本文引用: 张伦伦, 任 高, 邹北骥, 刘青萍. 基于术语词典的中医医案实体抽取研究[J]. 湖南中医药大学学报, 2024, 44(6): 1110-1116.

基于术语词典的中医医案实体抽取研究

张伦伦[1],任 高[1],邹北骥[2],刘青萍[1]*

1.湖南中医药大学信息科学与工程学院,湖南 长沙 410208;2.中南大学计算机学院,湖南 长沙 410083

〔摘要〕**目的** 针对中医医案开展症状、病因病机、治法、用药、处方、取穴6类实体的抽取研究,为中医医案知识图谱构建和中医智能辅助诊疗提供基础。**方法** 根据中医医案文本的特点,提出一个可以动态更新的术语词典方法用于分词,并在中医脑系疾病医案和ChineseBLUE/cEHRNER、ChineseBLUE/cMedQANER、CBLUE/CMeEE 3个公开数据集上验证该方法的有效性。**结果** 使用术语词典的模型在准确率、精确率、召回率和*F*1值上均高于未使用术语词典的模型,在测试集和验证集上,*F*1值分别为92.07%和93.04%。**结论** 融合动态更新的术语词典分词方法的模型,能够增强中医领域特定术语和新实体的识别能力,提高中医医案关键信息识别的准确率,推进中医药知识的传承与发展。

〔关键词〕中医医案;脑系疾病;术语词典;实体抽取;IDCNN-CRF模型

〔中图分类号〕R2 〔文献标志码〕A 〔文章编号〕doi:10.3969/j.issn.1674-070X.2024.06.026

Research on entity extraction of TCM medical cases based on terminology dictionary

*ZHANG Lunlun[1], REN Gao[1], ZOU Beiji[2], LIU Qingping[1]**

1. School of Informatics, Hunan University of Chinese Medicine, Changsha, Hunan 410208, China;

2. School of Computer Science And Engineering, Central South University, Changsha, Hunan 410083, China

〔**Abstract**〕 **Objective** To extract six categories of entities including symptoms, etiology and pathogenesis, treatment principles, medication, prescriptions, and acupoint selection from TCM medical cases, so as to lay the foundation for the construction of TCM medical case knowledge graphs and intelligent assistance in TCM diagnosis and treatment. **Methods** Based on the characteristics of TCM medical case texts, a dynamically updatable terminology dictionary method was proposed for word segmentation, and its effectiveness was validated on medical cases of TCM neurological disorders, as well as three publicly available datasets: ChineseBLUE/cEHRNER, ChineseBLUE/cMedQANER, and CBLUE/CMeEE. **Results** The model using the terminology dictionary achieved higher accuracy, precision, recall, and *F*1 values compared to the model without using the terminology dictionary. The *F*1 values on the test set and validation set were 92.07% and 93.04%, respectively. **Conclusion** The model integrating the dynamically updatable terminology dictionary segmentation method can enhance the recognition ability of specific terms and new entities in the TCM field, improve the accuracy of key information identification in TCM medical cases, and promote the inheritance and development of TCM knowledge.

〔**Keywords**〕 medical cases of Chinese medicine; neurological disorders; terminology dictionary; entity extraction; IDCNN-CRF model

〔收稿日期〕2023-12-15

〔基金项目〕湖南省教育厅科学研究优秀青年项目(22B0385);2022年度学科建设"揭榜挂帅"项目(22JBZ051);湖南省中医药管理局智慧中医工程技术重点研究室。

〔通信作者〕* 刘青萍,女,博士,副教授,硕士研究生导师,E-mail:liuliu@hnucm.edu.cn。

中医医案是中医医家学术思想和临床经验的总结[1]，详细记录了诊疗过程中的辨证、立法和处方等内容[2]，主要包括患者的病史、症状、治法、处方、用药等信息，以及患者不同阶段的患病情况[3]。开展中医医案研究可以更好地探索不同医家的学术思想和诊疗经验，有助于中医药的传承与创新发展[4]。

