

# 基于 BERT-BILSTM 的医疗文本关系提取方法

龚汝鑫, 余肖生

(三峡大学 计算机与信息学院, 湖北 宜昌 443002)

**摘要:**健康医疗文本关系提取可充分利用医疗资源,为构建医院系统和相关知识图谱奠定基础,但健康医疗文本上下文联系紧密,内容结构复杂,使用传统的机器学习方法无法充分学习并利用文本中所包含的信息,且由于未针对文本中包含的医疗领域专业名词进行处理,使研究所需的重要实体流失,导致准确率不高。因此,提出了一种基于 BERT 和 BILSTM 融合的健康医疗文本关系提取方法,在预处理阶段进行医疗关键词提取,使用 BERT 语言模型进行词嵌入,再结合 BILSTM 和注意力机制进行特征处理,最后使用 Softmax 分类器输出类别概率值,确定实体间关系类别。基于两个临床医疗数据集的实验验证结果,与单向 LSTM、CNN、BIGRU 等模型进行比较分析,BERT-BILSTM-ATT 模型表现最优,精确率提高 3.35% 以上、召回率提高 1.28% 以上、F1 值提高 2.58% 以上,基于 BERT 和 BILSTM 融合的健康医疗文本关系提取方法能准确有效地预测健康医疗文本中实体之间存在的关系类别。

**关键词:**关系提取;双向长短期记忆神经网络;注意力机制;健康医疗文本;BERT

**中图分类号:**TP391.1

**文献标识码:**A

**文章编号:**1673-629X(2022)04-0186-07

doi:10.3969/j.issn.1673-629X.2022.04.032

## Relation Extraction Method of Medical Texts Based on BERT-BILSTM

GONG Ru-xin, YU Xiao-sheng

(School of Computer and Information, Three Gorges University, Yichang 443002, China)

**Abstract:**Relation extraction method can make full use of medical resources in healthy medical texts and lay the foundation for the construction of hospital system and related knowledge graph. However, the context of healthy medical texts are closely related and the content structure is complex. Traditional machine learning methods cannot fully learn and use the information in the texts, and the medical domain terms are not processed in the texts. The important entities needed in the research are lost, resulting in low accuracy. Therefore, we propose a relation extraction method of healthy medical texts based on BERT and BILSTM. In the preprocessing stage, medical keywords are extracted, words are embedded by using the BERT language model, and then features are processed by BILSTM and attention mechanism. Finally, the Softmax classifier is used to output the probability value of the category to determine the relation category between entities. Based on the experimental results of two clinical data sets, compared with unidirectional LSTM, CNN, BIGRU and other models, BERT-BILSTM-ATT model showed the best performance, with the precision increased by more than 3.35%, the recall increased by more than 1.28%, and the F1-Score increased by more than 2.58%. The proposed relation extraction method of healthy medical texts based on BERT and BILSTM can accurately and effectively predict the relation categories between entities in healthy medical texts.

**Key words:**relation extraction; BILSTM; attention mechanism; healthy medical texts; BERT

## 0 引言

健康医疗大数据是国家重要的基础性战略资源,其应用发展将带来健康医疗模式的深刻变化,有利于提升健康医疗服务效率和质量<sup>[1]</sup>。其中,健康医疗文本数据是健康医疗大数据的重要组成部分之一,多为包含患者病史、现状、诊断、检查、治疗信息的非结构化长文本,并且上下文结构联系紧密、内容专业性强,容

易出现医疗信息抽取难度大、利用率低的情况<sup>[2]</sup>。

2010 年之前多采用机器学习方法来处理文本中的信息,但时间成本投入大,且未利用语言之间的相关性,无法充分学习、利用所包含的信息<sup>[3-4]</sup>。而后利用神经网络紧密联系上下文的特点,让 CNN、RNN 及变体 LSTM、GRU 在文本提取和分类研究中取得不错表现<sup>[5-7]</sup>。为减少重要信息的流失,张志昌等人将

收稿日期:2021-04-28

修回日期:2021-08-30

基金项目:国家重点研究发展计划资助项目(2016YFC0802500)

作者简介:龚汝鑫(1998-),女,硕士研究生,CCF 会员(H1699G),从事健康医疗大数据分析技术研究;通讯作者:余肖生(1973-),男,博士,副教授,CCF 会员(98980M),从事健康医疗大数据分析研究。

