基于BiLSTM-CRF的中医文本命名实体识别*

肖 瑞[1**]，胡冯菊[2]，裴 卫[1]

（1. 湖北中医药大学信息工程学院 武汉 430065；2. 湖北中医药大学第一临床学院 武汉 430065）

摘 要：中医药文本命名实体识别在中医药文本挖掘中占有重要地位，本文通过BiLSTM-CRF方法实现对中医医案文本进行命名实体识别，不仅实现了基本命名实体识别，通过对数据集按照中草药、疾病和症状三个类别进行标记，还能够进行命名实体类别识别。对中医药相关医案进行规整的10292条句子进行序列标注，基于word2vec的向量构建，从而进行模型训练迭代，得到了准确率为97.23%，召回率为89.47%，F值为88.34%的中医药命名实体识别模型。各类别识别中，中草药类别识别精准率为94.41%，召回率为94.36%，F值为94.38%；疾病类别精准率为80.92%，召回率为80.92%，F值为80.92%；症状类别精准率为75.68%，召回率为81.68%，F值为78.56%，人工测试模型效果较好，能够对医案数据进行实体识别。命名实体识别模型较多，但用于中医药相关命名实体识别模型数量微乎其微，构建中医药相关命名实体识别模型，将更加有效的推动中医药文本挖掘发展。

关键词：文本挖掘 中医药 命名实体 LSTM

doi: 10.11842/wst.20190513001 中图分类号：R285 文献标识码：A

1 概述

医案又称诊籍、病案、脉案、脉语，是医生临床诊治患者的记录，记载医案不仅是医生工作的一个重要环节，也是医学思想、理论水平、技术能力乃至医德医风的体现[1]。从中医继承和学习的角度而言，在浩瀚的中医典籍海洋中选取医案进行研究，不失为是一条行之有效的途径[2]。

命名实体识别[3]（Named Entity Recognition，NER）是自然语言处理(Natural Language Processing，NLP)工作中具有挑战性的任务之一，也是当前的研究热点。通过它可以准确地从文本中识别出人名、机构名、地名、时间、日期、货币、百分号等各类实体信息。鉴于中医药文本的特殊性，导致常规的通用命名实体识别模型对中医药文本识别并不友好，因此，以中医药经典医案为基础，以文本挖掘方法训练一套具有中医药特色，适应中医药行业的命名实体识别模型，将对中医药数据文本挖掘研究工作带来积极影响。

2 研究背景及研究现状

近年来，随着计算机信息技术的日益成熟，计算机辅助在各行业和学科领域都取得了飞快发展，在中医药方面的应用也是硕果累累。面对海量的中医药文本文献知识，自然人的工作精力和学习能力有限，因此学者们尝试通过自然语言处理辅助完成汇总中医知识的过程，将知识提炼出来，提取其中有用的诊疗信息[4]。有学者指出未来服务于现代中医个体诊疗体系的中医知识库，需要通过领域知识的交叉融合，借助先进的人工智能技术和知识管理理念，构建一个具备知识获取、应用、评价等知识管理功能的大型、通用、智能决策型知识库[5]。通过构建中医药相关命名实体识别模型，更是能在根源是解决知识库知识获取

收稿日期：2019-05-13

修回日期：2020-01-30

* 2017年湖北中医药大学“青苗计划”项目[No.2017ZZX016]：基于中医电子病历的慢性乙型肝炎诊断预测算法研究，负责人：肖瑞；国家中医药管理局2018年度中医药法制化建设项目[No.GZY-FJS-2018-162]：互联网虚假违法中医医疗广告监测，负责人：肖瑞。