脑血管疾病是指发生在脑部血管，因颅内血液循环障碍而造成脑组织损害的一组疾病，包括脑卒中、脑出血、脑梗死等[5]。本研究以中医脑系疾病医案文本为研究对象，识别症状、病因病机、治法、用药、处方、取穴 6 类实体[6]，作为知识图谱中的实体节点[7]，基于术语词典和深度学习模型开展中医脑系疾病医案实体抽取研究[8-10]，并在公开数据集上进行模型的对比研究，验证本研究所提方法的有效性，为中医医案知识图谱自动构建等研究打下坚实的基础，为中医临床诊疗提供新的思路与方法。

1　资料与方法

1.1　数据来源

本研究选取《名中医治愈脑血管病医案集》[11]为数据来源。该书包含邓铁涛、李聪甫、石学敏、丁筱兰等多位名医的医案。严格按照以下纳入和排除标准，筛选出 299 篇医案作为本研究的数据来源。

纳入标准：(1)以中医为切入点的医案；(2)符合中医脑系疾病诊断标准及症状表现；(3)诊疗过程、相关诊疗信息连续。

排除标准：(1)治疗方案中同时包含西医等其他疗法；(2)医案内容不完整，仅包含处方而缺少用药、用法用量等信息；(3)相关诊治信息过少；(4)医案内容不包含相关实体。

在上述筛选工作完成基础上，本研究将所选医案按 7∶1∶2 分别划分为训练集、验证集和测试集，并通过设置随机种子确保每次获得的训练集、验证集和测试集内容一致[12]。

1.2　数据处理

数据获取后，本研究从序列标注、分词、构建术语词典、构建字表和标签表等详细描述数据处理工作。

1.2.1　序列标注　序列标注目的是标记出原始医案文本中各个实体所属的类型，标记内容主要包含实体、实体类型、实体开始位置、实体结束位置等信息。本研究使用“精灵标注助手 2.0.4”作为标注工具，图 1 为多名中医专业研究生采用 BIO 标注方法对医案进行标注和相互校对后的结果[13]。

在中医理论指导下，本研究根据中医脑系疾病医案数据的特点，最终标注了 6 类实体，分别为：症状(以下简称“ZZ”)、病因病机(以下简称“BYBJ”)、治法(以下简称“ZF”)、用药(以下简称“YY”)、处方(以下简称“CF”)、取穴(以下简称“QX”)。采用 BIO 标注方法，将实体的开始位置标记为“B-X”，中间和结束位置标记为“I-X”，其他非实体部分，统一标记为“O”[14]。实体类型标注信息如表 1。

图 1　中医脑系疾病医案标注示例

表 1　实体类型标注信息

实体类别	标记符号	开始标记	中间/结束标记	举例
症状	ZZ	B-ZZ	I-ZZ	眩晕耳鸣
病因病机	BYBJ	B-BYBJ	I-BYBJ	肝肾阴虚
治法	ZF	B-ZF	I-ZF	化痰通络
用药	YY	B-YY	I-YY	黄芪、党参
处方	CF	B-CF	I-CF	天麻钩藤饮
取穴	QX	B-QX	I-QX	风池、百会
其他	O	O	O	不属于以上任何实体

以中医脑系疾病医案中“右侧半身不遂，语言不利，神志不清”为例，标注结果示例如表 2 所示。标注完成后本研究进一步进行数据预处理工作，将标注后的字和标签一一对应，生成新的数据，表 3 为字和标签的对应关系示例。BIOES 标注方法通过引入“E”和“S”标签，为实体标注提供了更丰富的信息[15]。同时，“S”标签能够标注单个字符构成的实体，极大程度提高了实体抽取的准确性和效率[16]。因此，为了使模型能够更准确地捕捉实体位置，本研究在数据预处理阶段将 BIO 标注转化为 BIOES 标注。

表 2　标注结果的示例

序号	实体类别	开始位置	结束位置	实体信息
T1	ZZ	9	15	右侧半身不遂
T2	ZZ	16	20	语言不利
T3	ZZ	21	25	神志不清

表 3 字和标签对应关系示例

字	标签	字	标签	字	标签
王	O	。	O	言	I-ZZ
×	O	右	B-ZZ	不	I-ZZ
×	O	侧	I-ZZ	利	I-ZZ
，	O	半	I-ZZ	，	O
男	O	身	I-ZZ	神	B-ZZ
，	O	不	I-ZZ	志	I-ZZ
5	O	遂	I-ZZ	不	I-ZZ
5	O	，	O	清	I-ZZ
岁	O	语	B-ZZ	。	O

