



DOI:10.12404/j.issn.1671-1815.2404910

引用格式: 谭平, 刘惠娜, 韦昌法. 基于 BiLSTM-CRF 和 Neo4j 的脾胃病知识图谱构建[J]. 科学技术与工程, 2025, 25(22): 9436-9444.

Tan Ping, Liu Huina, Wei Changfa, et al. Spleen and stomach disease knowledge graph construction based on BiLSTM-CRF and Neo4j[J]. Science Technology and Engineering, 2025, 25(22): 9436-9444.

# 基于 BiLSTM-CRF 和 Neo4j 的脾胃病知识图谱构建

谭平<sup>1</sup>, 刘惠娜<sup>2</sup>, 韦昌法<sup>1\*</sup>

(1. 湖南中医药大学信息科学与工程学院, 长沙 410208; 2. 湖南中医药大学医学院, 长沙 410208)

**摘要** 为了推动中医药文本数据的分析和挖掘,实现知识的智能化提取和处理。通过采用 BIO(begin, inside, outside) 序列标记法、BiLSTM-CRF 模型和人工定义规则的方法完成知识抽取任务。利用 Python3.6 中 Py2neo 库与 Neo4j 数据库等技术,构建了基于 Neo4j 的脾胃病知识图谱,并使用 Flask 框架开发了中医药脾胃病命名实体识别系统。结果表明:模型(BiLSTM-CRF)在测试集上取得了高性能和良好的泛化能力,其准确率、精确率、查全率和  $F_1$  分数分别为 96.19%、86.64%、88.82% 和 87.71%。构建出的知识图谱包含了方剂或中成药、中药、临床表现等 8 种节点标签以及 10 种关系类型,可支持中医药治疗脾胃病西医诊断、中医证候、中医治则等节点及各节点之间关系的查询与发现。可见 BiLSTM-CRF 模型在中医药脾胃病命名实体识别方面展现出了良好的通用性,它在处理复杂文本结构和领域术语上表现出了出色的能力,为中医药脾胃病知识提取和知识图谱构建研究提供了有力支持。

**关键词** 中医药; 脾胃病; BIO 序列标记法; 人工定义规则; BiLSTM-CRF 模型; 知识图谱  
中图法分类号 TP391.1 R256.3; 文献标志码 A

## Spleen and Stomach Disease Knowledge Graph Construction Based on BiLSTM-CRF and Neo4j

TAN Ping<sup>1</sup>, LIU Hui-na<sup>2</sup>, WEI Chang-fa<sup>1\*</sup>

(1. School of Informatics, Hunan University of Chinese Medicine, Changsha 410208, China;  
2. Medical School, Hunan University of Chinese Medicine, Changsha 410208, China)

**[Abstract]** In order to advance the analysis and mining of TCM (traditional Chinese medicine) text data and achieve intelligent extraction and processing of knowledge, the BIO (begin, inside, outside) sequence labeling method, the BiLSTM-CRF model, and manually defined rules were adopted to complete the knowledge extraction task. Utilizing the Py2neo library in Python 3.6 and the Neo4j database, a spleen and stomach disease knowledge graph was constructed based on Neo4j, and a TCM spleen and stomach disease named entity recognition system was developed using the Flask framework. The results show that the BiLSTM-CRF model achieves high performance and good generalization ability on the test set, with accuracy, precision, recall, and  $F_1$  scores of 96.19%, 86.64%, 88.82%, and 87.71%, respectively. The constructed knowledge graph includes eight types of node labels, such as prescriptions or patent medicines, Chinese medicines, and clinical manifestations, as well as ten types of relationships. It supports the querying and discovery of nodes and relationships among Western medical diagnosis, TCM syndromes, and TCM treatment principles for spleen and stomach diseases. It is concluded that the BiLSTM-CRF model demonstrates excellent generalizability in named entity recognition of TCM spleen and stomach disease. It exhibits outstanding performance in handling complex text structures and domain-specific terminology, providing strong support for the research on knowledge extraction and knowledge graph construction in Traditional Chinese Medicine for spleen and stomach diseases.

**[Keywords]** traditional chinese medicine; spleen and stomach diseases; BIO sequence tagging method; manually defined rules; BiLSTM-CRF model; knowledge graph

随着大数据和人工智能等相关技术的快速发展<sup>[1-2]</sup>, 医疗领域的数据日益增多, 其中包含了大量

有价值的医疗数据<sup>[3]</sup>。然而这些数据通常是非结构化的, 里面蕴涵着丰富而复杂的医学术语和领域

收稿日期: 2024-07-01; 修订日期: 2025-04-28

基金项目: 湖南省自然科学基金(2025JJ90089); 湖南省研究生科研创新立项不资助项目(LXBZZ2024175); 湖南中医药大学信息科学与工程学院科研资助专项(DK2023KY05)

第一作者: 谭平(2001—), 男, 汉族, 湖南邵阳人, 硕士研究生。研究方向: 医学自然语言处理。E-mail: 3207642976@qq.com。

\* 通信作者: 韦昌法(1982—), 男, 壮族, 广西巴马人, 博士, 教授。研究方向: 中医智能辅助诊疗与医学人工智能。E-mail: weichangfa@hnuem.edu.cn。

投稿网址: www.stae.com.cn

知识,使得其中的信息提取和理解变得非常具有挑战性。传统的信息处理方式效率低下,难以满足临床研究者的需求,因此,如何高效整合和利用这些珍贵的医疗资源成为当前急需解决的问题。

