EMB_ATT	78.17	79.32	80.13
BiGRU-CNN	76.84	78.32	79.91
H-BiGRU	78.44	79.63	80.17
PF-CNN	77.43	78.97	80.02
ELMo-CNN-BiGRU	80.86	81.61	82.69

表4 F1值对比结果

Table 4 F1 value comparison result			%
模型	IMDB数据集	yelp数据集	sentiment140数据集
CNN	72.18	74.63	76.52
BiGRU	74.73	76.78	77.37
BiLSTM+EMB_ATT	78.02	79.21	79.91
BiGRU-CNN	76.77	78.27	79.86
H-BiGRU	78.39	79.51	80.04
PF-CNN	77.32	78.73	79.98
ELMo-CNN-BiGRU	80.79	81.45	82.47

由表3、表4可知:BiGRU由于能够获取评论文本的语义信息和双向的上下文信息,较CNN在准确率和F1值上获得更好的结果;BiGRU-CNN在BiGRU和CNN串联处理特征信息的基础上,融入注意力机制,较CNN和BiGRU模型更优;ELMo-CNN-BiGRU模型采用动态词向量技术进行词嵌入,通过BiGRU和CNN双通道并行提取特征,能够同时获得原始文本信息的全局特征和局部特征,与采用串联结构的BiGRU-

CNN模型相比,评论文本情感分类准确率和F1值均有所提高;ELMo-CNN-BiGRU采用自注意力机制,对原始评论文本信息进行权重处理,能够对重要信息投入更多关注,有效解决一词多义问题,与H-BiGRU、BiLSTM+EMB_ATT、PF-CNN模型对比,在评论文本情感分类方面具有更好的性能表现;ELMo-CNN-BiGRU模型在3个数据集上的准确率和F1值均为最优,与对比模型中情感分类性能最优的H-BiGRU模型相比,准确率分别提升了2.42、1.98、2.52个百分点,F1值分别提升了2.40、1.94、2.43个百分点。

4 结束语

本文提出一种ELMo-CNN-BiGRU双通道文本情感分类模型。在输入层,融合静态词向量和动态词向量进行堆叠嵌入,使得文本信息表示能够较好地根据上下文调整语义。利用自注意力机制为文本信息分配权重,获得文本内部的词依赖关系,并将其输入到BiGRU-CNN双通道中,更全面地获取文本数据的局部与全局特征。最终使用softmax分类器,实现文本情感分类。实验结果表明,在IMDB、yelp和sentiment140这3个数据集上,ELMo-CNN-BiGRU模型相较于CNN、BiGRU、BiLSTM+EMB_ATT、PF-CNN、BiGRU-CNN和H-BiGRU模型具有更优的情感分类性能。下一步将建立细粒度情感分类模型,并且通过融合注意力机制,识别结构更复杂的评论文本的情感倾向。

参考文献

[1] 何力,郑灶贤,项凤涛,等. 基于深度学习的文本分类技术研究进展[J]. 计算机工程, 2021, 47(2): 1-11.
HE L, ZHENG Z X, XIANG F T, et al. Research progress of text classification technology based on deep learning[J]. Computer Engineering, 2021, 47(2): 1-11. (in Chinese)

[2] XU D L, TIAN Z H, LAI R F, et al. Deep learning based emotion analysis of microblog texts[J]. Information Fusion, 2020, 64: 1-11.

[3] MA X H. A novel hybrid model by using convolutional neural network and long short-term memory for text sentiment analysis[J]. IEEE Access, 2020, 8: 527-533.

[4] 周锦峰,叶施仁,王晖. 基于深度卷积神经网络模型的文本情感分类[J]. 计算机工程, 2019, 45(3): 300-308.
ZHOU J F, YE S R, WANG H. Text sentiment classification based on deep convolutional neural network model[J]. Computer Engineering, 2019, 45(3): 300-308. (in Chinese)

[5] JIN N, WU J X, MA X, et al. Multi-task learning model based on multi-scale CNN and LSTM for sentiment classification[J]. IEEE Access, 2020, 8: 77060-77072.