1.2.2 分词 分词又称为词语切分，是将连续的文本切分为一个个独立的词汇单元的过程。对于中医医案等专业性强、结构复杂且没有明显词汇边界的文本而言，分词是一个重要的文本预处理步骤，对于确定词的边界和整体上提高实体抽取的准确性具有重要作用[17]。

本研究分别采用 Jieba 默认的精确模式和所构建的术语词典方式对相关中医医案文本进行分词，并将两种方式的分词结果进行对比与分析。

1.2.3 构建术语词典 结合中医专业的特点、行业标准以及待标注术语情况，本研究构建了包含半身不遂、口舌歪斜、气虚血瘀、息风化痰、活血通络、半夏白术天麻汤、天麻、钩藤等 11 611 个词语在内的术语词典。术语词典通过以下 6 种方式获取术语：(1)从标注后的训练集中提取 3 976 个术语；(2)从《方剂学》[18]教材中提取 422 个术语；(3)从《中药学》[19]教材中提取 559 个术语；(4)从《中医临床诊疗术语 第 1 部分：疾病》[20](GB/T 16751.1—2023)中提取 2 069 个术语；(5)从《中医临床诊疗术语 第 2 部分：证候》[21](GB/T 16751.2—2021)中提取 3 211 个术语；(6)从《中医临床诊疗术语 第 3 部分：治法》[22](GB/T 16751.3—2023)中提取 1 809 个术语。

1.2.4 构建字表和标签表 为了更好地表示字和标签以及 id 值之间的映射关系，本研究构建了字表和标签表。首先统计每个字和每个标签出现的频率，对统计结果从高到低排序，赋予每个字和每个标签唯一、不重复的 id 值[23]，并在构建术语词典时添加“PAD”和“UNK”标识。其中，“PAD”标识是在同一批次中某个句子与最长句子不等长时进行填充；“UNK”标识表示未知字符，可以将测试集、验证集中未在词典中出现的字符，统一表示为“UNK”的 id。

1.2.5 数据处理流程 本研究的数据处理流程分为 5 个步骤，具体如下：(1)对中医医案进行标注，并将标注后的实体标签信息和文字对应，得到“字和标签”的对应关系；(2)将标注后的中医医案随机打乱顺序，按照 7:1:2 的方式划分为训练集、验证集和测试集；(3)检验文本的 BIO 编码是否正确，并将BIO 编码转化为 BIOES 编码；(4)通过统计字和标签，创建字映射和标签映射，赋予字和标签唯一且不重复的数值，并将数据集中的每句话均转化为字列表、字典映射列表、标签映射列表、分词列表 4 个集合；(5)对转化后的集合采取统一的批次管理，以每批次中最长的句子为基准，采用填充的方式使其他句子的维度和基准维度保持一致。

1.3 实验模型

1.3.1 迭代膨胀卷积神经网络 迭代膨胀卷积神经网络(iterated dilated convolutions neural network，IDCNN)是一种改进的卷积神经网络，可以解决通用卷积神经网络中池化层信息丢失的问题[24]。与传统池化层相比，IDCNN 采用膨胀卷积，通过增加空洞，扩大感受野，使每个卷积输出时可以包含更大范围的信息，并在一定程度上避免了池化层带来的信息损失。同时，IDCNN 包含多个膨胀卷积块，每个块是一个多层的膨胀卷积神经网络，通过引入迭代宽度实现对输入数据的跳跃式处理，进一步增强网络的信息获取能力[25]。因此，IDCNN 特别适用于处理自然语言处理任务中句子的长依赖关系。

1.3.2 条件随机场 条件随机场(conditional random field，CRF)是解决标注问题的经典算法，尤其是序列标注中的链式 CRF[26]。通常，序列标注任务中的标签之间存在前后依赖关系，这与输入序列前后关系有关。以标签“Yi”为例，其与前一个标签“Yi-1”、后一个标签“Yi+1”以及输入序列“X”都有关联。CRF 是一种判别式模型，通过学习输入序列和输出序列之间的关系，达到捕捉句子中标签的前后依赖性的目的。此外，CRF 在实际应用中可以接收 IDCNN 的输出结果，确保最终预测结果的准确性。