知识图谱<sup>[4]</sup>作为一种强大的知识表示和管理工具,能够精准地表征复杂的知识结构和关联,尤其适用于中医领域的知识特点<sup>[5]</sup>。自谷歌 2012 年提出这一概念以来,就在制造<sup>[6]</sup>、旅游<sup>[7]</sup>和公共基础设施<sup>[8]</sup>等领域展现出了巨大的潜力。而在医疗领域,Xing 等<sup>[9]</sup>基于文本挖掘构建了罕见病知识图谱;Liu 等<sup>[10]</sup>利用电子病历构建了类风湿性关节炎知识图谱;柴嘉琪等<sup>[11]</sup>基于文献数据探索性构建了中医药治疗糖尿病周围神经病变的知识图谱;牟丽君<sup>[12]</sup>和杨晶尧等<sup>[13]</sup>分别构建了中医古方知识图谱和中医医案知识图谱。然而,这类知识图谱在实际的临床应用中价值有限,存在着数据缺失较多<sup>[14]</sup>、知识实时更新差、无效数据和冗余信息过多<sup>[15]</sup>等问题,并且添加新的知识和复杂数据的处理需要大量的人工工作。因此,深度学习技术的兴起为知识图谱的构建提供了新的方法,并被广泛应用于序列标注和命名实体识别等任务中,为医学实体识别提供了强有力的技术支撑。如吴佳泽等<sup>[16]</sup>则通过预训练模型及条件随机场实现了对中医医案文本的命名实体识别。

中国传统医学的主要组成部分是中医药,其积累了上千年的临床经验和治疗方法<sup>[17]</sup>,在治疗脾胃病上拥有着丰富的经验和知识。基于脾胃病的知识图谱构建<sup>[18]</sup>有助于整合分散的医疗数据和知识,为脾胃病等疾病的诊断、治疗和研究提供了新的方法和思路。

为了探索基于 Neo4j 的中医药脾胃病知识图谱的构建方法,并探讨其在脾胃病诊断、治疗和研究中的潜在应用价值,现侧重于使用双向长短期记忆模型(bidirectional long short-term memory-conditional random field, BiLSTM-CRF)和人工定义规则的方法进行命名实体识别,以处理中医药脾胃病文本中的复杂结构和特定领域术语。期望能为中医药脾胃病领域的知识整合与应用提供新的思路和技术支持,促进中医药脾胃病在临床实践中的应用和推广。

## 1 脾胃病知识图谱构建方法

### 1.1 数据和知识来源

本文研究所用的数据和知识来源于第二届“博览医书”杯全国中医药知识图谱构建与应用大赛,使用该比赛提供的数据集,分别是 medical. dev、medical. test、medical. train 和 train. txt。

### 1.2 BIO 序列标记法

在自然语言处理中,序列标记技术通常用于对给定的输入序列(如文本句子)中的每个词或字符分配一个特定的标记或类别<sup>[19]</sup>。这些标记通常包括实体的类型、开始位置和结束位置。本文研究采用 BIO 序列标记法<sup>[20]</sup>,其标签含义如表 1 所示。实体的起始位置会标记为 B-X,其中 X 代表实体类型,中间和结束位置会标记为 I-X,其他的非实体部分统一标记为 O<sup>[21]</sup>。经标注后,输入序列的每个字符会与实体标签信息一一对应,形成“字符+标签”对应形式的数据。由于 BIO 标注提供模型训练所需的真实标签,故标记数据的质量,直接决定了模型预测的准确性和泛化能力的好坏<sup>[22]</sup>。

表 1 BIO 标注模式

Table 1 BIO sequence labeling method

标签	代表英文	标签含义
B	Begin	实体开始位置
I	Inside	实体中间或结束位置
O	Other	非实体,用于标记无关字符

### 1.3 BiLSTM-CRF 模型

本文研究将双向长短期记忆网络(bidirectional long short-term memory, BiLSTM)与 CRF(conditional random field, CRF)网络结合,形成双向长短期记忆模型(BiLSTM-CRF)<sup>[23]</sup>。BiLSTM-CRF 模型除了可以使用过去输入的特征和句子级标签信息外,还可以使用未来的输入特征,使其能更好地处理序列标注任务,提高标记的准确性。BiLSTM-CRF 模型架构如图 1 所示,总体分为 3 层。

第一层是词嵌入层,旨在将每个输入字符  $x$  转换为字符嵌入。将输入层中的每个词或字符映射到一个高维实数向量空间中,这些向量可以捕捉词语或字符之间的语义和语法信息。

第二层是 BiLSTM 层,通常作为序列数据处理的一部分,其利用正向 LSTM 和反向 LSTM 分别计

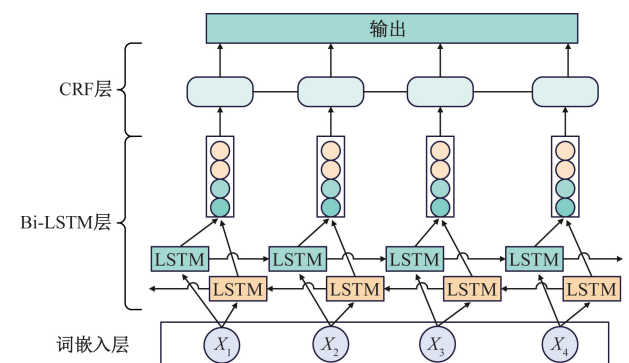


图 1 BiLSTM-CRF 模型结构图

Fig. 1 BiLSTM-CRF model structure diagram

算每个词考虑左侧和右侧词时对应的向量,并将两个向量相结合,形成词向量输出到 CRF 层。BiLSTM 层的输出为 BIO 标记中的每一个标签分值,其可以挑选分值最高的一个作为该单元的标签,但是标签不能保证每次都能预测正确<sup>[24]</sup>。