BIGRU 与注意力机制相结合,重点处理影响力较强的文本信息<sup>[8-9]</sup>。

现有的关系提取方法没有针对性地考虑如何获取和充分利用医疗领域文本包含的信息,该文提出一种 BERT 和双向长短期记忆神经网络 (bidirectional long short-term memory, BiLSTM) 融合的关系提取方法,来获取健康医疗文本中实体间隐含的价值信息。在预处理阶段,针对医疗词语进行关键词提取,减小医疗实体提取难度;再使用 BERT 模型进行词嵌入,将词向量、位置向量、句子级特征输入到 BiLSTM 中进行处理,最后结合注意力机制来优化特征向量,从而充分理解和利用文本中的重要信息,提高准确率。

## 1 相关工作

关系提取的实质是关系分类,即确定两实体间的关系类别。2010 年之前多采用人工标注特征结合浅层分类模型来进行分类,Kambhatla 等人采用模式匹配方法,利用基于文本特征的统计模型进行关系提取,取得一定成果<sup>[3]</sup>。2010 年之后,文本分类方法逐渐向深层模型过渡,Liu 等人使用 CNN 模型处理特征时,结合实体语义信息,将多个同义词使用同一向量作为模型输入值<sup>[4,10]</sup>。Mikolov 等人利用 RNN 将词语与上下文信息联系起来,实验结果表明,RNN 模型比 CNN 模型更适合用于文本分类研究<sup>[11]</sup>。LSTM-RNN<sup>[12]</sup>、LSTM-CNN<sup>[13]</sup>、LSTM-GRU<sup>[14-15]</sup>等混合神经网络模型在关系提取、情感分类等任务中的准确率进一步提高。在单向神经网络结构取得一定成果后,研究目光转向 BiLSTM 和 BIGRU,关举鹏等人使用 BiLSTM 模型提取文本特征,减少了人工制定特征带来的麻烦,还进一步提升了关系抽取效果<sup>[7]</sup>。双向循环神经网络可充分学习词语和上下文语义信息,为减少噪声带来的影响,可集成其他研究方法,发扬和弥补各自优缺点。

2015 年 Bahdanau 等人使用注意力机制进行机器翻译,让注意力机制在自然语言处理领域得到认可<sup>[16]</sup>。朱星嘉等人将注意力机制与 LSTM 相结合,重点处理对于关系分类影响力强的词语,减少重要信息的流失<sup>[17-20]</sup>。为进一步完善模型架构,在预训练阶段进行优化,Shi 等人使用 BERT 模型进行词嵌入,缓解了 Word2Vec 语言模型存在的一词多义问题,还加入位置向量、句子特征向量拼接成新向量作为模型输入,丰富了语义表征<sup>[21-24]</sup>。

健康医疗文本有内容上下文联系紧密、名词专业化等特点,相较于情感分析、人物关系等文本,更加难以处理。Ozlem 等人在医疗文本中定义了 6 大常见医疗实体关系,通过确定实体间关系类别来获取隐含信息<sup>[25]</sup>。Frunza 等人使用朴素贝叶斯和 SVM 模型对疾

病治疗、预防、副作用三者之间的语义关系进行分类,结合生物学文献和临床医疗知识,得到了更加准确的结果<sup>[26]</sup>。Sahu 等人首次使用 CNN 结合多种向量表征方式进行中医医学关系提取,减小了对专家定义特征质量的依赖<sup>[5]</sup>。Zhang 等人利用 CNN 和 RNN 优势,构成 CNN-RNN 模型对生物学文本中实体进行关系分类<sup>[6]</sup>。Chikka 等人使用 BiLSTM 模型来提取治疗和医疗问题之间的关系,结合基于规则的方法,可对小样本数据进行关系分类<sup>[27]</sup>。张志昌等人使用 BIGRU 模型结合注意力机制,增强特征识别的性能,减少误差带来的影响<sup>[8-9]</sup>。武小平等人将把 BERT 作为语言训练模型,相较于常见语言模型,新加入位置向量、句子特征向量,来丰富词性<sup>[28]</sup>。在文本处理领域,可采用神经网络模型作为基础架构,与注意力机制相结合,充分学习、利用文本信息,但准确率还有提升空间。

## 2 基于 BERT 与 BiLSTM 融合的关系提取方法

在医疗领域,存在大量非结构化文本数据,采用 CNN、RNN 进行处理,会出现文本利用率低,重要信息大量流失的情况。因此,提出了一种基于 BERT 和 BiLSTM-ATT 融合的关系提取模型(如图 1 所示),来充分学习和利用健康医疗文本中的重要信息。利用 BERT 语言模型生成词向量嵌入到 BiLSTM 层,不仅能让词语紧密联系上下文,还能缓解中文词语存在的一词多义问题;在 BiLSTM 层,获取正反两方向的隐含特征,让词语紧密联系上下文;再结合注意力机制考虑输入文本中部分词语重要性,合理分配权重;最后使用 Softmax 分类器得到关系概率。