1.3.3 模型结构 图 2 为模型结构图，本研究以中医医案中“右侧半身不遂”为例，详细介绍模型实体抽取的过程，具体如下。

(1)数据预处理阶段。将“右侧半身不遂”一句转化为字列表、word_to_id 映射列表、label_to_id 映射列表和分词列表 4 个集合。其中，字列表包含句子中的每个字，word_to_id 映射列表表示每个字对应的 id 值，label_to_id 映射列表表示字对应的标签 id 值。分词列表通过加载术语词典，对句子进行分词，并按规则将切分后的词转化为数值信息，为模型的输入

增加实体特征信息。将同一批次中未达到基准长度的文本通过 padding 填充后，进入 embedding 层。

(2)IDCNN 层。将第一步的结果作为 IDCNN 层的输入信息，通过迭代宽度获取更长上下文信息，映射到 k 维，其中 k 表示数据集中标注的标签数。通过特征提取，得到每个字在各个标签中的值，选择最大值作为该字的标签。

(3)CRF 层。已抽取到的实体在经过 IDCNN 层后，可能存在不规范的标签特征。CRF 层通过在训练过程中学习正确的约束规则，对 IDCNN 层输出的标签进行约束和纠正，确保最终预测结果的正确性。

2 实验

2.1 评价指标

本研究选取准确率、精确率、召回率和 *F*1 值作为模型性能的评价指标。其中 *TP* 表示模型预测为正类的正样本，*TN* 表示模型预测为负类的负样本，*FP* 表示模型预测为正类的负样本，*FN* 表示模型预测为负类的正样本。

准确率：预测正确的标签和总标签的比例，其计算公式为：

$$Accuracy=\frac{TP+TN}{TP+FP+TN+FN} \tag{1}$$

精确率：预测结果为正例样本中真实结果为正例的比例，其计算公式为：

$$Precision=\frac{TP}{TP+FP} \tag{2}$$

召回率：真实结果为正例的样本中预测结果为正例的比例，其计算公式为：

$$Recall=\frac{TP}{TP+FN} \tag{3}$$

*F*1 值：反映模型的稳健性，可以同时兼顾到精确率和召回率，*F*1 值越大则表示模型的性能越好，其计算公式为：

$$F1=\frac{2TP}{2TP+FN+FP}=\frac{2\times Precision\times Recall}{Precision+Recall} \tag{4}$$

2.2 参数设置

本研究配置的环境：Python 3.6.5、TensorFlow 1.13.1。

本研究的实验参数：分词维度为 20，WordEmbed ding 维度为 100，梯度裁剪为 5，dropout 为 0.5，batch_size 为 120，学习率为 0.001，优化器选择 Adam，最大轮训次数设置为 100。

