

# 结合 CNN 和 BiGRU 的双通道短文本意图识别算法<sup>①</sup>



王超, 孙喆喆, 徐飞, 马媛媛, 文雯, 汪露

(西安工业大学 计算机科学与工程学院, 西安 710021)

通信作者: 王超, E-mail: 1501873640@qq.com

**摘要:** 在短文本意图识别领域, 卷积神经网络 (CNN) 因其在局部信息提取方面的优异性能而备受关注. 然而, 由于其难以捕捉短文本语料的全局特征, 因此存在一定局限性. 针对该问题, 本文结合 TextCNN 和 BiGRU-att 的优点提出一个双通道短文本意图识别模型, 利用局部特征和全局特征更好地识别短文本的意图, 弥补模型对文本整体特征的不足. AB-CNN-BGRU-att 模型首先利用 ALBERT 多层双向 Transformer 结构对输入的文本向量化, 再将向量分别送入 TextCNN 和 BiGRU 网络模型以获取局部和全局特征. 将这两种特征进行融合, 并通过全连接层并输入 Softmax 函数得到意图标签. 实验结果表明, 在 THUCNews\_Title 数据集上, 本文提出的 AB-CNN-BGRU-att 算法准确率 (Acc) 达到了 96.68%,  $F1$  值达到了 96.67%, 相较于其他常用意图识别模型表现出更佳的性能.

**关键词:** 意图识别; ALBERT; BiGRU; 双通道

引用格式: 王超, 孙喆喆, 徐飞, 马媛媛, 文雯, 汪露. 结合 CNN 和 BiGRU 的双通道短文本意图识别算法. 计算机系统应用, 2024, 33(5): 136-143. <http://www.c-s-a.org.cn/1003-3254/9497.html>

## Dual Channel Short Text Intent Recognition Algorithm Combining CNN and BiGRU

WANG Chao, SUN Yong-Yong, XU Fei, MA Yuan-Yuan, WEN Wen, WANG Lu

(School of Computer Science and Engineering, Xi'an Technological University, Xi'an 710021, China)

**Abstract:** In the field of short-text intent recognition, convolutional neural networks (CNN) have garnered considerable attention due to their outstanding performance in extracting local information. Nevertheless, their limitations arise from the difficulty in capturing the global features of short-text corpora. To address this issue, this study combines the strengths of TextCNN and BiGRU-att to propose a dual-channel short-text intent recognition model, aiming to better recognize the intent of short texts by leveraging both local and global features, thereby compensating for the model's inadequacies in capturing overall text features. The AB-CNN-BGRU-att model initially utilizes an ALBERT multi-layer bidirectional Transformer structure to vectorize the input text and subsequently feeds these vectors separately into TextCNN and BiGRU network models to extract local and global features, respectively. The fusion of these two types of features, followed by passing through fully connected layers and inputting into the Softmax function, yields the intent labels. The experimental results demonstrate that on the THUCNews\_Title dataset, the proposed AB-CNN-BGRU-att algorithm achieves an accuracy (Acc) of 96.68% and an  $F1$  score of 96.67%, exhibiting superior performance compared with other commonly used intent recognition models.

**Key words:** intention recognition; ALBERT; BiGRU; dual channel

① 基金项目: 新型网络与检测控制国家联合地方工程实验室基金 (GSYSJ2018013)

收稿时间: 2023-11-27; 修改时间: 2023-12-29; 采用时间: 2024-01-05; csa 在线出版时间: 2024-04-01

CNKI 网络首发时间: 2024-04-03

人机问答的系统中自然语言理解 (natural language understanding, NLU) 是基础任务, 在意图识别环节可以得到准确结果, 可以降低整体自然语言理解的复杂度, 给予问句的文本对应标签后再进行处理. 相较于没有或者效果不佳的意图识别环节, 准确的分类可以极大程度的减轻后续的任务量. 而意图识别任务从输入和输出结果上来看, 可以将该功能模块划分到自然语言处理 (natural language processing, NLP) 的文本分类任务中, 模块的目标是将用户输入的对话文本经过算法、模型得到概率分数, 最终将这句文本分类到对应的意图标签中, 为下游的任务做铺垫.

随着神经网络技术的不断进步, 以卷积神经网络为基础的意图识别模型已得到大量的研究, 各类模型也取得了较好的成果, 但是卷积神经网络仅使用局部特征, 忽略了文本全句带来的信息. 因此, 为了进一步结合语料的局部特征和全局特征来分析问句, 本文提出了一种多池化类型的 TextCNN 和 BiGRU-att 相结合的双通道短文本意图识别模型 AB-CNN-BGRU-att (ALBERT-TextCNN-BiGRU-attention) 模型. 利用双通道得到不同特征, BiGRU-att 得到文本的全局特征, 同时在 TextCNN 的传统池化层中, 使用不同大小卷积核和不同池化策略相结合, 得到更加丰富的局部特征. 实验结果证明了使用 AB-CNN-BGRU-att 模型进行意图分类任务, 相比于其他现流行的意图识别模型在 THUC-News\_Title 数据集上性能有所提升.