第三层为标注层,即 CRF 层和输出层。以 BiLSTM 层输出的词向量作为输入,对句中命名实体进行序列标注。第二层 BiLSTM 的输出可能会得到不合语法的标记序列,而 CRF 层可以从训练数据集中学习到正确的约束,并对 BiLSTM 层的输出进行约束和修正,以保证预测结果的正确性。例如输出的结果中,句子中第一个单词的标签应该以 B 或 O 开头。当存在连续标签时,实体的第一个标签应该以“B-”开头,而不是“I-”,且应为同一命名实体类别标签,如“B-CP I-CP”有效、“B-CP I-OT”无效<sup>[25]</sup>。

#### 1.4 评价指标

本研究使用准确率 (accuracy,  $A$ )、精确率 (precision,  $P$ )、召回率 (recall,  $R$ ) 和 ( $F_1$ -score,  $F_1$ ) 分数作为模型训练效果优劣的评估指标<sup>[26]</sup>,其分数越趋近 100%,模型的训练效果就越好。

准确率是衡量正确预测的样本占样本总数的比例,计算公式为

$$A = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \quad (1)$$

精确率是指在所有预测为正的样本中实际为正样本的概率,计算公式为

$$P = \frac{TP}{TP + FP} \times 100\% \quad (2)$$

召回率是指模型能够正确判断为正样本的数量与所有实际正样本数量之比,计算公式为

$$R = \frac{TP}{TP + FN} \times 100\% \quad (3)$$

$F_1$ 反映了模型的综合评估效果,可以同时考虑到准确率和召回率,计算公式为

$$F_1 = \frac{2PR}{P + R} \times 100\% \quad (4)$$

式中:TP 为真正例;FP 为假正例;TN 为真反例;FN 为假反例。

#### 1.5 知识图谱概念

知识图谱本质上是一种语义网络,即结构化的语义知识库,逻辑上可分为模式层与数据层<sup>[27]</sup>,如表 2 所示。数据层由一系列的事实组成,常见的存储形式是三元组。模式层构建在数据层之上,是知识图谱的核心,其中蕴含了实体之间的概念及关系。通过这种逻辑结构,知识图谱可以表示和存储各种实体及其关系,从而形成一个丰富的语义网络。将知识图谱应用在医疗领域可以整合领域内

表 2 知识图谱逻辑结构

Table 2 Logical structure of the knowledge graph

结构层次	主要内容	示例
模式层	有关的数据模型	概念及关系
数据层	具体数据	三元组

的大量数据,帮助医生更好地了解疾病之间的关系、病因、症状和治疗方法。

#### 1.6 知识图谱构建

##### 1.6.1 流程设计

本文研究的流程设计如图 2 所示。首先采用 BIO 序列标记法构建出训练集、验证集和测试集。接着利用 BiLSTM-CRF 算法模型在训练集上训练,并通过验证集调节超参数,使得训练的模型各个评价指标最佳。最终在测试集上进行测评,得到测评结果。然后,按照知识图谱自顶向下的构建方法,定义实体、关系及属性。接着利用评价指标最佳的 BiLSTM-CRF 模型和人工定义规则的方法从非结构化和半结构化数据中抽取实体、关系及属性。最后,将抽取的知识融合并导入 Neo4j 图数据库,完成脾胃病知识图谱的构建,实现关键词检索、临床辅助决策和图谱可视化。

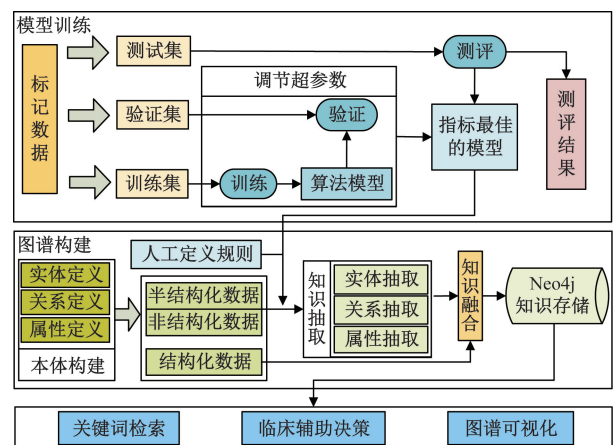


图 2 脾胃病知识图谱构建流程图

Fig. 2 Flowchart of spleen and stomach disease knowledge graph construction

##### 1.6.2 数据处理

本研究使用 BIO 标记法对中医诊断、临床表现和方剂或中成药等 10 种实体类别分别进行标注,具体实体类别和标签信息如表 3 所示。

以文本“恶心呕吐、食欲下降、上腹不适、暖气、烧心、反酸等”为例,应用 BIO 标签后,如表 4 所示。文本中的每个字符都被标记为临床表现 (CP) 或其他 (O)。如“B-CP”表示临床表现的起始位置,“I-CP”表示临床表现的中间或者结束位置,“O”表示无关字符。