2.3 整体结果分析

实验 1：采用 IDCNN-CRF-Unloaded、Bert-BiLSTM-CRF、IDCNN-CRF-Loaded 模型在中医脑系疾

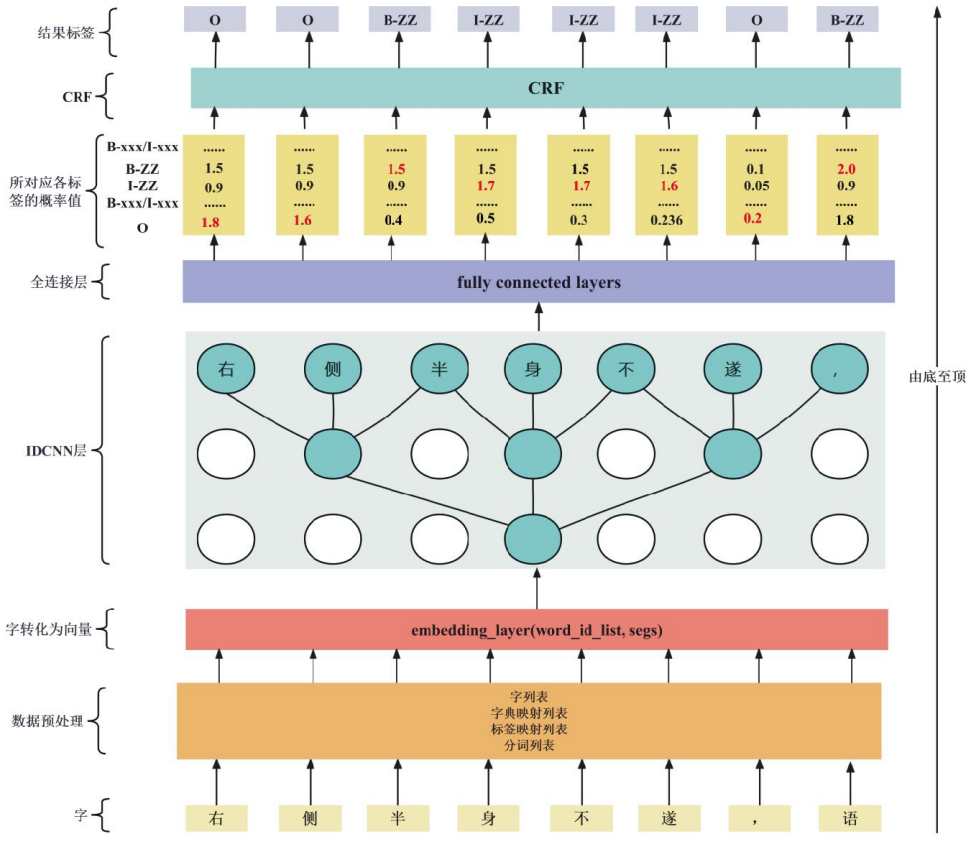

图 2　模型结构图

病医案训练集上进行训练，并在测试集和验证集上测验。如表 4，使用术语词典的模型（IDCNN-CRF-Loaded）的各项指标值均超过 90%，其中测试集、验证集 *F*1 值分别达到了 92.07%、93.04%。IDCNN-CRF-Loaded 模型表现最好，其次是 Bert-BiLSTM-CRF 模型，而 IDCNN-CRF-Unloaded 模型性能相对较差。使用术语词典的模型，无论在测试集还是验证集上，4 种评价指标均超过未使用术语字典的模型（IDCNN-CRF-Unloaded），其中测试集、验证集分别高出 5.79%、5.47%。

实验 2：采用 IDCNN-CRF-Unloaded、Bert-BiLSTM-CRF、IDCNN-CRF-Loaded 模型，在公开数据集 ChineseBLUE/cEHRNER 上训练。从表 4 分析结果可知，使用术语词典的 IDCNN-CRF-Loaded 模型的 4 种评价指标都超过了未使用术语字典的模型（IDCNN-CRF-Unloaded），且在测试集和验证集上，*F*1值分别达到了 89.60%、84.13%。

实验 3：采用 IDCNN-CRF-Unloaded、Bert-BiLSTM-CRF、IDCNN-CRF-Loaded 模型，在公开数据集 ChineseBLUE/cMedQANER 上训练。从表 4 分析结果可知，使用术语词典的模型（IDCNN-CRF-Loaded）的 4 种评价指标都超过了未使用术语字典的 IDCNN-CRF Unloaded 模型，且在测试集和验证集上，*F*1 值分别达到了 82.88%、83.53%。

实验 4：采用 IDCNN-CRF-Unloaded、Bert-BiLSTM-CRF、IDCNN-CRF-Loaded 模型，在公开数据集 CBLUE/CMeEE-V2 上训练。从表 4 分析结果可知，使用术语词典的模型（IDCNN-CRF-Loaded）的 4 种评价指标都超过了未使用术语字典的 IDCNN-CRF-Unloaded 模型，且在测试集和验证集上，*F*1 值分别达到了 78.36%、78.48%。

由上述 4 个实验可知，实验 2、实验 3 和实验 4 的分析结果与实验 1 相比稍差，其原因如下：（1）实验 2、实验 3 和实验 4 所用的公开数据集标注的质量较差，很多实体的词语间存在标点符号，在一定程度上会影响关键信息的识别；（2）训练集、测试集和验证集存在划分不平衡的情况，实验 2 中表现比较明显。通过模型性能比较可知：IDCNN-CRF-Loaded 模型在大多数情况下表现最好，其次是 Bert-BiLSTM-CRF 模型，而 IDCNN-CRF-Unloaded 模型性能相对较差。上述结果表明，在中医医案实体抽取任务中，IDCNN-CRF-Loaded 模型是最佳选择。