** 通讯作者：肖瑞，讲师，主要研究方向：数据挖掘。

问题，从而完成大量文献文本知识获取难题，因此，构建高准确率实用型中医药相关命名实体模型成为中医药文本研究重点。

在中医药相关命名实体研究方面，冯丽芝[6]针对面向命名实体抽取的大规模中医临床病历库的构建问题，实现了结构化病历数据、条件随机场(Conditional RandomFields，CRFs)和Bootstrapping等三种自动化批量语料标注方法。尹迪[7]等人针对传统的序列化标注方法的不足，提出了一种新的基于联合模型的中文嵌套命名实体识别方法。刘凯[8]针对中医临床病历的命名实体，如症状、疾病和诱因等的抽取问题，通过手工标注的413份病历数据(以中文字为特征)与4类特征模版，将条件随机场(CRF)、隐马尔科夫模型(HMM)和最大熵马尔科夫模型(MEMM)用于中医病历命名实体抽取的实验，并进行比较分析。王世坤[9]针对明清古医案中症状、病机的自动识别标注问题，采用了基于条件随机场(CRF)的方法，提出数据清洗以及缩减合并词性以减少特征空间规模。袁玉虎[10]针对中医临床病历中的现病史部分展开症状术语抽取方法研究。原旎[11]通过应用深度表示的方法实现临床上的现病史数据的自动标识。

本文采用BiLSTM-CRF（Bi-directional Long Short Term Memory networks-conditional random field，基于条件随机场的双向长短时记忆网络）方法实现对中医医案文本进行命名实体识别，用于识别中医药文献或医案中的中草药名、疾病名称以及中医症状名称等，从而组成实体识别三元组，以提取中医药相关文献或医案中的信息。

3 资料与方法

3.1 数据来源

中医医学著作不仅是中华医学瑰宝，是前人医学经验的传承，也是后辈研究学习的重要来源。在中医医学各类著作中，中医药医案更是集中医药特色的理、法、方、药综合运用的一种具体反映形式，其中包含了大量的命名实体，而名老中医医案著作不仅规则，而且命名更加规范，便于进行文本挖掘。本文研究的数据主要来自于部分名老中医医案著作，包括《李培生老中医经验集》[12]、《增补评注柳选医案》[12]、《章次公医案》[14]、《姚贞白医案》[15]、《吴佩衡医案》[16]和《名老中医之路》[17]等，其中将《名老中医之路》作为最终测试集，用以验证模型优劣效果。研究过程中，将纳入训练集的各医案文本数据进行整合，以句号作为间隔符将原医案文本内容进行切分，在切分后的语句中剔除字数长度低于10的语句，得到了9409条语句作为训练集，对训练集中的所有语句按照中草药词汇、症状词汇和疾病词汇三种类别进行分类，分类完成后统计得知训练集语句中含有中草药词汇20682个，症状词汇9246个，疾病词汇2047个；用相同的方法将作为测试集的《名老中医之路》文献文本数据进行语句切分及剔除，得到了883条句子作为测试集，采用与训练集中语句相同的分类方法进行分类统计，得知测试集语句中含有中草药词汇1630个，症状词汇857个，疾病词汇152个。训练集与测试集的语句数据量比约为10.7∶1，词汇数据量比约为12.1∶1，略低于预期但可以开展基础研究工作。

3.2 数据预处理

3.2.1 词典构建

词汇是自然语言的基石，是语言更高层面自动分析的基础[18]。在命名实体识别研究过程中，词典数据的质量直接决定了分词效果的质量，从而影响到实体识别结果的质量。医疗行业内专业术语数量庞大，特别是中医药中专业相关术语众多，然而目前现有的大部分分词词库以通用词库居多，并不包含这些行业专业术语，以至于分词工作在专业术语中开展的效果并不理想。为了提高中医药专业词汇在分词工作中的准确性，需要建立中医药特色专业术语的相关词典，包含症状、疾病、证、中药、方剂等中医药行业特色专业术语内容。本文构建的中医药术语词典基于中华人民共和国国家标准GB/T15657-1995《中医病证分类与代码》[19]中包含的“证”、“疾病”类专业名词，中华人民共和国国家标准GB/T 31773-2015《中药方剂编码规则及编码》[20]包含的“方剂”类专业名词，另外，还包含有从国际疾病分类(ICD-10)[21]、《秦伯未医学名著全书》中摘录疾病与方剂专业名词。在此基础上，还收集了一些从医院临床一线医护人员反馈得到的常用专业术语以及中医药著作的相关术语信息作为补全内容，从而综合的构建症状、疾病、证、中药、方剂等专业名词词典。