表 3 实体类型标签信息  
Table 3 Entity type label information

实体类型	标签符号	开始标签	中间/结束标签	示例
中医诊断	TCMD	B-TCMD	I-TCMD	黄疸
临床表现	CP	B-CP	I-CP	形体瘦弱
方剂或中成药	PACPM	B-PACPM	I-PACPM	逍遥散加味
中药	CHM	B-CHM	I-CHM	夏枯草
西医诊断	WMD	B-WMD	I-WMD	病毒性肝炎
中医适宜技术	TCMT	B-TCMT	I-TCMT	穴位敷贴
中医证候	TCMS	B-TCMS	I-TCMS	脾虚血瘀
西药治疗	WMT	B-WMT	I-WMT	氟西汀片
其他治疗	OT	B-OT	I-OT	心理支持疗法
中医治则	POTCM	B-POTCM	I-POTCM	调肝理脾
其他	O	O	O	不包括在上述实体中

表 4 数据和标签之间的对应关系

Table 4 The correspondence between data and labels

字符	标签	字符	标签	字符	标签
恶	B-CP	降	I-CP	气	I-CP
心	I-CP	、	O	、	O
呕	I-CP	上	B-CP	烧	B-CP
吐	I-CP	腹	I-CP	心	I-CP
、	O	不	I-CP	、	O
食	B-CP	适	I-CP	反	B-CP
欲	I-CP	、	O	酸	I-CP
下	I-CP	暖	B-CP	等	O

表 5 模型参数设置

Table 5 Model parameter settings

参数	参数含义	设置值
minibatch	每个小批次的样本数量	64
epochs	训练的轮数	60
hidden_dim	隐藏状态的维度	300
optimizer	优化器类型	Adam
lr	学习率	0.001
clip	梯度裁剪的阈值	5
dropout	神经元保留比例	0.5
embedding_dim	字符嵌入的维度	300

1.6.3 超参数设置

大多数学习算法都需要设定参数,不同的参数配置往往会影 响模型学习的性能。本文研究采用 Python 3.6 和 TensorFlow1.2.0 框架,通过控制变量法<sup>[28]</sup>进行多次调参实验,并在验证集上观测衡量模型取不同参数时的训练效果。最终确定超参数取表 5 中的值时,模型的训练效果达到最佳。

1.6.4 本体层构建

本体构建是指在模式层规范化描述指定领域内的概念及其关系,其构成了知识图谱的基本架构<sup>[29]</sup>。本文研究将“中医诊断”“临床表现”等 10 种实体作为研究的本体。其中,“方剂或中成药”和“中药”可以抽象为“药品”作为一级本体,而“方剂或中成药”和“中药”则作为二级本体。同理“西药治疗”“其他治疗”“中医治则”可以抽象为一级本体“治疗方式”。为完善脾胃病相关知识,故增加了“穴位”这一本体,由基于人工定义规则的实体识别方法<sup>[30]</sup>进行提取。最终设计了表 6 所示的本体结构。

1.6.5 实体关系构建

本文研究采用知识图谱中的三元组模型,来描述脾胃病中的知识和关系。将实体间的对应关系描述为:西医诊断-其他治疗的关系,西医诊断-西药治疗的关系,西医诊断-中医治疗的关系,西医诊断-

表 6 脾胃病知识图谱本体结构

Table 6 Ontology structure of spleen and stomach disease knowledge graph

一级本体	二级本体	实例
西医诊断		病毒性肝炎
中医诊断		黄疸
临床表现		形体瘦弱
中医适宜技术		穴位敷贴
中医证候		脾虚血瘀
穴位		足三里
治疗方式	西药治疗	氟西汀片
	其他治疗	心理支持疗法
	中医治则	调肝理脾
药品	方剂或中成药	逍遥散加味
	中药	夏枯草

方剂或中成药的关系,西医诊断-中药的关系,西医诊断-中医诊断的关系,西医诊断-中医适宜技术的关系,西医诊断-中医证候的关系,西医诊断-临床表现的关系和西医诊断-穴位的关系共 10 条关系。基于这些关系,本文研究设计出了脾胃病知识图谱三元组结构,如表 7 所示。

1.7 知识存储及可视化

本文研究使用 Neo4j 图数据库完成知识图谱数据层的存储和可视化。Neo4j 主要是按节点和边组织数据进行数据的存储,其中节点表示知识图谱中

表7 脾胃病知识图谱三元组结构表

Table 7 Spleen and stomach disease knowledge graph triplet structure table

实体 1	实体关系类型	中文含义	实体 2
西医诊断	rels_Western_Medicine_Treatment	西医诊断与西药治疗的关系	治疗方式
西医诊断	rels_Other_Treatments	西医诊断与其他治疗的关系	治疗方式
西医诊断	rels_Principles_of_Traditional_Chinese_Medicine	西医诊断与中医治疗的关系	治疗方式
西医诊断	rels_Prescriptions_and_Chinese_Patent_Medicines	西医诊断与方剂或中成药的关系	药品
西医诊断	rels_Chinese_Herbal_Medicine	西医诊断与中药的关系	药品
西医诊断	rels_Traditional_Chinese_Medicine_Syndromes	西医诊断与中医证候的关系	中医证候
西医诊断	rels_Acupoint	西医诊断与穴位的关系	穴位
西医诊断	rels_Traditional_Chinese_Medicine_Diagnosis	西医诊断与中医诊断的关系	中医诊断
西医诊断	rels_Traditional_Chinese_Medicine_Techniques	西医诊断与中医适宜技术的关系	中医适宜技术
西医诊断	rels_Clinical_Presentation	西医诊断与临床表现的关系	临床表现

的实体,边表示实体之间的关系<sup>[31]</sup>。本文研究将文本中提取到的“中医诊断”等 11 种实体及其关系整理为结构化数据,经知识融合后,使用 Python 的 Py2neo 库将数据导入 Neo4j,进行知识的存储和图谱的构建。其中 Neo4j 强大的查询功能可以对脾胃病的相关信息进行有效的检索与分析。