2.4 实体类别结果分析

本研究基于 IDCNN-CRF-Loaded 模型，分析中医治疗脑系疾病医案中 6 种实体的训练结果。IDCNN-CRF-Loaded 模型在各类实体之间的识别结果如表 5—6 所示。其中，表 5 为测试集之间是否使用术语词典结果的对比，表 6 为验证集之间是否使用术语词典结果的对比。在 6 类实体结果上分析，使用术语词典的模型的 *F*1 值高于未使用术语词典的模型。

2.5 中医医案在线实体抽取系统

本研究采用 Flask 框架，结合中医领域术语词典和在中医脑系疾病医案数据集上训练得到的 IDCNN-CRF-Loaded 模型，设计与实现了中医医案在线实体抽取系统。系统界面分为输入中医医案的左侧部分、显示实体类别的中间部分以及输出识别结果的右侧部分（如图 3）。通过该系统，用户可以准确地识别中医医案中的症状、病因病机、治法、用药、处方、取穴 6 类实体，快捷地对术语词典中识别出的新实体进行核验。对于术语词典中已包含的实体，

表 4 模型实验结果对比

数据集	模型	准确率/%		精确率/%		召回率/%		*F*1 值/%	
		测试集	验证集	测试集	验证集	测试集	验证集	测试集	验证集
中医脑系疾病医案	IDCNN-CRF-Unloaded	95.29	95.11	85.95	86.57	86.61	88.58	86.28	87.57
	Bert-BiLSTM-CRF	95.72	95.78	87.45	88.26	88.63	91.52	88.04	89.86
	IDCNN-CRF-Loaded	97.69	97.56	90.80	92.30	93.83	93.80	92.07	93.04
ChineseBLUE/cEHRNER	IDCNN-CRF-Unloaded	97.30	96.04	86.10	77.46	86.23	78.91	86.17	78.17
	Bert-BiLSTM-CRF	97.30	95.56	83.33	76.20	87.04	76.75	85.15	76.47
	IDCNN-CRF-Loaded	97.97	97.13	87.96	83.36	91.30	84.91	89.60	84.13
ChineseBLUE/cMedQANER	IDCNN-CRF-Unloaded	94.57	94.36	78.56	78.72	79.03	78.03	78.79	78.37
	Bert-BiLSTM-CRF	94.63	93.73	77.71	78.46	81.10	74.43	79.37	76.39
	IDCNN-CRF-Loaded	96.12	95.86	82.85	82.15	82.91	84.96	82.88	83.53
CBLUE/CMeEE	IDCNN-CRF-Unloaded	82.32	82.17	66.64	66.27	64.96	65.92	65.79	66.09
	Bert-BiLSTM-CRF	84.74	84.41	70.88	70.97	72.47	72.13	71.67	71.55
	IDCNN-CRF-Loaded	88.99	88.64	78.77	78.84	77.96	78.13	78.36	78.48

表 5　测试集之间的对比

实体类型/测试集	准确率/%		精确率/%		召回率/%		*F*1 值/%	
	未加载	加载	未加载	加载	未加载	加载	未加载	加载
症状	81.63	95.00	77.73	87.36	77.39	92.17	77.56	89.70
病因病机	79.10	87.87	74.44	89.88	74.03	83.43	74.24	86.53
治法	84.85	92.42	75.61	78.20	76.86	85.95	76.23	81.89
用药	97.84	98.78	95.20	95.27	97.00	98.44	96.09	96.83
处方	89.10	95.02	84.93	91.55	84.93	89.04	84.93	90.28
取穴	90.53	90.95	88.50	91.40	89.29	90.18	88.89	90.79