3.2.2 分词

分词是自然语言处理的一条分支，是计算机在对中文进行处理过程中，将一条自然语言按照一定的标

准规范划分成若干个独立字或者词汇的过程。在英文的行文方式中，各个单词之间是以空格作为自然分界符的，而中文和大部分西方语言不同，行文中对于字、句和段能通过明显的分界符来基本划分，唯独对于词而言并没有一个形式上的分界符，句子均以字串的形式出现，因此对中文文本在处理时的首要任务就是进行分词工作，也就是将字串转变成词串的过程。虽然在此问题上，英文也同样存在短语的划分问题，不过在词这一级别，中文分词比之英文分词的难度和复杂程度都要高的多。对于在词的切分和属性研究，包括术语语义研究、字频、词频统计和字典编纂等方面，分词工作具有重要的语义[22]。本文利用python语言提供的第三方中文分词包，jieba分词，通过补充完善自定义的中医药专业名词的词典，完善后的词典中词汇量总计41110个，并以此为基础对各典籍的摘要数据进行分词处理。

3.2.3 序列标注

序列标注简单而言就是给定一串序列，对序列中存在的每个元素打上相应标记或标签，通过标签可以客观的对这一串序列进行深度分析。比如，某患者在某三级甲等中医院的一份医案主诉为“头痛畏寒，恶风，大便偏干，口苦”，我们希望在中医药特色诊疗的基础上，识别这份医案中所涉及到的中医症状，因此对这医案中的这句话的序列标注如表1所示，序号为1的是原始主诉句子，序号为2的为分词后主诉句子，序号为3的是进行序列标注后的主诉句子，“/s”用来表示句子进行分割的地方，这里标注采用BIO[23]（Begin, Intermediate, Other）的表示方法，其中“B”表示词语首字，“I”表示词语非首字，“O”表示非关注词汇或标点。

表1 序列标注

序号	数据
1	头痛畏寒，恶风，大便偏干，口苦
2	头痛/s 畏寒/s，/s 恶风/s，/s 大便偏干/s，/s 口苦
3	头[B]痛[I]畏[B]寒[I]，[O]恶[B]风[I]，[O]大[B]便[I]偏[I]干[I]，[O]口[B]苦[I]

本文设计的模型，主要目的在于识别中医药文献或医案中的中草药名、疾病名称以及中医症状名称等，由此三要素组成基本的实体识别三元组，用以提取中医药相关文献或医案中的信息。

第一步，对所涉及的训练和测试数据集的中医药文献进行命名实体标注，完成总体文本数据集的初步提取。

第二步，在初步提取的标注过程中，按照中草药、疾病、症状三种分类情况标注各命名实体所属类别（由于按照中草药、疾病、症状的三种分类而言分类结果相互独立，因此各命名实体所属类别相对唯一，不存在一个实体具有多个标注情况）。

第三步，对于训练和测试文本数据标注均采用BIO标注方法，在具体标注过程中，选用如表2所示的标记方式以便区分，其中中草药名称词首采用B-chm(Begin - Chinese herbal medicine)形式表示，词的其他部分均用I-chm(Intermediate - Chinese herbal medicine)形式表示，症状名称词首用B-sym(Begin - Symptom)形式表示，词的其他部分用I-sym(Intermediate - Symptom)形式表示，疾病名称词首用B-dis(Begin - disease)形式表示，词的其他部分用I-dis(Intermediate - disease)形式表示，非实体用O(Other)形式表示。

最后，对已经按要求完成相应标注的词，使用word2vec方法进行嵌入，生成300维字向量矩阵，供模型训练。

表2 CRF使用BIO标签集

实体标记	开始标记	中间和结束标记
中草药	B-chm	I-chm
症状	B-sym	I-sym
疾病	B-dis	I-dis
非实体标记	O	O

4 实验方法

4.1 BiLSTM-CRF

BiLSTM-CRF模型是在结合双向LSTM和CRF两个模型的基础上，综合两者的优点改进而来，其模型结构如图1所示。模型总体分为三部分，第一部分是输入层，本部分的任务是负责将词进行向量化映射，形成字向量矩阵，本文中是将序列标记后的数据通过word2vec方法生成300维字向量，用$\vec{B}$进行表示。第二部分是隐含层，也就是双向LSTM层，分别是前向LSTM神经网络层和后向LSTM神经网络层，在本部分中，将输入层的生成的字向量$\vec{B}$作为双向LSTM神经网络每个时间节点的初始输入值，在具体输入过程中，依次把输入向量序列的顺序序列和逆序序列分别