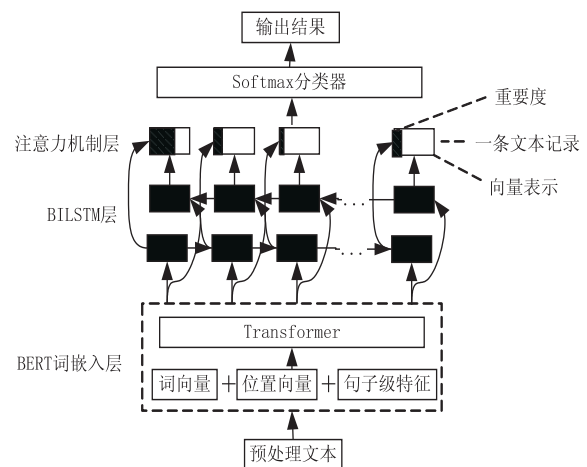


图1 基于 BERT 与 BiLSTM 融合的关系提取模型架构

### 2.1 BERT 词嵌入层

基于已有的 Word2Vec、GPT 等语言模型,2018 年

Google 团队提出 BERT 语言模型,由多层 Transformer 语言架构组合而成,凭借庞大语料库和超强算力,在文本处理任务中取得优异表现<sup>[29]</sup>。BERT 模型嵌入值由词向量、位置向量和句子特征向量组成,能保证文本中词语正确排序,并获得句子级表征能力,从而丰富向量表征信息,有利于后续任务的进行<sup>[16]</sup>。健康医疗文本的上下文联系较为紧密,实体需要结合上下文进行处理才能获取精确的信息,且中文词语可在不同句子中表示不同意思,存在一词多义的情况,使用 BERT 模型代替常用的 Word2Vec 模型进行词嵌入,能很好地缓解这些问题。

该文采用的是 BERT 中文预训练模型 (BERT<sub>BASE</sub>),Transformer 层数为 12,隐藏层维度为 768,多头注意力机制个数为 12,总参数为 110 M。以句子“患者出现胸闷、气短,诊断为高血压,口服速效救心丸后缓解”为例,展示 BERT 模型处理后的文本向量。

## 2.2 BILSTM 层

1997 年 Hochreiter 提出 RNN 变体—LSTM,由门控记忆单元组成,包括输入门、遗忘门、输出门。计算过程见式(1)~式(6):

$$f_t = \sigma(W_f h_{t-1} + U_f x_t + b_f) \quad (1)$$

$$i_t = \sigma(W_i h_{t-1} + U_i x_t + b_i) \quad (2)$$

$$a_t = \tanh(W_a h_{t-1} + U_a x_t + b_a) \quad (3)$$

$$c_t = c_{t-1} \cdot f_t + i_t \cdot a_t \quad (4)$$

$$o_t = \sigma(W_o h_{t-1} + U_o x_t + b_o) \quad (5)$$

$$h_t = o_t \cdot \tanh(c_t) \quad (6)$$

其中,  $h_{t-1}$  表示上一单元的隐藏状态,  $x_t$  表示本单元输入信息,  $f_t$  表示对上一单元隐藏状态的遗忘概率,决定保留多少过去状态信息;  $h_{t-1}$  和  $x_t$  使用不同激活函数构成本单元输入值,由上一单元部分记忆和本单元输入值共同构成本单元记忆  $c_t$ ,  $c_t$  和之前未处理的隐藏状态信息  $o_t$  共同构成本单元隐藏状态信息  $h_t$ ,决定向后传递多少信息。利用 LSTM 的链式结构选择记忆信息,缓解了 RNN 存在的长距离依赖问题,避免信息快速流失<sup>[30]</sup>。

BILSTM 由正向和反向 LSTM 组成,同时从两个方向开始训练,能有效获取过去和未来信息,从而缓解了单向 LSTM 只能序列化处理文本的情况。BILSTM 模型处理过程如图 2 所示,以句子“患者出现胸闷、气短,诊断为高血压,口服速效救心丸后缓解”为例,获取 LSTM 处理后的向量;  $\{h_{i1}, h_{i2}, h_{i3}, h_{i4}, h_{i5}, h_{i6}, h_{i7}, h_{i8}, h_{i9}\}$  和  $\{h_{j1}, h_{j2}, h_{j3}, h_{j4}, h_{j5}, h_{j6}, h_{j7}, h_{j8}, h_{j9}\}$  分别代表“患者”“出现”“胸闷”“气短”“诊断”“高血压”“口服”“速效救心丸”“缓解”在正向 LSTM 和反向 LSTM 中的向量,对正向和反向的隐藏向量进行叠加,得到向量  $\{h_1, h_2, h_3, h_4, h_5, h_6, h_7, h_8, h_9\}$ ,从而做到基于上下文进行处理,使 BILSTM 中每个字词包含整句的信息。由于健康医疗文本数据存在多模态、碎片化、记录时间不规律等特点,仅使用 LSTM 模型效果有限,故与其他方法结合进行集成分析,扩展其在医疗领域应用范围。