表 6　验证集之间的对比

实体类型/验证集	准确率/%		精确率/%		召回率/%		*F*1 值/%	
	未加载	加载	未加载	加载	未加载	加载	未加载	加载
症状	82.52	95.12	75.00	84.32	77.62	90.70	76.29	87.39
病因病机	87.03	95.86	73.76	88.89	81.89	94.49	77.61	91.60
治法	88.24	91.88	69.23	79.59	73.47	79.59	71.29	79.59
用药	97.32	97.38	96.74	98.00	96.61	97.03	96.67	97.51
处方	86.33	88.49	76.32	93.75	82.86	85.71	79.45	89.55
取穴	95.34	97.46	95.33	98.13	94.44	97.22	94.88	97.67

系统会自动提示;对于新识别出的实体,系统会自动将其标记为红色,待进一步核验或修改识别结果。用户确认无误后,可以直接获得识别结果。如需修改识别错误的实体,点击确定后,系统会自动将修正后的正确实体添加到术语词典中。

图 3　中医医案在线实体抽取系统

3　讨论

本研究使用 3 种深度学习模型,提出了加载术语词典的实体抽取方法。对比不同模型在 4 个数据集上的表现,观察到在 4 个实验中,使用术语词典的模型 IDCNN-CRF-Loaded 在测试集和验证集上的性能普遍优于未使用术语词典的模型 IDCNN-CRF-Unloaded。在中医脑系疾病医案训练集上,使用术语词典的模型在测试集和验证集上的 *F*1 值分别达到了 92.07%和 93.04%,相比未使用术语词典的模型分别高出了 5.79%和 5.47%。在其他 3 种公开数据集上,采用术语词典的模型同样取得了更高的评价结果。通过比较不同模型之间的性能差异,本研究发现,在中医脑系疾病医案数据集和其他公开数据集上,IDCNN-CRF-Loaded 模型表现最好,其次是 Bert-BiLSTM-CRF 模型,而 IDCNN-CRF-Unloaded 模型性能相对较差。上述结果表明,引入术语词典在垂直领域开展实体抽取研究,能有效提升模型对术语的识别和分类能力。

采用 IDCNN-CRF-Loaded 模型对中医脑系疾病医案数据集中的 6 类实体进行识别和分析,在识别不同类别的实体时,对病因病机和症状两类实体而言,加载术语词典后,模型的识别效果明显提升,这主要因为病因病机和症状的表达方式与常规词语有较大差异。例如,在中医胸痹医案中,病因病机常以饮食不节、情志失调、禀赋异常等四字短语形式出现;症状常用肌肤不仁、舌强语謇、四肢厥冷等四字术语。加载术语词典可以帮助模型更好地识别和理解上述特定术语,从而提高对病因病机和症状这两类实体识别的准确性。针对用药和取穴两类实体而言,加载术语词典对模型的性能提升影响较小,主要因为用药和取穴涉及的词语与常规词汇更为接近,药物或穴位名称如麻黄、桂枝、太子参或百会、风池、足三里等,较为常见且易于被模型学习和识别。

本研究可为后续关系抽取和知识图谱构建奠定基础,但也存在一些不足:本研究主要围绕中医脑系疾病展开,未涉及其他疾病,今后将考虑拓展研究范

围,结合其他疾病深入探索,进一步丰富样本类型和数量;同时,当前研究是以中医学为切入点进行,未涉及藏族医学、蒙古族医学、维吾尔族医学等知识,今后将尝试融入其他医学宝贵知识,旨在提升模型的普适性和应用价值,推动中医药知识的传承与发展,为中医临床诊疗提供新路径与方法。

参考文献

[1] 王桂彬, 庞 博. 名老中医隐性知识发现与医案解构模式研究[J]. 中华中医药杂志, 2023, 38(5): 2230-2234.

[2] 郝瑞森. 基于多元融合方法的刘燕池教授道术传承研究[D]. 北京: 北京中医药大学, 2022.

[3] 林 玲, 沈绍武, 付文娇, 等. 中医病案知识结构及其要素演变比较研究[J]. 时珍国医国药, 2023, 34(10): 2554-2557.

[4] 李 纲, 潘荣清, 毛 进, 等. 整合 BiLSTM-CRF 网络和词典资源的中文电子病历实体识别[J]. 现代情报, 2020, 40(4): 3-12, 58.

[5] 木其尔, 詹 青, 陈 伟. 重症脑病的中医药治疗进展[J]. 世界科学技术: 中医药现代化, 2020, 22(11): 4025-4032.