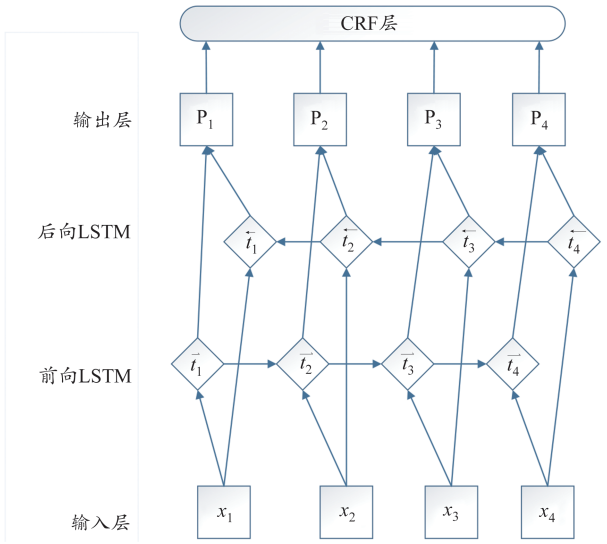

图1　BiLSTM-CRF算法模型结构图

作为前向LSTM层和后向LSTM的输入数据。假设在某一时刻i，模型将正向LSTM输出的隐状态序列$(\overrightarrow{t_1},\overrightarrow{t_2},...,\overrightarrow{t_i})$与反向LSTM输出的隐状态序列$(\overleftarrow{t_1},\overleftarrow{t_2},...,\overleftarrow{t_i})$进行位置拼接，得到$T_i=[\overrightarrow{t_i}\cdot\overleftarrow{t_i}]$，从而得到完整的隐状态序列$(t_1,t_2,...,t_i)$。第三部分是标注层，也就是输出层和CRF层。在BiLSTM网络的输出层中，为每一个输入的数据打一个标签的预测分值，也就是图中输出层的P_i，其中P_i是一个向量，表示代表句子$X(x_1,x_2,...,x_i)$的x_i的概率，对应BIO标记法所定义的标记$Tag(tag_1,tag_2,...,tag_j)$，j表示标记的维度，对于BiLSTM的输出矩阵P而言，其中$P_{i,j}$代表句子$X(x_1,x_2,...,x_i)$的x_i映射到tag_j的非归一化概率，即使没有CRF层，也可以训练一个BiLSTM命名实体识别模型，由于BiLSTM的输出为BIO标记中的每一个标签分值，可以挑选分值最高的一个作为该单元的标签，虽然可以得到句子X中每个单元的正确标签，但是不能保证标签每次都是预测正确的。在输出层后增加CRF层，也叫逻辑回归层，主要功能是进行语句的序列标注，加强文本间信息的相关性，可以为最后预测的标签添加一些约束来保证预测的标签是合法的，在训练数据训练过程中，这些约束可以通过CRF层自动学习到。CRF层中引入了转移矩阵M，而$M_{i,j}$表示从tag_i转移到tag_j的转移概率，从而利用此前标注过的信息对一个新的位置进行标注。例如，记预测标签序列为$Y(y_1,y_2,...,y_i)$，对于句子$X(x_1,x_2,...,x_i)$而言，模型预测标签的打分为$S(X,Y)=\sum_{n=1}^{i}(M_{Y_nY_{n+1}}+P_{n,Y_i})$，其中P的值是有LSTM的输出概率进行决定的，M的值取决于CRF的变化句子M，最终使用Softmax函数进行归一化。由于序列标注后，每个字都有一个标记，在对句子进行预测后将产生一个标注结果，通过测试集预测标记与序列标注的标记，运用评价标准中所述公式，即可计算模型的准确率等相关衡量指标。