## 1 相关工作

人机问答中存在大量的短文本, 询问句通常是口语表述, 整句字数少, 包含信息不够丰富, 这使一般的意图识别模型很难达到一个让人满意的结果. 以医疗问答为例, “我头疼去哪找医生?”这种简短的、口语表达在询问医生的时候, 表达对应的意图是希望知道“所属科室”而不是问“医院”或者一个地名. 自然语言理解的结果会直接给整体问答的答案带来影响, 如果上游对问句理解不准确就会导致答非所问的结果.

Kim<sup>[1]</sup>将用于计算机视觉的卷积神经网络用于对文本的处理上, 提出后来作为文本分类经典的 TextCNN 模型. 将句子矩阵用向量表示, 不同尺寸的卷积核进行卷积操作类似  $n$ -gram 的特征抽取, 将得到的特征进行一个下采样的操作, 将压缩后的特征放入全连接层得到分类的标签. 在 TextCNN 提出以后, 大家发现适用

于图像处理的神经网络在文本领域也可以有不错的效果, 也是给了众多学者一个新的方向和灵感. 代令令<sup>[2]</sup>提出针对中文文本分类使用 fastText 不仅可以提高系统准确度同时还可以缩短分类时间, 这对问答系统很重要. 张倩<sup>[3]</sup>也同样意识到缩短意图识别的时间十分重要, 张倩提出结合自注意力机制的 GRU 用于意图识别, 在保险问题数据集上对比其他方法不仅准确度达到最佳, 所用时间也大幅缩短. 侯雪亮等<sup>[4]</sup>和 Wang 等<sup>[5]</sup>意识到经典的 TextCNN 模型无法很好地处理短文本问题. 侯雪亮等选择将短文本关键词重构后得到的特征作为 FastCNN 的输入, 将原始的特征值作为 TextCNN 模型的输入, 使用多神经网络混合的方法提高模型对短文本的分类能力. Wang 等则是提出了采用非线性滑动法和  $n$ -gram 语言模型获得丰富的文本特征, 经实验表明, 作者提供的方法可以显著改善短文本分类过程中的分类结果. 马思丹等<sup>[6]</sup>对 Word2Vec 进行改进, 先得到文本关键词的重叠部分和非重叠部分, 再通过计算文本的相似度用 KNN 来进行分类. 实验表明, 马思丹等提出的方法, 对比传统的分类模型效果有所提升. 孙红等<sup>[7]</sup>和迟海洋等<sup>[8]</sup>在输入的词向量模型上, 认为静态词向量无法很好地表征句子, 所以在提出的 BBGA 模型中, 对嵌入层没有选择 Word2Vec 来表示词向量, 而是选用语义表征能力更强的 BERT 来训练, 在嵌入层让模型得到更加丰富的信息, 将训练得出的词向量放入 BiGRU 得到句子的全局特征, 最后利用注意力机制突出关键信息得到分类标签. 孙红等的模型在 THUC-News 数据集上的分类效果优于 TextCNN 模型, 迟海洋等的模型也在自己的数据集上证明了该方法能提高意图识别的准确性. 郭云雪<sup>[9]</sup>提出了将 BERT 模型和 Word2Vec 模型融合作为词嵌入层的模型, 丰富的语句特征同样提升了模型的性能. 温超东等<sup>[10]</sup>提出了 ALBERT-BiGRU 模型, 同样, 作者在词向量的选择上, 也没有选择传统的静态词向量, 但是在动态词向量的选择上, 作者考虑到复杂的 BERT 模型对硬件的要求, 从而选用了更为轻减的 ALBERT 语言模型. 同时, 作者考虑到所涉及领域的多为中长距离词, 选择了使用 BiGRU 为模型主体, 保留长语句的语义信息. 在效果上, ALBERT 作为嵌入层的效果和 BERT 类似, 同样明显优于 Word2Vec 和 GloVe. 李飞鸽等<sup>[11]</sup>和黄星宇<sup>[12]</sup>使用更为轻减的 ALBERT 语言模型获取动态词向量, 将向量送入 TextCNN 层处理适应各自领域的意图识

别任务. 曾诚等<sup>[13]</sup>也同样使用 ALBERT 模型做为嵌入层, 将得到的特征作为 BiGRU 层的输入, 以 3 大视频软件的弹幕为数据集进行对比实验, F1 值优于其他模型.

但是, 在现在的研究中, 有学者<sup>[14-18]</sup>意识到, 虽然短文本语料更在意局部特征, 但是也不能完全忽视了全局特征对结果的影响, 提出了多通道的方法, 同时兼顾了局部特征和全文的信息, 双通道分别对来自嵌入层的输入进行处理, 再融合特征进行分类. 吴迪等<sup>[19]</sup>提出在双通道的模型基础上, 改进嵌入层, 同时使用 ELMo 和 GloVe, 将静态词向量和动态词向量相结合. 实验表明, 该方法在 IMDB 等数据集上相比其他常用模型均有提升.

在以上的研究基础上, 我们首先结合文献研究提出一种适合意图识别的双通道模型, 兼顾文章局部特征和全局特征, 嵌入层使用 ALBERT 预训练语言模型让模型整体更加轻便. 其次, 在双通道模型中, 提出一种改进的 TextCNN 模型, 结合多种池化策略混合池化方式得到更丰富的局部特征. 最后通过实验验证了本

文提出的算法相比其他算法具有一定的优越性.