## 2 脾胃病知识图谱构建结果

### 2.1 模型训练结果

模型识别的准确率决定了知识图谱构建的质量,BiLSTM-CRF 模型的最终性能评测结果如表 8 所示,其在验证集和测试集上均取得了较高的分数,精确率超过 95%, $F_1$  超过 86%。这表明模型在识别各项实体上表现优异,证明了其在脾胃病实体抽取方面具有高效性和可靠性。

本文研究也对于 10 个不同的实体类别进行了分析,其评测结果如表 9 所示。CHM、POTCM、WMD 与 WMT 等实体的各项评价指标相对较好,这可能与它们在数据集中占比较高有关,说明数据集中实体类型的分布可能会对模型的性能产生影响。通常实体类型的样本量越大可能就会产生更好的结果,使模型在识别这些实体上具有高的准确度。然而,OT 实体的表现相对较差,这可能是由于该类别数据量较少或特征较为复杂所致,也可能是受到样本质量和数据标注质量等其他因素的影响,说明模型对该类实体的识别有待改进。

### 2.2 脾胃病知识图谱

本文研究利用训练好的模型去提取非结构化的数据,并使用 Python 中的 py2neo 库构建知识图谱的节点和关系,经知识融合后。将已经提取并整理好的结构化数据导入 Neo4j 数据库,从而完成基于 BiLSTM-CRF 模型的脾胃病知识图谱构建。使用 Neo4j 的 Cypher 查询语句(MATCH(n) RETURN n)查询所构建的知识图谱,如图 3 所示。

表8 模型评测结果

Table 8 Model evaluation results

模型	A/%	P/%	R/%	$F_1$ /%
Dev set	95.77	85.23	88.04	86.61
Test set	96.19	86.64	88.82	87.71

表9 各个实体类别评测结果

Table 9 Results for each entity category

实体类别	P/%		R/%		$F_1$ /%	
	Dev set	Test set	Dev set	Test set	Dev set	Test set
CHM:	94.63	95.43	96.22	96.43	95.42	95.92
CP:	79.27	80.88	82.89	83.65	81.04	82.24
OT:	50.00	66.67	50.00	60.00	50.00	63.16
PACPM:	80.61	79.87	86.93	83.01	83.65	81.41
POTCM:	90.00	100.00	60.00	73.33	72.00	84.62
TCMD:	90.62	96.77	87.88	90.91	89.23	93.75
TCMS:	71.09	75.40	77.78	81.20	74.29	78.19
TCMT:	78.05	80.49	79.01	81.48	78.53	80.98
WMD:	93.02	93.52	95.69	96.65	94.34	95.06
WMT:	92.16	89.09	90.38	94.23	91.26	91.59

根据表 10,构建出的图谱总共包含 2 849 个实体。药品实体有 961 个,临床表现有 1 135 个,而穴位实体仅有 7 个。这表明了脾胃病在治疗方面涉及的药物种类繁多和临床上的症状表现多样,意味着药品节点可能存在着较大的复杂性和信息量,而临床表现的节点可能承载了丰富的病情描述和临床信息,对诊断和治疗决策具有重要意义。穴位实体数量较少说明了脾胃病在穴位治疗方面的局限性,涉及的信息和关联较为有限。通过分析,可以知道不同的实体类型在数据集中的分布情况和具体含义,这对于进一步的数据处理和模型训练具有指导意义。

### 2.3 脾胃病命名实体识别系统

本文研究采用基于规则和词典的方法<sup>[32]</sup>开发了一个针对中医药脾胃病的命名实体识别系统,并使用 Flask 框架实现。该系统界面左侧输入中医药脾胃病病例,中间显示实体类别,右侧展示识别结