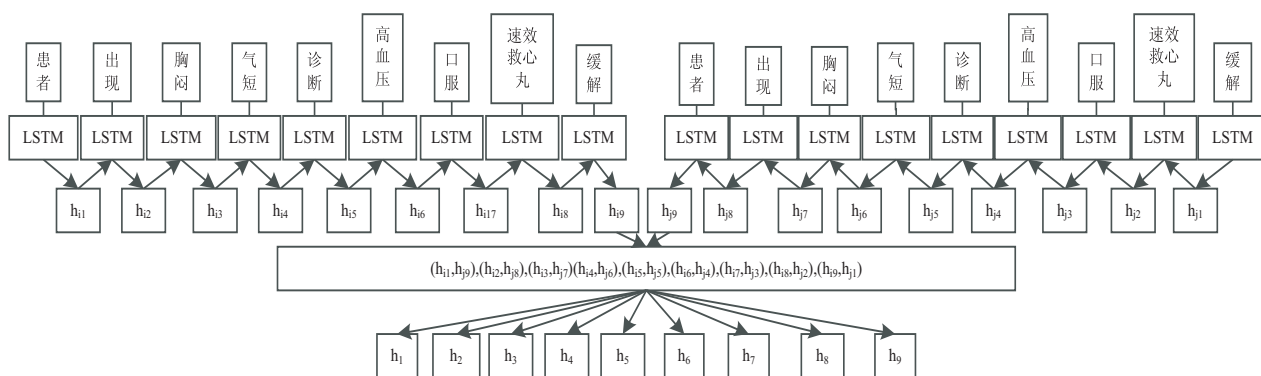


图 2 BILSTM 模型示例图

## 2.3 注意力机制层

在健康医疗文本中,每个词语对于实验目的贡献不相上下,若统一分配权重,会遗漏部分重要信息,降低准确率,故结合注意力机制对权重进行调整,快速筛选并重点处理高价值信息。计算过程见式(7)~式(9):

$$u_t = \tanh(W_u h_t + b_u) \quad (7)$$

$$a_t = \frac{\exp(u_t^T u_w)}{\sum_i \exp(u_i^T u_w)} \quad (8)$$

$$s_t = \sum_{i=1}^n a_i h_i \quad (9)$$

其中,  $u_t$  为随机初始化矩阵,以键值对的形式计算当前词语与句子中其他词语的相似度,得到占比权重  $a_t$ ,权重  $a_t$  和 BILSTM 层输出向量  $h_t$  乘积的累加和得到新的表示向量  $s_t$ <sup>[16]</sup>。

注意力机制模型见图 3,以句子“患者出现胸闷、气短,诊断为高血压,口服速效救心丸后缓解”为例,根据 BILSTM 层输出向量  $\{h_1, h_2, h_3, h_4, h_5, h_6, h_7, h_8,$

$h_9$ }, 得到对应权重  $\{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9\}$ , 乘积累加得到新的隐藏状态向量  $s_t$ , 最后使用 Softmax 函数进行分类, 得到预测标签。

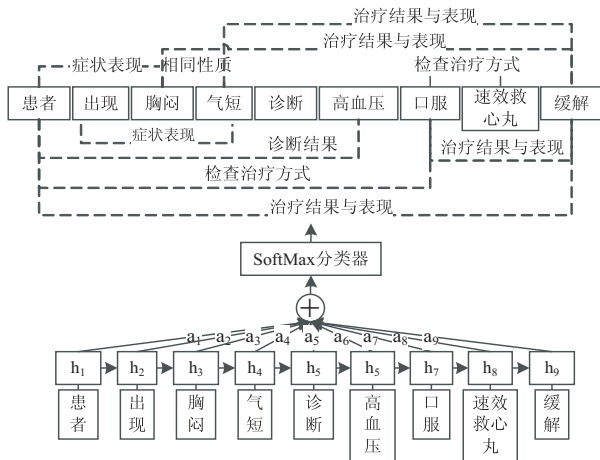


图3 注意力机制模型示例图

## 2.4 模型训练方法

BERT-BILSTM-ATT模型以健康医疗文本数据集、实验参数作为输入, 利用BERT模型把文本处理成向量形式, 再使用BILSTM结合注意力机制获取关系类别, 算法如下:

输入: 健康医疗文本数据集、实验参数;