[6] 蔡林森,彭超,陈思远,等. 基于多样化特征卷积神经网络的情感分析[J]. 计算机工程, 2019, 45(4): 169-174, 180.
CAI L S, PENG C, CHEN S Y, et al. Sentiment analysis based on multiple features convolutional neural networks[J]. Computer Engineering, 2019, 45(4): 169-174, 180. (in Chinese)

[7] WU D, ZHANG J P, ZHAO Q C. A text emotion analysis

- method using the dual-channel convolution neural network in social networks[J]. *Mathematical Problems in Engineering*, 2020(3): 1-10.
- [8] LI W, ZHU L Y, SHI Y, et al. User reviews: sentiment analysis using lexicon integrated two-channel CNN-LSTM family models[J]. *Applied Soft Computing*, 2020, 94: 106435.
- [9] UMER M, ASHRAF I, MEHMOOD A, et al. Sentiment analysis of tweets using a unified convolutional neural network-long short-term memory network model[J]. *Computational Intelligence*, 2021, 37(1): 409-434.
- [10] 李卫疆, 漆芳. 基于多通道双向长短期记忆网络的情感分析[J]. *中文信息学报*, 2019, 33(12): 119-128.
- LI W J, QI F. Sentiment analysis based on multi-channel bidirectional long short term memory network[J]. *Journal of Chinese Information Processing*, 2019, 33(12): 119-128. (in Chinese)
- [11] 王义, 沈洋, 戴月明. 基于细粒度多通道卷积神经网络的文本情感分析[J]. *计算机工程*, 2020, 46(5): 102-108.
- WANG Y, SHEN Y, DAI Y M. Sentiment analysis of texts based on fine-grained multi-channel convolutional neural network[J]. *Computer Engineering*, 2020, 46(5): 102-108. (in Chinese)
- [12] USAMA M, AHMAD B, SONG E M, et al. Attention-based sentiment analysis using convolutional and recurrent neural network[J]. *Future Generation Computer Systems*, 2020, 113: 571-578.
- [13] 陈思远, 彭超, 蔡林森, 等. 一种用于特定目标情感分析的深度网络模型[J]. *计算机工程*, 2019, 45(3): 286-292.
- CHEN S Y, PENG C, CAI L S, et al. A deep network model for specific target sentiment analysis[J]. *Computer Engineering*, 2019, 45(3): 286-292. (in Chinese)
- [14] NISTOR S C, MOCA M, MOLDOVAN D, et al. Building a twitter sentiment analysis system with recurrent neural networks[J]. *Sensors*, 2021, 21(7): 2266.
- [15] 彭祝亮, 刘博文, 范程岸, 等. 基于BLSTM与方面注意力模块的情感分类方法[J]. *计算机工程*, 2020, 46(3): 60-65, 72.
- PENG Z L, LIU B W, FAN C G, et al. Sentiment classification method based on BLSTM and aspect attention module[J]. *Computer Engineering*, 2020, 46(3): 60-65, 72. (in Chinese)
- [16] 高玮军, 杨杰, 张春霞, 等. 基于AT-DPCNN模型的情感分析研究[J]. *计算机工程*, 2020, 46(11): 53-60.
- GAO W J, YANG J, ZHANG C X, et al. Sentiment analysis research based on AT-DPCNN model[J]. *Computer Engineering*, 2020, 46(11): 53-60. (in Chinese)
- [17] BASIRI M E, NEMATIS 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.
- [18] LIU J, LIU P Y, ZHU Z F, et al. Graph convolutional networks with bidirectional attention for aspect-based sentiment classification[J]. *Applied Sciences*, 2021, 11(4): 1528.
- [19] PETERS M E, NEUMANN M, IYYER M, et al. Deep contextualized word representations[EB/OL]. [2021-06-17]. <https://arxiv.org/abs/1802.05365>.
- [20] 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, LÜ X Q, et al. Attention enhanced Bi-directional LSTM for sentiment analysis[J]. *Journal of Chinese Information Processing*, 2019, 33(2): 105-111. (in Chinese)
- [21] 胡玉琦, 李婧, 常艳鹏, 等. 引入注意力机制的BiGRU-CNN情感分类模型[J]. *小型微型计算机系统*, 2020, 41(8): 1602-1607.
- HU Y Q, LI J, CHANG Y P, et al. BiGRU-CNN sentiment classification model with attention mechanism[J]. *Journal of Chinese Computer Systems*, 2020, 41(8): 1602-1607. (in Chinese)
- [22] 胡均毅, 李金龙. 基于情感评分的分层文本表示情感分类方法[J]. *计算机工程*, 2020, 46(3): 46-52, 59.
- HU J Y, LI J L. Sentiment evaluation based hierarchical text representation method for sentiment classification[J]. *Computer Engineering*, 2020, 46(3): 46-52, 59. (in Chinese)
- [23] 周泳东, 章韵, 曹艳蓉, 等. 基于特征融合分段卷积神经网络的情感分析[J]. *计算机工程与设计*, 2019, 40(10): 3009-3013, 3029.
- ZHOU Y D, ZHANG Y, CAO Y R, et al. Sentiment analysis based on piecewise convolutional neural network combined with features[J]. *Computer Engineering and Design*, 2019, 40(10): 3009-3013, 3029. (in Chinese)