## 2 AB-CNN-BGRU-att 模型结构

以常见的 BERT-TextCNN 模型为例, TextCNN 主要在卷积层和池化层提取信息, 在卷积层使用卷积核滑动大小为取词长短, 在池化层使用不同池化策略得到局部特征. 但是 TextCNN 忽略了问句整句文字的信息, 为了得到全局特征, 本文在 BERT-TextCNN 的基础上, 添加了 BiGRU 网络, 同时用更加轻量化的 ALBERT 代替了原模型中的预训练语言模型 BERT, ALBERT 在减少了模型参数的同时, 也能像 BERT 一样输出丰富语义的词向量, 这些丰富的向量分别传入 TextCNN 和 BiGRU 中, 在 TextCNN 中通过不同大小卷积核和不同池化策略提取丰富的局部特征, 在 BiGRU 中得到全局特征, 得到局部特征和全局特征后, 进而将两种特征融合, 经过 Dropout 和 Softmax 输出概率值实现多分类. 模型架构如图 1 所示.

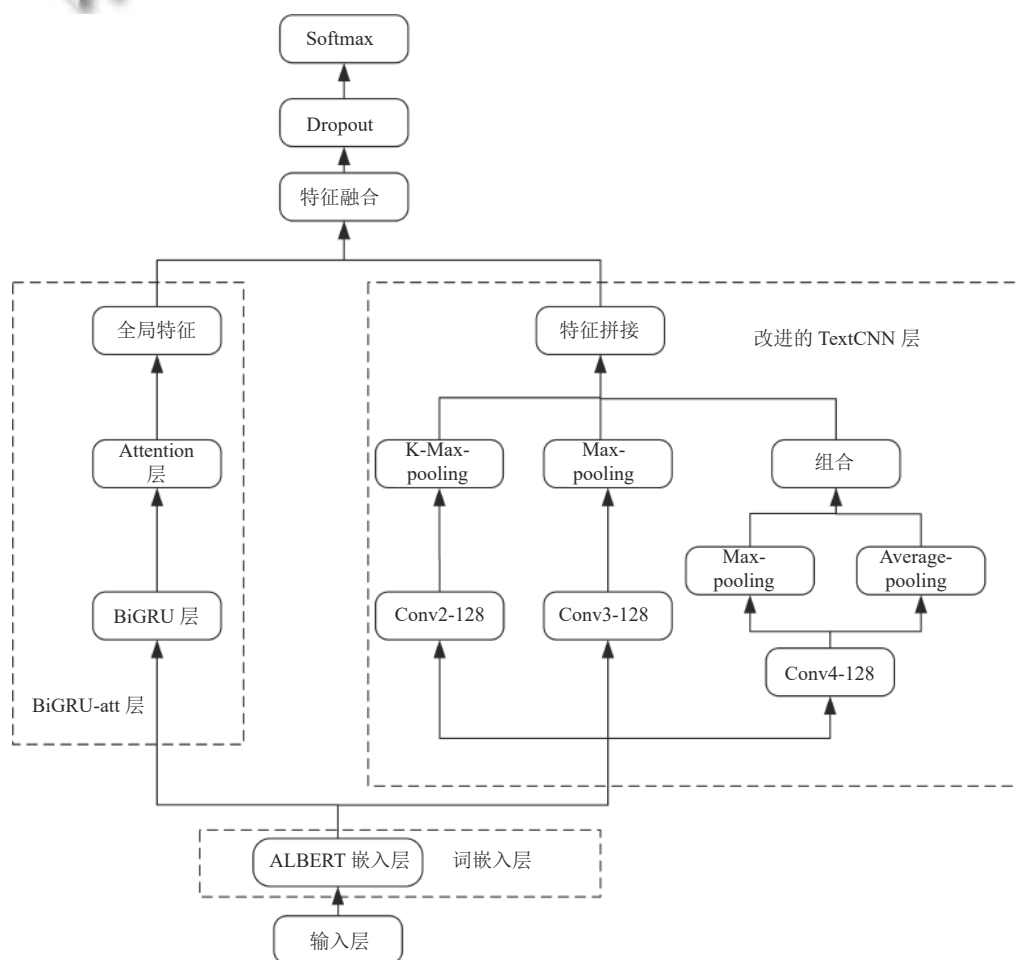


图 1 AB-CNN-BGRU-att 模型架构图



### 2.1 词嵌入层

常用的词嵌入模型有很多,比如 Word2Vec、GloVe 和 BERT 等,其中 BERT 已经在大量 NLP 学者的实验中得到肯定,是最优的几种模型之一。ALBERT 预训练语言模型是 BERT 的变种,简化了原生的 BERT,但两者效果相似。例如在 ALBERT 官方论文中给出的数据显示,在几个具有代表性的现有任务中,ALBERT-large 的性能已经和 BERT-base 接近,但是后者参数量却是前者的 6 倍,时间也是前者的近 3 倍。所以,本文在选取模型的词嵌入层时选择了 ALBERT,其模型结构如图 2 所示。