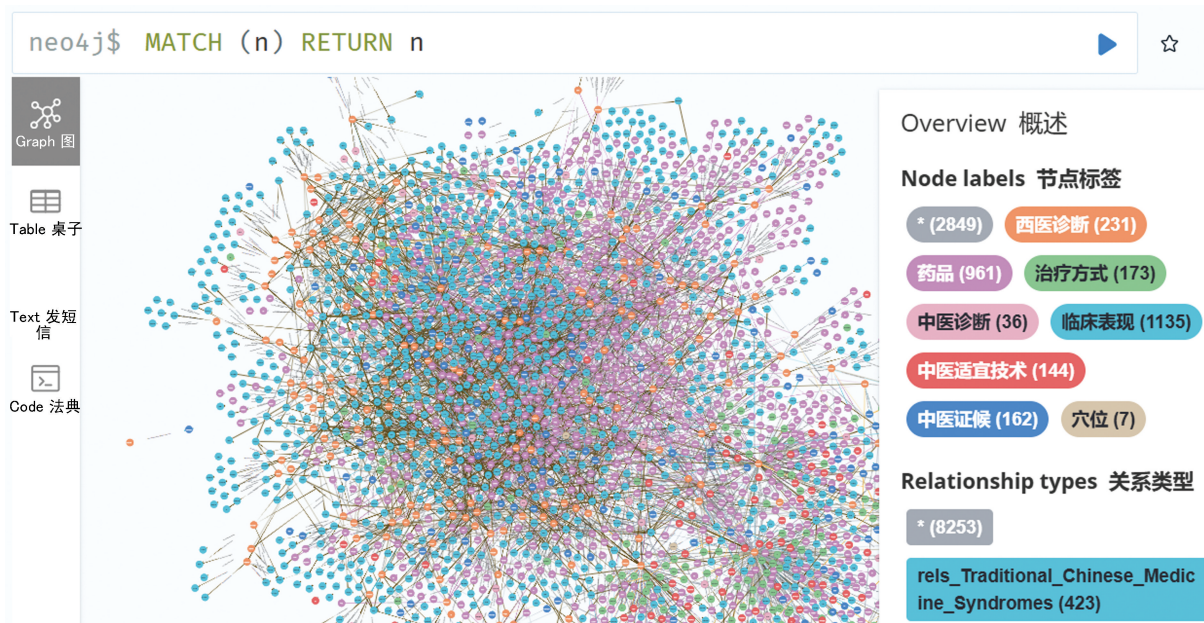


图3 脾胃病知识图谱

Fig. 3 Knowledge graph of spleen and stomach diseases

表10 脾胃病知识图谱实体类型统计表

Table 10 Statistical table of entity types of spleen and stomach disease knowledge graph

实体类型	中文含义	实体数量	举例
Traditional_Chinese_Medicine_Diagnosis	中医诊断	36	痞满
Drugs	药品	961	健脾和胃汤
Western_Medicine_Diagnosis	西医诊断	231	功能性消化不良
Cure_ways	治疗方式	173	疏肝理气,心理疏导
Clinical_Presentation	临床表现	1 135	上腹部不适,食欲不振
Traditional_Chinese_Medicine_Techniques	中医适宜技术	144	辨证针刺
Traditional_Chinese_Medicine_Syndromes	中医证候	162	肝郁气滞
Acupoints	穴位	7	天枢,足三里
总计		2 849	约3 000 实体量级

果。该系统能够帮助医生从脾胃病病例中自动提取“穴位”“中医诊断”等实体。

系统的测试效果如图4所示,以一则医案<sup>[33]</sup>为例,测试结果输出中药,能够识别出大部分的实体,对脾胃病医案的识别有较好效果。

### 3 结论

针对中医药文本进行命名实体识别,实现了非结构化数据的知识抽取。在此基础上,通过使用 Neo4j 图数据库完成了中医药脾胃病知识图谱的构建,并开发了相应的命名实体识别系统,得出以下结论。

(1)通过 BiLSTM-CRF 模型和人工定义规则相结合的方法,解决了中医药文本中复杂结构和特定术语识别的难题,提供了一种高效的知识抽取方法。通过对比 8 种实体类型的数量,发现数据集中

实体类型的分布对模型的性能有一定的影响,样本量更多的实体类型往往能取得更好的结果,实验也证明了 BiLSTM-CRF 模型在中医药脾胃病命名实体识别任务中的有效性和实用性。这项工作对于推动中医药领域的信息提取、知识图谱构建以及基于自然语言处理的应用具有重要意义,该研究的方法和思路也为其他单病种知识图谱的构建提供了有益的参考和借鉴。

(2)通过 Neo4j 图数据库构建了中医药脾胃病知识图谱,实现了领域知识的重新组织和展示,为医师经验传承和创新奠定基础。其不仅包括了和脾胃病有关的“中医诊断”“临床表现”等 10 个实体和实体之间的关系,还涵盖了疾病与“穴位”之间的关联。这种以图的形式呈现知识,并通过节点和关系的方式构建的语义连接,多层次、多维度的知识表示模式为中医药脾胃病的研究和应用提供了全



图4 中医药脾胃病命名实体识别系统

Fig. 4 Traditional chinese medicine spleen and stomach disease named entity recognition system

新的视角。通过输入特定的检索词信息,知识图谱可以展示查询结果并且同时输出具有高度关联的其他信息,可为医患获取信息和解决措施提供专业指导,支持医生做出科学且精准的医疗决策。这种快速、智能和数字化的呈现方式在临床实践中具有实际应用价值,对于提高诊疗效率和用户体验具有重要意义。

(3)开发了中医药脾胃病命名实体识别系统,该系统能够帮助医疗企业和机构快速有效地从病例中自动提取有关实体,使得研究者和临床医生能够更深入地分析中医药文献中的脾胃病有关实体信息。在脾胃病病历的审阅中,利用该系统可以快速识别提取临床表现、证候等关键实体信息,帮助医生更快地了解病情并做出诊断。同时,在促进脾胃病知识管理和优化方面能够发挥重要作用,实现中医药领域的可持续发展和创新发展的目标,并有利于为建立相关数据库、医药理论研究和新药研发提供可靠的数据支持,促进了中医药信息化的发展。

(4)本文研究也存在一些不足之处。一是模型训练的时间过长,模型识别的效果依赖参数设置和人工标注的数据,后续还需要不断优化算法和工具,进一步完善和拓展知识图谱的内容,以便更好地满足临床实践的需求。二是构建的中医药脾胃病命名实体识别系统能在特定领域和任务中取得良好的效果,但其编制过程非常耗时且难以涵盖所有的语言现象,为扩展其适用范围,对于不同的系统则需要该领域专家重新书写规则。