在进行具体的命名实体识别过程中，对于一个句子$X(x_1,x_2,.....,x_i)$，识别流程为：首先对句子进行划分，然后对划分得到的字进行高维特征抽取，通过学习特征到标注结果的映射，可以得到特征到任意标签的概率，通过概率对比可以得到每一个字x_i所对应标签，从而得到句子$X(x_1,x_2,.....,x_i)$的预测标签序列$Y(y_1,y_2,.....,y_i)$，再按照BIO标注法对预测的标签序列$Y(y_1,y_2,.....,y_i)$进行顺序规整，从而达到实现命名实体识别的目的。

4.2　评价标准

在评价模型训练效果的优劣情况时一般重点考核识别准确度（ACC, accuracy）、精准率（P, precision）、召回率（R, Recall）和F-balanced等重要参数指标，各参数值越趋近100%，说明模型训练效果越好。为计算出各重要参数值，引入TP、FP、TN、FN等指标，如表3所示，其中TP表示预测为正样本，实际也为正样本的特征数。FP表示预测为正样本，实际为负样本的特征数。TN表示预测为负样本，实际也为负样本的特征数。FN表示预测为负样本，实际为正样本的特征。

表3　评价标准

真实情况	预测结果	
	正	负
正	TP	FN
负	FP	TN

其中准确率（ACC）计算公式为：$ACC=\frac{TP+TN}{TP+TN+FP+FN}$，是最常见的评价指标正确率越高，训练模型分类器效果越好。精确率（P）计算公式为：$P=\frac{TP}{TP+FP}$，表示的是预测为正的样本中有多少是真实正样本。召回率（R）计算公式为：$R=\frac{TP}{TP+FN}$，表示的是真实为正的样本中有多少被

预测为正样本。本文采用F-balanced计算公式为 $F = 2*\frac{R*P}{R+P}$。

5 实验结果

本次模型综合训练结果如表4所示，模型综合测评实验结果准确率达97.23%，而对于各个类别命名实体测试结果如表5所示，分类测试模型中中草药类别精准率最高，达94.41%，疾病类别精准率达80.92%，症状类别精准率最低达75.68%，影响症类别准确率的主要由于中草药训练数据较疾病和症状总量较大而症状类训练数据相对较少。

对训练好的模型进行测试，测试结果如表6所示，其中医案为某医生诊断病人后的医案点评数据，通过模型的预测结果为药物、症状和疾病三个类型，总体而言效果较好，能够用于大部分中医药相关文献数据命名实体识别。

表4 模型总评测结果

准确率	精准率	召回率	F-balanced
97.23%	87.25%	89.47%	88.34%

表5 各类别命名实体评测结果

命名实体类别	精准率	召回率	F-balanced
中草药	94.41%	94.36%	94.38%
疾病	80.92%	80.92%	80.92%
症状	75.68%	81.68%	78.56%

综合测试如图2所示，图2为本文中医药文本实体识别模型构建的系统，测试输入文本为网络医案文本，数据较杂，测试结果输出中草药，症状和疾病，能够识别大部分命名实体，对传统医案识别有较好的效果。

表6 单句测试

类别	数据
医案	本例胸痛系肝阳偏亢，气机不畅、瘀滞内停所致，症见胸闷时痛，心慌等症状，王老师认为，治疗应以疏肝、调畅气机为主，气机调畅则气血自能平和流畅，在处方中王老师用金铃子散加味，配柴胡、龙骨牡蛎、调畅气机、配合丹参、红花等活血，全方疏肝理气结合养心通脉，对其它各种气滞兼血瘀者也有较好疗效
药物	金铃子，柴胡，龙骨，牡蛎，丹参，红花
症状	胸痛，肝阳偏亢，气机不畅，瘀滞内停，胸闷，心慌，气机调，气滞，血瘀
疾病	无