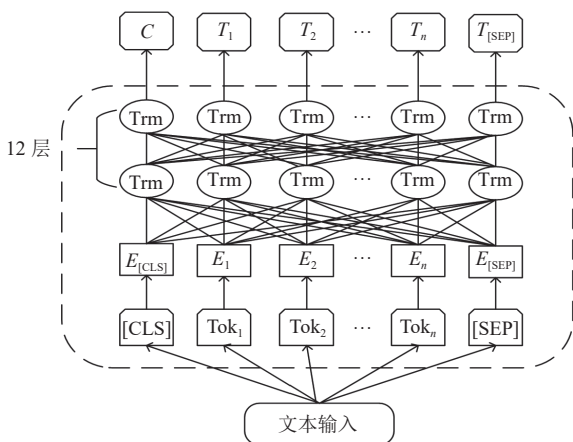


图 2 ALBERT 模型结构

输入的文本进入词嵌入层,会在开始和结尾分别打上[CLS]和[SEP]的标签,用于确定一句话的位置。将文本序列化得到的  $E_n$  向量经过 Transformer 编码器提取向量特征得到  $T_n$ 。ALBERT 和 BERT 相同,两者均使用了 Transformer 中的编码部分,该部分由多个相同的网络层组成,每层都是由残差网络连接“Multi-head attention”层和“FeedForward”层组成。该环节中,“Multi-head attention”层是核心模块,文本序列中的每一个 query、key 和 value 组成输入向量  $Q$ 、 $K$ 、 $V$ ,经过“Multi-head attention”,具体计算公式如式(1)–式(3)所示。

$$head_t = \text{Attention}(QW_t^Q, KW_t^K, VW_t^V) \quad (1)$$

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_t}}\right)V \quad (2)$$

将输出矩阵拼接:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(head_1, head_2, \dots, head_h)W^0 \quad (3)$$

其中,  $W^0$  为权重矩阵,目的是调整最后矩阵的维度,与序列长度保持一致,  $W_t^Q$ 、 $W_t^K$ 、 $W_t^V$  是每个  $Q$ 、 $K$ 、 $V$  向量的权重矩阵,  $d_t$  是每个  $Q$ 、 $K$ 、 $V$  向量的维度大小。

### 2.2 BiGRU-att 层

为获取文本语料的全局特征,得到模型对文本大意的了解,本文在 ALBERT 得到词向量后连接 BiGRU 层,在该层针对全句获取更多特征信息。GRU 和 LSTM 均是针对 RNN 梯度消失问题提出的改进模型,两者皆提出利用门控机制来选择记忆或者遗忘信息,从而解决 RNN 存在的梯度消失问题。

同时 GRU 又是 LSTM 的改进模型,在 LSTM 的 3 个门的基础上进行改进,将门的个数减为两个,用更新门来代替 LSTM 输入门和遗忘门的功能。大量实验验证在数据量不是特别庞大时,GRU 的性能和 LSTM 相差不大,同时因为门个数减少,参数更少的 GRU 训练的时间更短,具体实验详见第 3.5 节,在实验数据的支持下,文本模型在提取全局特征的工作上,选择了 GRU 网络为基础。GRU 更新门和重置门的相关计算公式如式(4)和式(5)所示:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad (4)$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (5)$$

其中,  $z_t$  是 GRU 的更新门,  $r_t$  是 GRU 的重置门,  $W_z$  和  $W_r$  分别是更新门和重置门的权重矩阵,  $\sigma$  是 Sigmoid 函数,作用是将目标数据变为 0–1 范围内的数值。  $z_t$  和  $r_t$  代表了从  $h_{t-1}$  到  $h_t$  一个时间步的影响程度。

GRU 模型当前时刻的输出与上一时刻有关,具体计算公式如式(6)所示:

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t]) \quad (6)$$

其中,  $\tilde{h}_t$  记录的是当前时刻,  $W$  为候选隐含状态的权重矩阵,  $*$  是对位元素相乘。当  $r_t$  接近 0 时,丢弃上一时间步  $h_{t-1}$  的隐藏状态,反之,当  $r_t$  接近 1 时,保留上一时间步  $h_{t-1}$  的隐藏状态。

GRU 模型最后输出的结果与当前时刻的输出和上一时刻的隐藏状态都有关系,具体计算公式如式(7)所示:

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (7)$$

其中,  $h_t$  为最后输出结果。当  $z_t$  越接近 1 的时候,代表对当前节点的候选隐藏状态记忆的越多。当  $z_t$  越接近 0 的

时候,代表对原本隐藏状态遗忘的越多。

GRU 中的隐藏状态是单向传递,只在意输入的当前状态,但是文本信息中上下文对当前的状态都有影响,不能单纯地考虑当前的输出,GRU 的单向传播无法得到后来信息对先前状态的影响,为了得到完备的输出信息,本文使用 GRU 的变体 BiGRU. BiGRU 由两个方向相反的 GRU 叠加实现,输出信息受前后两个方向的结果影响,最后的输出结果如式 (8) 所示, BiGRU 的模型结构如图 3 所示。

$$h_i^{(i)} = [\vec{h}_i^{(i)}, \overleftarrow{h}_i^{(i)}] \quad (8)$$

其中,  $\vec{h}_i^{(i)}$  表示第  $i$  个文本经过前向 GRU 得到的信息,  $\overleftarrow{h}_i^{(i)}$  表示第  $i$  个文本经过后向 GRU 得到的信息.  $h_i^{(i)}$  则是这个文本经过 BiGRU 得到的最后的结果。