## 参 考 文 献

- [1] Carini C, Seyhan A A. Tribulations and future opportunities for artificial intelligence in precision medicine[J]. Journal of Translational Medicine, 2024, 22(1): 411-421.
- [2] Seo S, Lee J W. Applications of big data and ai-driven technologies in CADD(computer-aided drug design)[J]. Methods in Molecular Biology, 2024, 2714: 295-305.
- [3] Wang M, Sushil M, Miao B Y, et al. Bottom-up and top-down paradigms of artificial intelligence research approaches to healthcare data science using growing real-world big data[J]. Journal of the American Medical Informatics Association, 2023, 30(7): 1323-1332.
- [4] 王彩云, 郑增亮, 蔡晓琼, 等. 知识图谱在医学领域的应用综述[J]. 生物医学工程学杂志, 2023, 40(5): 1040-1044.  
Wang Caiyun, Zheng Zengliang, Cai Xiaoqiong, et al. Overview of the application of knowledge graphs in the medical field[J]. Journal of Biomedical Engineering, 2023, 40(5): 1040-1044.
- [5] 刘金垒, 惠小珊, 张振鹏, 等. 基于中医诊疗指南的冠心病知识图谱构建[J]. 中国实验方剂学杂志, 2023, 29(7): 208-215.  
Liu Jinlei, Hui Xiaoshan, Zhang Zhenpeng, et al. Knowledge graph construction of coronary heart disease based on traditional Chinese medicine diagnosis and treatment guidelines[J]. Chinese Journal of Experimental Traditional Medical Formulas, 2023, 29(7): 208-215.
- [6] 刘浩, 张建业, 吕张成, 等. 面向数控机床设计知识图谱构建的实体识别[J]. 科学技术与工程, 2023, 23(13): 5655-5661.  
Liu Hao, Zhang Jianye, Lü Zhangcheng, et al. Entity recognition for CNC machine tool design knowledge graph[J]. Science Technology and Engineering, 2023, 23(13): 5655-5661.
- [7] 徐春, 苏明钰, 孙彬. 基于 ChatGLM 和提示微调的旅游知识图

- 谱构建[J]. 科学技术与工程, 2024, 24(31): 13484-13492.
- Xu Chun, Su Mingyu, Sun Bin. Tourism knowledge graph construction based on ChatGLM and prompt-tuning[J]. Science Technology and Engineering, 2024, 24(31): 13484-13492.
- [8] 姚卫华, 葛悦, 邓兵兵, 等. 城镇燃气管网应急处置知识图谱构建与应用[J]. 科学技术与工程, 2024, 24(10): 4343-4351.
- Yao Weihua, Ge Yue, Deng Bingbing, et al. Research on the construction and application of emergency response knowledge graph for urban gas pipeline networks[J]. Science Technology and Engineering, 2024, 24(10): 4343-4351.
- [9] Xing H, Zhang D, Cai P, et al. RDBridge: a knowledge graph of rare diseases based on large-scale text mining[J]. Bioinformatics, 2023, 39(7): btad440.
- [10] Liu F, Liu M, Li M, et al. Automatic knowledge extraction from chinese electronic medical records and rheumatoid arthritis knowledge graph construction [J]. Quantitative Imaging in Medicine and Surgery, 2023, 13(6): 3873-3890.
- [11] 柴嘉琪, 谈钰濛, 向兴华, 等. 基于文献数据的知识图谱构建——以中医药治疗糖尿病周围神经病变为例[J]. 中国实验方剂学杂志, 2024, 30(6): 144-150.
- Chai Jiaqi, Tan Yumeng, Xiang Xinghua, et al. Construction of knowledge graph based on literature data by taking treatment of diabetic peripheral neuropathy with traditional Chinese medicine as an example[J]. Chinese Journal of Experimental Traditional Medical Formulas, 2024, 30(6): 144-150.
- [12] 牟丽君. 中医古方知识图谱的构建与应用研究[D]. 上海: 华东师范大学, 2024.
- Mou Lijun. Research on the construction and application of traditional chinese medicine formula knowledge graph[D]. Shanghai: East China Normal University, 2024.
- [13] 杨晶尧, 杨方燕, 朱文慧, 等. 基于医案知识图谱的中医诊疗规律发现研究[J]. 世界中西医结合杂志, 2023, 18(4): 821-831, 835.
- Yang Jingyao, Yang Fangyan, Zhu Wenhui, et al. Discovery of TCM diagnosis and treatment rules from medical records by knowledge graph[J]. World Journal of Integrated Traditional and Western Medicine, 2023, 18(4): 821-831, 835.
- [14] 杨旭华, 高良煜. 基于动态分层强化学习的知识图谱推理[J]. 小型微型计算机系统, 2025, 46(5): 1081-1088.
- Yang Xuhua, Gao Liangyu. Knowledge graph reasoning based on dynamic hierarchical reinforcement learning[J]. Journal of Chinese Mini-Micro Computer Systems, 2025, 46(5): 1081-1088.
- [15] 戴高阳, 孟小艳, 张容祯. 基于深度学习的畜牧业知识图谱构建[J]. 计算机与数字工程, 2024, 52(6): 1746-1753, 1847.
- Dai Gaoyang, Meng Xiaoyan, Zhang Rongzhen. Construction of a knowledge graph for animal husbandry based on deep learning[J]. Computer & Digital Engineering, 2024, 52(6): 1746-1753, 1847.
- [16] 吴佳泽, 李坤宁, 陈明. 基于预训练模型及条件随机场的中医医案命名实体识别[J]. 中医药信息, 2023, 40(9): 38-45.
- Wu Jiazhe, Li Kunning, Chen Ming. Named entity recognition in traditional chinese medicine medical records based on pre-trained models and conditional random fields[J]. Journal of Traditional Chinese Medical Information, 2023, 40(9): 38-45.
- [17] 张欣怡, 张洪春. 论中医药治疗支气管扩张症[J]. 中华中医药杂志, 2023, 38(4): 1518-1521.
- Zhang Xinyi, Zhang Hongchun. Discussion on treatment of bronchiectasis with traditional Chinese medicine[J]. China Journal of Traditional Chinese Medicine and Pharmacy, 2023, 38(4): 1518-1521.
- [18] 宋源. 中医脾胃病知识图谱自动构建研究[D]. 苏州: 中国中医科学院, 2024.
- Song Yuan. Research on automatic construction of knowledge graph for traditional chinese medicine digestive system diseases [D]. Suzhou: China Academy of Chinese Medical Sciences, 2024.
- [19] 马艺洁, 赖海光, 刘子威, 等. 实体识别技术研究进展综述[J]. 太赫兹科学与电子信息学报, 2024, 22(5): 503-515.
- Ma Yijie, Lai Haiguang, Liu Ziwei, et al. Overview of the research progress in entity recognition technology [J]. Journal of Terahertz Science and Electronic Information Technology, 2024, 22(5): 503-515.
- [20] 陈勇刚, 刘康妮, 王帅. 基于 BiGRU-Attention 改进的航空设备故障知识图谱构建[J]. 航空学报, 2024, 45(18): 168-181.
- Chen Yonggang, Liu Kangni, Wang Shuai. Construction of knowledge graph for aerospace equipment fault based on improved BiGRU-attention[J]. Acta Aeronautica et Astronautica Sinica, 2024, 45(18): 168-181.
- [21] 许雅玺, 孟天宇, 王欣, 等. 航空不安全事件知识图谱构建[J]. 科学技术与工程, 2024, 24(32): 13925-13934.
- Xu Yaxi, Meng Tianyu, Wang Xin, et al. Research on construction of knowledge graph for aviation unsafe events[J]. Science Technology and Engineering, 2024, 24(32): 13925-13934.
- [22] 李成林, 赵珍威, 李国厚, 等. 基于知识图谱的农作物良种问答系统的设计与实现[J]. 河南科技学院学报(自然科学版), 2024, 52(1): 48-59.
- Li Chenglin, Zhao Zhenwei, Li Guohou, et al. Design and implementation of a recommendation system for improved crop varieties based on knowledge graph[J]. Journal of Henan Institute of Science and Technology (Natural Science Edition), 2024, 52(1): 48-59.
- [23] 周锦, 薛钰, 杨秉杰, 等. 基于深度学习 BiLSTM-CRF 模型的网络安全知识图谱实体识别方法[J]. 网络安全技术与应用, 2024(9): 41-44.
- Zhou Jin, Xue Yu, Yang Bingjie, et al. A method for entity recognition in cybersecurity knowledge graphs based on the deep learning BiLSTM-CRF model[J]. Network Security Technology & Application, 2024(9): 41-44.
- [24] 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.
- Xiao Rui, Hu Fengju, Pei Wei. Named entity recognition in traditional chinese medicine texts based on BiLSTM-CRF [J]. World Science and Technology-Modernization of Traditional Chinese Medicine, 2020, 22(7): 2504-2510.
- [25] 刘合兵, 张德梦, 熊蜀峰, 等. 融合 ALBERT 与规则的小麦病虫害命名实体识别[J]. 计算机科学与探索, 2023, 17(6): 1395-1404.
- Liu Hebing, Zhang Demeng, Xiong Shufeng, et al. Named entity recognition of wheat diseases and pests fusing albert and rules[J]. Journal of Frontiers of Computer Science and Technology, 2023,