输出: 健康医疗文本中实体间关系类别。

Step1: 使用BERT模型进行词嵌入, 词语对应向量为  $x_t$ 。

Step2: 使用BILSTM和注意力机制对数据集中每条健康医疗文本记录进行处理, 其中文本中单词数量为  $n$ 。

for hop=1 to  $n$ :

$$\vec{h}_t = \text{LSTM}(x_t, \vec{h}_{t-1})$$

$$\overleftarrow{h}_t = \text{LSTM}(x_t, \overleftarrow{h}_{t-1})$$

$$h_t = \vec{h}_t \oplus \overleftarrow{h}_t$$

$$u_t = \tanh(W_u h_t + b_u)$$

$$a_t = \frac{\exp(u_t^T u_w)}{\sum_i \exp(u_i^T u_w)}$$

$$s_t = \sum_{i=1}^n a_i h_i$$

end for

Step3: 根据最终的表示向量  $s_t$ , 使用Softmax函数计算各关系概率, 确定关系类别。

## 3 实验与结果分析

### 3.1 实验数据与预处理

健康医疗文本除了包含医学专业词语外, 还包含大量常用词语。若未提前对健康医疗文本中词语进行筛选, 直接使用jieba工具提取关键词, 会出现关键词中常用词语较多、医疗相关词语较少的情况, 妨碍获取文本中医疗实体。为减少高频无意义词语(如: 的、不能、月末等)对医疗信息获取的影响, 在预处理阶段先过滤掉部分无意义词语, 再进行医疗关键词提取, 降低医疗信息获取的困难程度, 使模型重点针对医疗相关实体进行处理。

#### 3.1.1 实验数据

通过查阅国内公开电子健康病历、咨询专家后, 将医疗相关实体分为12类(例如: 身体部位、疾病名称、药品名称等), 实体间关系分为10类(例如: 检查治疗方式、诊断结果等), 具体内容如图4所示。

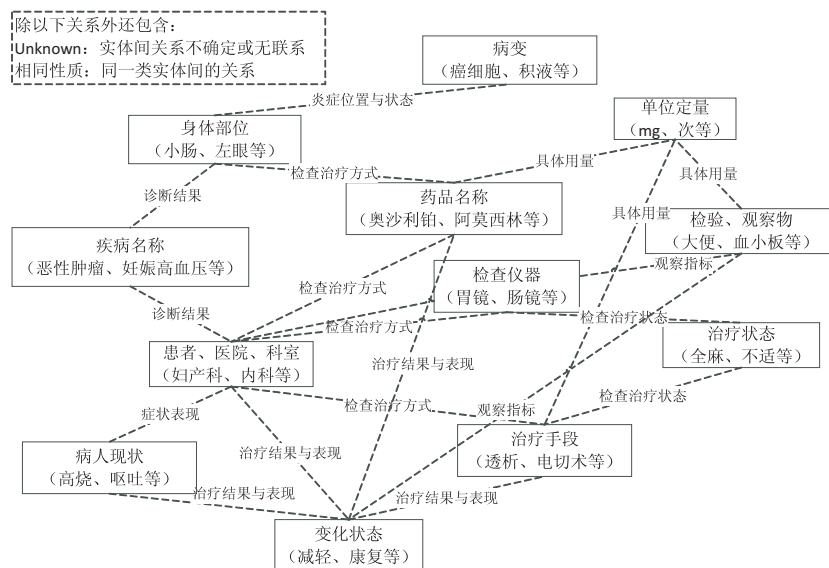


图4 医疗实体关系表

选取2个健康医疗文本数据集进行实验研究。Yidu-S4K数据集: 由“医渡云”医学数据智能平台根

据真实病历的分布情况人工编辑而成, 共包含8000条数据, 80%为肿瘤疾病文本, 每条记录包含现状、诊

断内容、治疗项目、恢复情况等;本地数据集:由 2017—2018 年间某市疾控中心的高血压患者数据组成,数据集已经过脱敏处理,包含记录 5 830 条,每条记录包含患者现状、诊断内容、治疗项目等。

### 3.1.2 对健康医疗文本的预处理过程

将寻医问药网(<https://www.xywy.com/>)词库作为文中的疾病词库,其中词库包含疾病、检查方式、科室、症状、治疗项目、身体部位等 12 类实体,共计 19 832 条数据,用于判断是否为医疗相关实体,有利于提取医疗关键词。

健康医疗文本处理步骤如图 5 所示,输入健康医疗文本,去除无用符号(如逗号、引号等),保留句号、分号作为句子隔断,再使用 jieba 工具进行分词处理。为减少无意义词语(如:年前、出现、不会等)对医疗领域文本处理的影响,以疾病词库为基础过滤掉部分词语,即存在于疾病词库则判断为关键词并保存,否则删除;使用 jieba 工具统计排名靠前的关键词作为研究实体,专家对实体间关系类别进行标注,最后将处理后的数据输入到文中模型中。