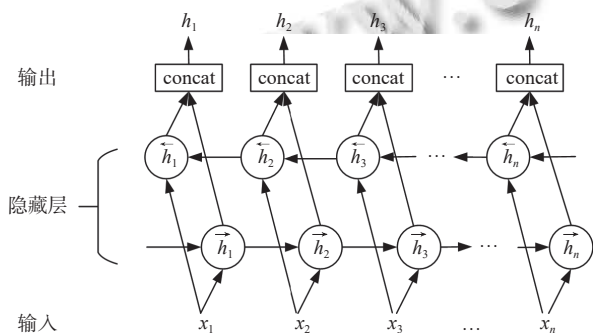


图 3 BiGRU 网络结构

注意力机制是为了给文本中的词对应的权重,使对用的结果可以根据权重的大小做出对重要特征的偏向,让模型将关注度聚焦在权重得分更高的词上提高模型分类的准确性. Attention 层首先对每个 BiGRU 模型输出向量相应位置计算各自的词语权重,该部分最终的句子表示为所求出的权重与对应位置特征向量的加权和. BiGRU-att 层让 BiGRU 模型在本身的基础上更好的聚焦了句子中权重得分高的重点词,从而更好地获得输入文本的全局特征。

### 2.3 改进的 TextCNN 层

原始的 TextCNN 模型有嵌入层、卷积层、池化层和全连接层.改进的 TextCNN 模型由多个大小不同的卷积层、不同池化策略的池化层和全连接层组成,模型结构如图 4 所示。

图 4 中改进的 TextCNN 层的卷积部分通过大小不同的卷积核对文本特征进行卷积,卷积核的宽度和 Token Embedding 的维度一致,向量矩阵的行信息代表

词,当卷积核的宽度和 Token Embedding 的维度不一致时,卷积核就无法提取完整的词信息.该矩阵通过非线性激活函数得到特征映射矩阵  $c = [c_1, c_2, \dots, c_n]$ ,特征公式如式 (9) 所示:

$$c_i = f(w \cdot x_{i:i+h-1} + b) \quad (9)$$

其中,  $f$  为激活函数,  $w$  为卷积核的权重矩阵,  $b$  为偏置项。

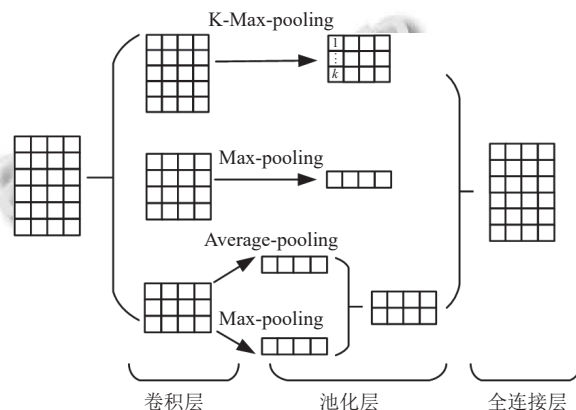


图 4 改进的 TextCNN 模型结构

短文本的问句中,句子本身长度有限,由于双通道的结构,TextCNN 模型更在意对文本的局部特征提取.询问词一般长度较短,本文模型在卷积时,对局部特征提取选取的卷积核大小也没有使用通用的 3、4、5,而是选取了大小分别为 2、3、4 的卷积核,同时,对于大小不一的卷积核在卷积时进行调整,采取不同的池化策略。

在卷积核大小为 2 的卷积操作后,模型采用 K-Max-pooling 池化方法,卷积后的特征经过 K-Max-pooling 层进行池化,获取采样过程中得分前  $K$  个的较大值.在这里本文没有选用文本分类中最常见的最大池化方法,因为在卷积核大小为 2,步长为 1 的卷积操作中,得到的信息是最丰富的,最大池化方法忽视了一些可能多次出现的特征,对于这些特征最大池化方法只能看到一次.在 K-Max-pooling 池化方法中,可以保留得分较高的  $K$  个特征,同时还可以保留部分特征间的相对顺序信息。

在卷积核大小为 3 的卷积操作后,模型采用最大池化方法,得到卷积后产生的最大特征,忽略文本中不重要的信息,抛弃弱的特征,减小噪音影响,突出关键词。

在卷积核大小为 4 的卷积操作中,模型分别采用最大池化策略和平均池化策略,然后将两次得到的特

征拼接. 在池化操作中, 最大池化和平均池化信息是互补的, 平均池化策略考虑到每一个词的信息, 最大池化策略只考虑得分最高的特征.