[6] 孔静静, 于 琦, 李敬华, 等. 实体抽取综述及其在中医药领域的应用[J]. 世界科学技术: 中医药现代化, 2022, 24(8): 2957-2963.

[7] 屈丹丹, 杨 涛, 朱 垚, 等. 基于字向量的 BiGRU-CRF 肺癌医案四诊信息实体抽取研究[J]. 世界科学技术: 中医药现代化, 2021, 23(9): 3118-3125.

[8] 杨延云, 杜建强, 聂 斌, 等. 一种面向中医文本的实体关系深度学习联合抽取方法[J]. 计算机应用与软件, 2023, 40(3): 217-222, 234.

[9] 杨锦锋, 关 毅, 何 彬, 等. 中文电子病历命名实体和实体关系语料库构建[J]. 软件学报, 2016, 27(11): 2725-2746.

[10] 胡蕴慧, 刘 朋, 熊皓舒, 等. 数智中药: 现代中药数智化升级与创新发展[J]. 中草药, 2024, 55(1): 1-11.

[11] 钟起哲. 名中医治愈脑血管病医案集[M]. 北京: 中国医药科技出版社, 1992.

[12] 蒋 翔, 马建霞, 袁 慧. 基于 BiLSTM-IDCNN-CRF 模型的生态治理技术领域命名实体识别[J]. 计算机应用与软件, 2021, 38(3): 134-141.

[13] 高佳奕, 刘 震, 杨 涛, 等. 基于条件随机场的中医临床医案症状命名实体抽取研究[J]. 世界科学技术: 中医药现代化, 2020, 22(6): 1947-1954.

[14] 肖 瑞, 胡冯菊, 裴 卫. 基于 BiLSTM-CRF 的中医文本命名实体识别[J]. 世界科学技术: 中医药现代化, 2020, 22(7): 2504-2510.

[15] WANG C Y, WANG H, ZHUANG H, et al. Chinese medical named entity recognition based on multi-granularity semantic dictionary and multimodal tree[J]. Journal of Biomedical Informatics, 2020, 111: 103583.

[16] 汤洁仪, 李大军, 刘 波. 基于 BERT-BiLSTM-CRF 模型的地理实体命名实体识别[J]. 北京测绘, 2023, 37(2): 143-147.

[17] 王世民. 基于深度学习的中文电子病历命名实体识别研究: 以脑卒中为例[D]. 武汉: 华中科技大学, 2020.

[18] 贾 波, 许二平. 方剂学[M]. 新世纪 3 版. 北京: 中国中医药出版社, 2021.

[19] 钟赣生,杨柏灿. 中药学[M]. 新世纪 5 版. 北京: 中国中医药出版社, 2021.

[20] 国家市场监督管理总局, 国家标准化管理委员会. 中医临床诊疗术语 第 1 部分: 疾病: GB/T 16751.1—2023[S]. 北京: 中国标准出版社, 2023.

[21] 国家市场监督管理总局, 国家标准化管理委员会. 中医临床诊疗术语 第 2 部分: 证候: GB/T 16751.2—2021[S]. 北京: 中国标准出版社, 2021.

[22] 国家市场监督管理总局, 国家标准化管理委员会. 中医临床诊疗术语 第 3 部分: 治法: GB/T 16751.3—2023[S]. 北京: 中国标准出版社, 2023.

[23] 祖 弦, 谢 飞, 刘啸剑. 融合词和文档嵌入的关键词抽取算法[J]. 计算机科学与探索, 2021, 15(2): 294-304.

[24] ZHANG R Y, ZHAO P Y, GUO W Y, et al. Medical named entity recognition based on dilated convolutional neural network[J]. Cognitive Robotics, 2022, 2: 13-20.

[25] HE Z R, LUO X N, ZHONG Y R, et al. Information extraction method based on dilated convolution and character-enhanced word embedding[C]//2020 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC). Chongqing, China. IEEE, 2020: 138-143.

[26] PATIL N, PATIL A, PAWAR B. Named entity recognition using conditional random fields[J]. Procedia Computer Science, 2020, 167(C): 1181-1188.

(本文编辑 周 旦)