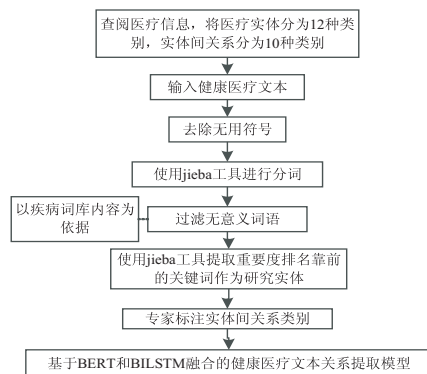


图 5 健康医疗文本预处理步骤

### 3.2 实验参数及评价指标

实验使用精确率 (precision)、召回率 (recall)、F1 值 (f1-score) 3 个标准作为评价指标,它们的计算公式如下:

$$\text{precision} = \frac{TP}{TP + FP} \quad (10)$$

$$\text{recall} = \frac{TP}{TP + FN} \quad (11)$$

$$\text{f1-score} = \frac{2 * p * r}{p + r} \quad (12)$$

表 1 数据集实验结果

Yidu-S4K 数据集					本地数据集				
relation	precision	recall	f1-score	support	relation	precision	recall	f1-score	support
unknown	0.781 0	0.764 3	0.772 6	280	unknown	0.934 6	0.877 2	0.905 0	228
诊断结果	0.869 8	0.873 8	0.871 8	214	诊断结果	0.920 0	0.970 0	0.950 0	171
检查治疗方式	0.927 3	0.927 3	0.927 3	275	检查治疗方式	0.925 9	0.961 5	0.943 4	87
相同性质	0.817 4	0.764 2	0.789 9	123	相同性质	0.934 3	0.914 3	0.924 2	140
治疗结果与表现	0.827 1	0.859 4	0.842 9	128	治疗结果与表现	0.948 3	0.887 1	0.916 7	62
检查治疗状态	1.000 0	0.927 5	0.962 4	69	检查治疗状态	0.925 9	0.961 5	0.943 4	26
症状表现	0.955 9	0.878 4	0.915 5	148	症状表现	0.938 8	0.945 7	0.942 2	276
炎症位置与状态	0.818 7	0.886 9	0.851 4	168	炎症位置与状态	0.911 1	0.820 0	0.863 2	50
具体用量	0.886 0	0.952 8	0.918 2	106	具体用量	1.000 0	0.944 4	0.971 4	36
观察指标	0.869 6	0.898 9	0.884 0	89	观察指标	0.843 1	0.955 6	0.895 8	90
macro avg	0.875 3	0.873 4	0.873 6	1 600	macro avg	0.930 9	0.925 4	0.927 1	1 166
weighted avg	0.866 0	0.865 0	0.864 0	1 600	weighted avg	0.928 4	0.927 1	0.926 9	1 166

### 3.3 实验结果分析

对 BERT-BiLSTM-ATT 模型进行迭代,得到各类实验结果和总体平均值,见表 1。由于 LSTM 模型参数量大,存在过拟合的风险,联合观察精确率和损失值动态变化来判断模型拟合状态,并做出参数调整;如图 6,迭代过程无异常波动,模型良好拟合。

各模型训练得到的结果如表 2 所示,其中 BERT-BiLSTM-ATT 模型效果最佳。对比未提取关键词的

BERT-BiLSTM-ATT 模型, BERT-BiLSTM-ATT 模型精确率提升 12.18% ~ 12.85%, 证明突出医疗词语有助于进行后续健康医疗文本关系提取实验。对比 Word2Vec-BiLSTM-ATT 模型, BERT-BiLSTM-ATT 模型精确率提升 3.36% ~ 3.76%, 证明在词嵌入过程中, BERT 语言模型优于 Word2Vec 语言模型。与 CNN 模型、BiLSTM-CNN 模型进行对比, 精确率大幅度提升, 说明本模型效果优于其他 CNN、RNN 神经网络模

型,能充分学习、利用健康医疗文本包含的信息,提高准确率;与 BERT-LSTM-ATT 模型进行对比,精确率提高 3.35% ~ 4.8%,说明双向 LSTM 模型结构优于单向结构,可获取更加准确的语义信息;与 BERT-

BIGRU-ATT 模型对比,精确率提高 4.07% ~ 4.45%,说明虽然 LSTM 结构比 GRU 结构复杂,但灵活性强,更加适合提取内容复杂的文本信息。总体来看,本模型对健康医疗文本关系提取效果较优。