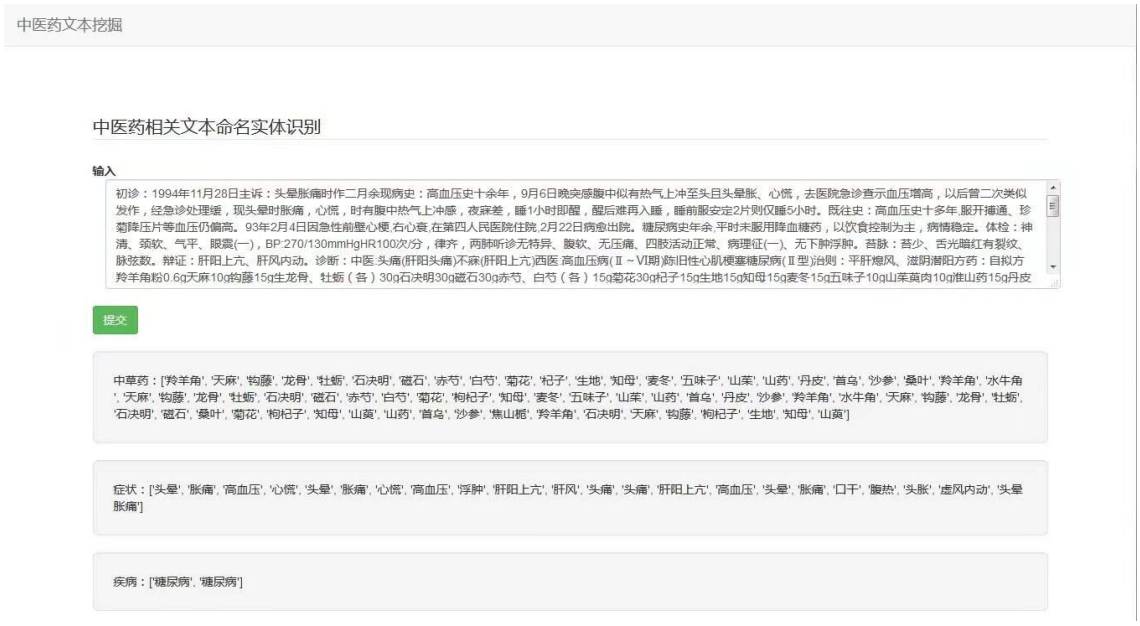

图2 中医药命名实体识别

6 结论与展望

本文采用BiLSTM-CRF方法对中药医案书籍人工标记，设计并训练命名实体识别模型，得到了准确率为97.23%，召回率为89.47%，F值为88.34%的中医药综合命名实体识别模型。各类别识别中，中草药类别识别精准率为94.41%，召回率为94.36%，F值为94.38%；疾病类别精准率为80.92%，召回率为80.92%，F值为80.92%；症状类别精准率为75.68%，召回率为81.68%，F值为78.56%，经人工测试后，模型实体识别效果良好，能够对所选择的医案数据进行较为准确的实体识别。但在单一类型实体识别中，症状类别和疾病类别的精准率、召回率和F值不高，主要是由于这两个类别的训练数据量不足，数据覆盖较小，从而对整体识别水平有一定影响。在模型实体识别类别上，由于数据来源辐射范围有限，仅划分了中草药、疾病和症状三种类别，并没有涵盖中医药相关文本实体的总类别，如针灸推拿中的腧穴、针灸术语等实体。在后期的研究中，将通过获取更多的中医药相关语料数据，采用更大的数据进行模型训练，从而提高各类别的精准率，对模型类别进一步划分，扩大可识别实体范围，提高可实用性和有效性。由于命名实体识别是自然语言处理的基础，并且中医药相关文献，例如医案，医集等数据众多，在临床中医方面，主要是靠医生通过手工整理处方，如何由相关人员进行数据挖掘，某类中医对某类疾病治疗的经验，由于人工整理效率较低，还存在一定的错误性，并且古医案较多，耗时耗力，基于中医药的命名实体识别，从某种程度上可以减少临床人员录入数据的负担，并且可以高效率的挖掘到更多的数据集，用于中医药临床疾病研究的初期跟着，从而促进中医药文本数据分析相关事业发展和中医临床经验水平的学习传播，在之后的工作中，将整合更加丰富的数据集训练更加精准的模型用于中医药事业发展。

参考文献

1 孟庆云. 宣明往范_昭示来学_论中医医案的价值_特点和研究方法. 中医杂志, 2006(08): 568-570.

2 何彦澄, 肖永华, 闫璞, 等. 中医医案分析方法评述. 中医杂志, 2018,59(13):1106-1109.

3 Chiu J P C, Nichols E. Named Entity Recognition with Bidirectional LSTM-CNNs. *Computer Science*, 2016.

4 柴华, 路海明, 刘清晨. 中医自然语言处理研究方法综述. 医学信息学杂志, 2015(10):58-63.