在通过 3 个大小不一的卷积核和对应的池化操作后, 得到了向量不同的局部特征, 为了避免过拟合, 模型在 TextCNN 的池化层后设置 Dropout 层, 提高泛化能力. 将改进的 TextCNN 层得到的组合特征和 BiGRU-att 层得到的全局特征拼接融合得到最终的特征向量, 通过最后的全连接层得到分类结果, 计算公式如式 (10) 所示:

$$Z = \text{Softmax}(W_Z \cdot F + b) \quad (10)$$

其中,  $Z$  为预测的意图标签结果,  $\text{Softmax}$  为激活函数,  $W_Z$  为全连接层权重,  $F$  为最终的特征向量,  $b$  为偏置项.

#### 2.4 AB-CNN-BGRU-att 算法

AB-CNN-BGRU-att 算法通过将输入的文本语料经过模型处理, 得到对应意图标签的概率进行意图识别. 具体算法流程如算法 1 所示.

算法 1. AB-CNN-BGRU-att 算法

输入:  $S=(s_1, s_2, s_3, \dots, s_n)$ ,  $S$  为输入文本序列.

输出: 意图识别标签结果.

1. 数据预处理, 导入训练集, 测试集
2. 加载 ALBERT 模型, 获得动态词向量 Token //Token 简称为  $T$
3.  $\text{conv}_{output_{1\sim n}} = \text{Conv}_{1\sim n}(T)$ ; //在 CNN 卷积层, 采用不同的卷积核进行卷积
4.  $\text{pooling}_{output_{1\sim n}} = \text{Pooling}(\text{Conv}_{output_{1\sim n}})$ ; //对不同卷积核卷积后的输出, 采用文章中提到的不同的池化策略
5.  $\text{cnn}_{output} = \text{Concat}(\text{pooling}_{output_{1\sim n}})$ ; //得到输入文本的局部特征
6.  $\text{forward} = \text{GRU}(T)$ ; //前向 GRU
7.  $\text{backward} = \text{GRU}(T)$ ; //后向 GRU
8.  $\text{bigru}_{output} = \text{Concat}(\text{forward}, \text{backward})$ ; //输入文本的全局特征
9.  $\text{output} = \text{Concat}(\text{cnn}_{output}, \text{bigru}_{output})$ ;
10.  $\text{dropout} = \text{Dropout}(\text{output})$ ; //为避免过拟合, Dropout 设置为 0.5
11.  $\text{dense} = \text{Dense}(\text{dropout})$ ; //对输出的张量进行降维
12.  $\text{out} = \text{Softmax}(\text{dense})$ ; //得到意图识别的标签结果
13. END

### 3 实验与结果分析

#### 3.1 实验数据集

实验使用第 1 个数据集为 THUCNews\_Title 数据集, 该数据集抽取了 THUCNews 的标题集合, 共计 20 万条, 文本长度大多数不超过 30 字符合短文本的要求. 数据集包含体育类、时政类、教育类、科技类等总计 10 个类别, 数据量分布平均.

实验选择的第 2 个数据集是阿里天池实验室的医疗搜索检索词意图分类数据集 (KUAKE-QIC) 来验证模型性能. 在医疗搜索中, 对搜索问题的意图分类可以极大提升搜索结果的相关性, 特别是医学知识具备极强的专业性, 对问题意图进行识别也有助于融入医学知识来做增强搜索结果的性能, KUAKE-QIC 就是针对这样的背景产生. 该数据集共有病情诊断、病因分析、治疗方案、就医建议、疾病描述等 11 种分类, 训练集数据 6931 条, 验证集数据 1955 条, 测试集数据 1994 条. 经过统计, 有 6684 条数据的长度小于 30 字, 约占总数据集的 96%, 符合实验数据集对短文本医疗数据的期待.

#### 3.2 实验设置

AB-CNN-BGRU-att 模型中改进版 TextCNN 参数设置如表 1 所示.

参数	值
卷积核大小	2, 3, 4
词向量维度	384
激活函数	ReLU
学习率	1E-5
Dropout	0.5
Batch size	128

AB-CNN-BGRU-att 模型中 BiGRU-att 参数设置如表 2 所示.

参数	值
隐藏层大小	256
词向量维度	384
激活函数	ReLU
Dropout	0.2
Batch size	128

#### 3.3 实验结果

通过对上述模型超参数的调节, 使用 THUCNews\_Title 对模型进行训练和验证, 模型的 loss 和 accuracy 均能很好的收敛.

验证结果如图 5 所示, THUCNews\_Title 测试集数据分布平均, 模型得到每一类别的结果并进行对比, 由图 5 可知, 模型对每一类的分类结果各评价指标均不低于 90%, 其中科技类, 股票类, 社会类 3 种分类下的文本各项评价指标较低, 分析数据集可以得出原因, 科技类的文本更加复杂不常见, 股票的文本以及社会文本和其他标签下的文本相近容易混淆, 导致了分类准确性不高.