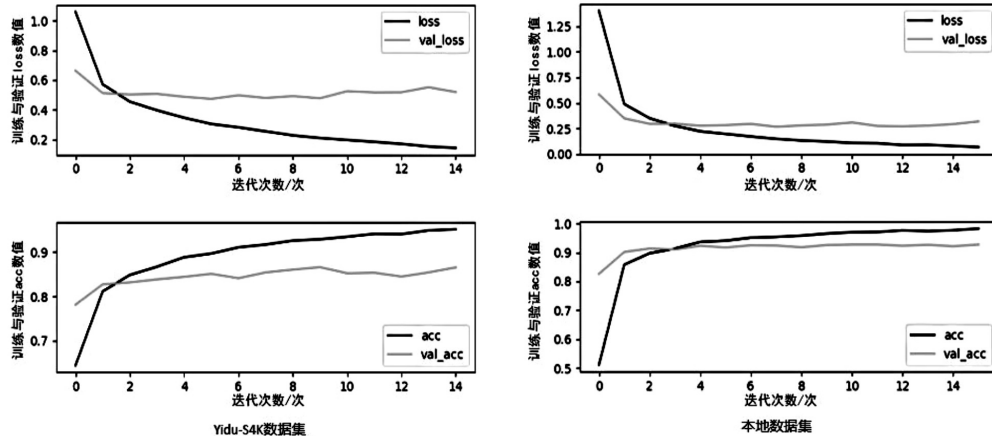


图6 loss 和 acc 变化过程

表2 各类模型实验结果

Methods	Yidu-S4K 数据集			本地数据集		
	precision	recall	f1-sorce	precision	recall	f1-sorce
CNN	0.555 4	0.393 6	0.460 7	0.347 2	0.285 8	0.312 5
BILSTM-CNN	0.307 1	0.356 2	0.329 8	0.264 5	0.321 2	0.290 1
BERT-BIGRU-ATT	0.841 1	0.862 4	0.851 6	0.891 2	0.895 3	0.893 2
BERT-LSTM-ATT	0.835 2	0.844 7	0.839 9	0.900 7	0.900 2	0.900 4
未提取关键词的 BERT-BILSTM-ATT	0.775 6	0.773 2	0.774 4	0.829 8	0.827 1	0.828 4
Word2Vec-BILSTM-ATT	0.843 6	0.842 2	0.842 9	0.900 6	0.891 0	0.895 8
BERT_BILSTM_ATT	0.875 3	0.873 4	0.873 6	0.930 9	0.925 4	0.927 1

## 4 结束语

针对健康医疗文本处理面临着利用不充分、信息获取难度大、准确率较低等问题,提出一种基于 BERT 和 BILSTM 融合的健康医疗文本关系提取方法,来对此类问题进行处理。与单向 LSTM、CNN、BIGRU 等模型进行对比,该模型分类性能表现最好,精确率提高 3.35% 以上,验证了基于 BERT 和 BILSTM 的融合模型能很好地获取实体间关系,满足医疗领域实际应用需要。但还存在一些问题,需进一步思考,如:不同病种数据对于健康医疗文本关系提取模型的影响;在健康医疗文本预处理阶段需专家进行标注,存在部分主观因素。

## 参考文献:

[1] 国务院办公厅. 国务院办公厅关于促进和规范健康医疗大数据应用发展的指导意见[EB/OL][2016-06-24]. [http://www.gov.cn/zhengce/zhengceku/2016-06/24/content\\_5085091.htm](http://www.gov.cn/zhengce/zhengceku/2016-06/24/content_5085091.htm).

[2] 胡佳慧,赵琬清,方安,等. 基于医疗大数据的临床文本处理与知识发现方法研究[J]. 中国数字医学,2020,15(7):11-13.

[3] KAMBHATLA N. Combining lexical, syntactic, and semantic features with maximum entropy models for information extraction[C]//Proceedings of the ACL interactive poster and demonstration sessions. Barcelona, Spain: Association for Computational Linguistics,2004:178-181.

[4] LIU C Y, SUN W B, CHAO W H, et al. Convolution neural network for relation extraction[C]//International conference on advanced data mining and applications. Hangzhou, China: Springer,2013:231-242.

[5] SAHU S K, ANAND A, ORUGANTY K, et al. Relation extraction from clinical texts using domain invariant convolutional neural network[C]//Proceedings of the 15th workshop on biomedical natural language processing. Berlin, Germany: Association for Computational Linguistics,2016:206-215.