5 马利, 刘保延, 谢琪, 等. 我国中医知识库研究回顾与展望. 中国数字医学, 2014,9(01):11-14.

6 冯丽芝. 面向命名实体抽取的大规模中医临床病历语料库构建方法研究. 北京:北京交通大学硕士论文, 2015.

7 尹迪, 周俊生, 曲维光. 基于联合模型的中文嵌套命名实体识别. 南京师大学报(自然科学版), 2014,37(3).

8 刘凯, 周雪忠, 于剑, 等. 基于条件随机场的中医临床病历命名实体抽取. 计算机工程, 2014,40(9).

9 王世昆, 李绍滋, 陈彤生. 基于条件随机场的中医命名实体识别. 厦门大学学报(自然科学版), 2009,48(3).

10 袁玉虎, 周雪忠, 张润顺, 等. 面向中医临床现病史文本的命名实体抽取方法研究. 世界科学技术-中医药现代化,2017,19(01):70-77.

11 原旎, 卢克治, 袁玉虎, 等. 基于深度表示的中医病历症状表型命名实体抽取研究. 世界科学技术-中医药现代化, 2018, 20(03): 355-362.

12 李培生. 李培生老中医经验集.

13 尤在泾. 增补评注柳选医案.

14 朱良春. 章次公医案.2015.

15 姚贞白. 姚贞白医案. 2013.

16 吴佩衡. 吴佩衡医案. 2009.

17 张奇文, 柳少逸, 郑其国. 名老中医之路. 中国中医药出版社出版, 2010.

18 徐琳, 赵铁军. 国家自然科学基金在自然语言处理领域近年来资助的已结题项目综述. 软件学报, 2005(10):1853-1858.

19 国家中医药管理局. GB/T 15657-1995 中医病证分类与代码. 1995.

20 全国中药标准化技术委员会. GB/T 31773-2015 中药方剂编码规则及编码. 2015.

21 Sacconi P, Glover P, Marriot R, *et al*. International Classification of Diseases, Tenth Revision Implementation. *The Health Care Manager*, 2018,37(1):39-46.

22 刘耀, 段慧明, 王惠临, 等. 中医药古文献语料库设计与开发研究. 中文信息学报, 2008,22(4):24-30.

23 买买提阿依甫, 吾守尔·斯拉木, 帕丽旦·木合塔尔, 等. 基于BiLSTM_CNN_CRF模型的维吾尔文命名实体识别. 计算机工程, 2018,44(08):230-236.

Chinese Medicine Text Named Entity Recognition Based on BiLSTM-CRF

Xiao Rui[1], Hu Fengju[2], Pei Wei[1]

(1. Hubei University of Chinese Medicine , Wuhan ,430065, China ;2.Hubei University of Chinese Medicine , Wuhan, 430065, China)

Abstract: Text named entity recognition of Chinese medicine occupies an important position in text mining of traditional Chinese medicine, this article through the BiLSTM - CRF method was carried out on the basis of traditional Chinese medicine text named entity recognition, not only has realized the basic named entity recognition, based on the data set according to the Chinese herbal medicine, the three categories and symptoms, also can used to identify the named entity classes. Sequence annotation was performed on 10292 sentences of TCM related medical cases, and vector construction was conducted based on word2vec to carry out model training iteration. Thus, a TCM named entity recognition model with accuracy rate of 97.23%, recall rate of 89.47% and F value of 88.34% was obtained. Among all kinds of recognition, the accuracy rate of Chinese herbal medicine category identification is 94.41%, recall rate is 94.36% and F value is 94.38%. The precision rate of disease category was 80.92%, recall rate was 80.92%, and F value was 80.92%. The accuracy rate of the symptom category was 75.68%, the recall rate was 81.68%, and the F value was 78.56%. There are many named entity recognition models, but the number of them used for TCM related named entity recognition is very small. Therefore, the establishment of TCM related named entity recognition model will promote the development of TCM text mining more effectively.

Keywords: Text mining, TCM, Named entity, LSTM

（责任编辑：闫 群，责任译审：钱灵姝）