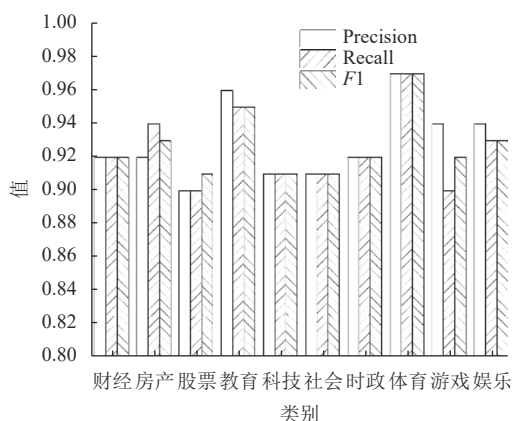


图5 模型验证结果

对于 KUAKE-QIC 数据集, 实验共训练 20 个 epochs, 模型的 loss 和 accuracy 均能很好地收敛. 测试集准确率达到 86.02%, 在 CBLUE3.0 排行榜中, 可以排到前三.

### 3.4 对比实验

为了验证本文模型选取的 BiGRU-att 层相比 BiLSTM-att 层效果更好, 本研究设置对照实验, 分别使用两层网络加入本文模型, 在网络隐藏层个数, Dropout 大小, epoch 个数等其他参数保持一致的基础上, 对每个 epoch 消耗的平均时间、训练总时间、最后的准确度和 F1 值进行对比, 两个网络每个 epoch 消耗的时间对比如图 6 所示, 两个网络总体对比如表 3 所示.

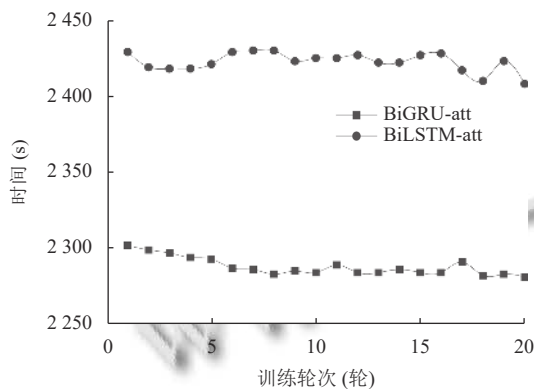


图6 网络用时对比

表3 BiGRU-att 和 BiLSTM-att 对比

网络层	平均时长 (s)	总时长 (s)	Acc (%)	F1 (%)
BiGRU-att	2288.9	45779	90.83	90.64
BiLSTM-att	2423.6	48472	90.45	90.41

在训练模型所需时间上, 使用 BiGRU-att 层明显优于使用 BiLSTM-att 层, 模型分别使用两个网络层得到最后的准确度和 F1 值也相近. 在考虑到训练代价和训练结果之后, 本文选取了 BiGRU-att 层作为提取全

局特征的网络层.

同时, 为验证本文模型与其他模型在同等条件下的优越性, 本文设置对比实验进行对比, 对比实验模型信息如下.

1) SAttBiGRU: 先使用 BiGRU 网络得到全局特征, 再使用 Self-Attention 对上游输出的特征进行加权达到增强文本特征的目的, 最后得到更加丰富的特征信息后进行分类.

2) Self-Attention-CNN: 在基础的 TextCNN 上结合 Self-Attention, 对 TextCNN 的嵌入层得到的向量使用 Self-Attention 进行加权, 让卷积层得到的文本信息更加紧密, 使用 Max-pooling 进行池化, 再通过全连接层经过 Softmax 输出分类结果.

3) BiGRU-MCNN: 分别使用 BiGRU 模型提取全局特征, 多通道 CNN 模型提取细致局部特征, 将两类特征融合, 再通过全连接层经过 Softmax 输出分类结果.

4) MC-AttCNN-AttBiGRU: 首先采用注意力机制分别对多通道 CNN 和 BiGRU 进行加权, 再将得到的文本特征向量进行拼接后通过全连接层经过 Softmax 进行分类.

对比实验结果如表 4 所示, 使用准确率 (Acc)、精度 (Pre)、召回率 (Recall) 和 F1 值作为评价指标. 结果表明, AB-CNN-BGRU-att 模型在各项性能上相较于其他模型均有一定优势. 对比现流行模型在使用相同数据集 THUCNews\_Title 的实验中, 本文模型的各项指标相对其他模型的最优仍有 1-2 个百分点的提升, 对比实验的数据证明了本文提出的 AB-CNN-BGRU-att 模型具有一定优越性.

表4 实验结果对比 (%)

模型	Acc	Pre	Recall	F1
SAttBiGRU	96.16	96.20	96.16	96.17
Self-Attention-CNN	94.85	94.89	94.85	94.85
BiGRU-MCNN	95.43	95.45	95.43	95.43
MC-AttCNN-AttBiGRU	95.93	95.98	95.93	95.93
AB-CNN-BGRU-att	96.68	96.68	96.67	96.67

### 3.5 消融实验

为了验证本文提出模型针对短文本分类的有效性, 对模型设置了消融实验, 分解了各个局部网络, 分别设置了 TextCNN、改进的 TextCNN、BiGRU-att、AB-CNN-BGRU-att, 消融实验的词嵌入层均为 ALBERT. 消融实验结果如表 5 所示.