[6] ZHANG Yijia, LIN Hongfei, YANG Zhihao, et al. A hybrid model based on neural networks for biomedical relation extrac-

- tion[J]. Journal of Biomedical Informatics, 2018, 81: 83–92.
- [7] 关鹏举, 曹春萍. 基于 BLSTM 的临床文本实体关系抽取[J]. 软件, 2019, 40(5): 159–162.
- [8] 张志昌, 周 炯, 张瑞芳, 等. 融合双向 GRU 与注意力机制的医疗实体关系识别[J]. 计算机工程, 2020, 46(6): 296–302.
- [9] ZHANG T, LIN H, TADESSE M M, et al. Chinese medical relation extraction based on multi-hop self-attention mechanism[J]. International Journal of Machine Learning and Cybernetics, 2021, 12: 355–363.
- [10] ZENG Daojian, LIU Kang, LAI Siwei, et al. Relation classification via convolutional deep neural network[C]//The 25th international conference on computational linguistics. Ireland, Dublin: Computer Science Bibliography, 2014: 2335–2344.
- [11] MIKOLOV T, CHEN K, CORRADO G, et al. Distributed representations of words and phrases and their compositionality[J]. Advances in Neural Information Processing Systems, 2013, 26: 3111–3119.
- [12] MIWA M, BANSAL M. End-to-end relation extraction using LSTMs on sequences and tree structures[C]//Proceedings of the 54th annual meeting of the association for computational linguistics. Berlin, Germany: Association for Computational Linguistics, 2016: 1105–1116.
- [13] 王海涛, 宋 文, 王 辉. 一种基于 LSTM 和 CNN 混合模型的文本分类方法[J]. 小型微型计算机系统, 2020, 41(6): 1163–1168.
- [14] 张玉环. 基于多种 LSTM 结构的文本情感分析[D]. 北京: 北京邮电大学, 2018.
- [15] 李 强, 李瑶坤, 夏书月, 等. 一种改进的医疗文本分类模型: LS-GRU[J]. 东北大学学报: 自然科学版, 2020, 41(7): 938–942.
- [16] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[C]//Proceedings of the 3rd international conference on learning representations. San Diego: [s. n.], 2015: 1125–1135.
- [17] 朱星嘉, 李红莲, 吕学强, 等. 一种改进的 Attention-Based LSTM 特征选择模型[J]. 北京信息科技大学学报: 自然科学版, 2018, 33(2): 54–59.
- [18] PENG Z, WEI S, TIAN J, et al. Attention-based bidirectional long short-term memory networks for relation classification[C]//Proceedings of the 54th annual meeting of the association for computational linguistics (volume 2: short papers). Berlin, Germany: Association for Computational Linguistics, 2016: 207–212.
- [19] 余本功, 许庆堂, 张培行. 基于 MAC-LSTM 的问题分类研究[J]. 计算机应用研究, 2020, 37(1): 40–43.
- [20] 周文辉, 刘亮亮, 张再跃. 融合多层注意力机制与双向 LSTM 的语义关系抽取[J]. 软件导刊, 2019, 18(7): 10–14.
- [21] SHI P, LIN J. Simple BERT models for relation extraction and semantic role labeling[J]. Computing Research Repository, 2019, 21: 114–122.
- [22] 王子牛, 姜 猛, 高建瓴, 等. 基于 BERT 的中文命名实体识别方法[J]. 计算机科学, 2019, 46(S2): 138–142.
- [23] 胡瑞雪, 曾 曦. 基于 BERT-LSTMCNN 的立场分析[J]. 信息技术, 2020, 44(2): 93–97.
- [24] IKA Novita Dewi, 蔡晓玲, 刘晓锋, 等. 结合类别关键词与注意力机制的药物相互关系抽取模型[J]. 华南理工大学学报: 自然科学版, 2021, 49(1): 10–17.
- [25] UZUNER O, MAILLOA J, RYAN R, et al. Semantic relations for problem-oriented medical records[J]. Artificial Intelligence in Medicine, 2010, 50: 63–73.
- [26] FRUNZA O, INKPEN D. Extraction of disease-treatment semantic relations from biomedical sentences[C]//Proceedings of the 2010 workshop on biomedical natural language processing. Uppsala, Sweden: Association for Computational Linguistics, 2010: 91–98.
- [27] CHIKKA R, KARLAPALEM K. A hybrid deep learning approach for medical relation extraction[C]//Conference on knowledge discovery and data mining, London, United Kingdom: Association for Computing Machinery, 2018: 755–761.
- [28] 武小平, 张 强, 赵 芳, 等. 基于 BERT 的心血管医疗指南实体关系抽取方法[J]. 计算机应用, 2021, 41(1): 145–149.
- [29] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers). Minneapolis, Minnesota: Association for Computational Linguistics, 2019: 4171–4186.
- [30] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735–1780.