由表 5 可以看出, 基础的 TextCNN 模型在分类结

果上效果并不理想,各项指标均没有达到90%,原因可能在于基础的TextCNN对于短文本的处理效果不佳,改进的TextCNN模型各项指标有了明显的提高,实验数据证明,使用不同大小卷积核和不同池化策略相结合,可以得到更加丰富的局部特征,从而提高模型性能。BiGRU-att模型的效果略逊于改进的TextCNN模型,但实验也得到了较为理想的结果,证明了全局特征在意图识别任务中也同样重要。AB-CNN-BGRU-att模型结合了局部特征和全局特征,实验结果比改进的TextCNN和BiGRU-att都有近两个百分点的提升。

表5 消融实验结果(%)

模型	Acc	Pre	Recall	F1
TextCNN	89.96	89.90	89.96	89.90
改进的TextCNN	94.85	94.89	94.85	94.85
BiGRU-att	94.00	94.17	94.00	94.90
AB-CNN-BGRU-att	96.68	96.68	96.67	96.67

#### 4 结论与展望

本文结合ALBERT、BiGRU-att以及改进的TextCNN模型,提出一种结合CNN和BiGRU-att的双通道短文本意图识别模型。该模型将ALBERT得到的向量分别送入BiGRU-att网络模型和TextCNN模型,在BiGRU-att网络模型得到全局特征,在采用了多种池化策略混合池化方式的TextCNN得到局部特征,将两种特征融合后经过全连接层输入Softmax得到分类结果。通过实验,在公开的数据集上,与以上4种模型在相同数据集的实验结果对比,最后的对比实验数据证明了本文提出的模型,比现流行的模型具有优越性,在各项评价指标的最优值上仍有一定的提升,实验数据证明本文提出的模型能够得到更准确的意图识别结果。

虽然本文的模型可以有效地识别出短文本的意图,但是算法模型仍然离不开大量已标注的数据集作为基础,未来的工作希望可以使用半监督甚至无监督的模式来训练模型,做到在减少人工标注工作量的同时,提高模型性能。

#### 参考文献

1 Kim Y. Convolutional neural networks for sentence classification. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha: ACL, 2014. 1746–1751.

2 代令令. 基于fastText的问答系统用户意图识别与关键词抽取研究[硕士学位论文]. 南宁: 广西大学, 2018.

3 张倩. 基于问答系统的问句意图识别方法研究[硕士学位论文]. 重庆: 重庆大学, 2020.

4 侯雪亮, 李新, 陈远平. 基于多神经网络混合的短文本分类模型. 计算机系统应用, 2020, 29(10): 9–19. [doi: 10.15888/j.cnki.csa.007493]

5 Wang HT, He J, Zhang XH, et al. A short text classification method based on  $n$ -gram and CNN. Chinese Journal of Electronics, 2020, 29(2): 248–254. [doi: 10.1049/cje.2020.01.001]

6 马思丹, 刘东苏. 基于加权Word2Vec的文本分类方法研究. 情报科学, 2019, 37(11): 38–42.

7 孙红, 陈强越. 融合BERT词嵌入和注意力机制的中文文本分类. 小型微型计算机系统, 2022, 43(1): 22–26.

8 迟海洋, 严馨, 周枫, 等. 基于BERT-BiGRU-Attention的在线健康社区用户意图识别方法. 河北科技大学学报, 2020, 41(3): 225–232. [doi: 10.7535/hbkd.2020yx03004]

9 郭云雪. 基于深度学习的人机对话中短文本意图识别[硕士学位论文]. 哈尔滨: 哈尔滨理工大学, 2020.

10 温超东, 曾诚, 任俊伟, 等. 结合ALBERT和双向门控循环单元的专利文本分类. 计算机应用, 2021, 41(2): 407–412.

11 李飞鸽, 王芳, 黄树成. 基于ALBERT与TextCNN的中文文本分类研究. 软件导刊, 2023, 22(4): 27–31. [doi: 10.11907/rjdk.221591]

12 黄星宇. 基于ALBERT的医疗问答系统设计与实现[硕士学位论文]. 成都: 电子科技大学, 2022.

13 曾诚, 温超东, 孙瑜敏, 等. 基于ALBERT-CRNN的弹幕文本情感分析. 郑州大学学报(理学版), 2021, 53(3): 1–8.

14 李洋, 董红斌. 基于CNN和BiLSTM网络特征融合的文本情感分析. 计算机应用, 2018, 38(11): 3075–3080.

15 李启行, 廖薇, 孟静雯. 基于注意力机制的双通道DACRNN文本分类模型. 计算机工程与应用, 2022, 58(16): 157–163. [doi: 10.3778/j.issn.1002-8331.2101-0196]

16 邢鑫, 孙国祥. 基于双通道词向量的ACRNN文本分类. 计算机应用研究, 2021, 38(4): 1033–1037.

17 康雁, 李晋源, 杨其越, 等. 基于双通道词向量的卷积胶囊网络文本分类. 计算机工程, 2019, 45(11): 177–182.

18 宋中山, 牛悦, 郑禄, 等. 多尺度CNN卷积与全局关系的中文文本分类模型. 计算机工程与应用, 2023, 59(20): 103–110. [doi: 10.3778/j.issn.1002-8331.2209-0093]

19 吴迪, 王梓宇, 赵伟超. ELMo-CNN-BiGRU双通道文本情感分类模型. 计算机工程, 2022, 48(8): 105–112.

(校对责编: 张重毅)