- 17(6): 1395-1404.
- [26] 李先旺, 黄忠祥, 贺德强, 等. 汽车故障知识图谱构建及应用研究[J]. 科学技术与工程, 2024, 24(4): 1578-1587.  
Li Xianwang, Huang Zhongxiang, He Deqiang, et al. Research on the construction and application of failure knowledge graph in automobile field[J]. Science Technology and Engineering, 2024, 24(4): 1578-1587.
- [27] 方思怡. ISO 国际标准知识图谱的构建方法研究[J]. 标准科学, 2024(12): 73-77.  
Fang Siyi. Research on the construction method of ISO international standard knowledge graph [J]. Standards Science, 2024(12): 73-77.
- [28] 陈婕卿, 竹志超, 张锋, 等. 中文电子病历命名实体识别方法研究[J]. 医学信息学杂志, 2024, 45(4): 78-84.  
Chen Jieqing, Zhu Zhichao, Zhang Feng, et al. Study on named entity recognition of chinese electronic medical records[J]. Journal of Medical Informatics, 2024, 45(4): 78-84.
- [29] 李忠伟, 李明轩, 李永, 等. 基于鸟类迁徙关联要素数据的知识图谱构建[J]. 计算机工程与设计, 2024, 45(12): 3704-3711.  
Li Zhongwei, Li Mingxuan, Li Yong, et al. Construction of knowledge graph based on bird migration associated factor data [J]. Computer Engineering and Design, 2024, 45(12): 3704-3711.
- [30] 姜彬峰. 基于知识图谱的大数据实体识别方法[J]. 电脑知识与技术, 2023, 19(22): 74-77.  
Jiang Binfeng. A method for entity recognition in big data based on knowledge graphs [J]. Computer Knowledge and Technology, 2023, 19(22): 74-77.
- [31] Rajabi E, Kafaie S. Building a disease knowledge graph [J]. Studies Health Technology Information, 2023, 302: 701-705.
- [32] 李冬梅, 罗斯斯, 张小平, 等. 命名实体识别方法研究综述[J]. 计算机科学与探索, 2022, 16(9): 1954-1968.  
Li Dongmei, Luo Sisi, Zhang Xiaoping, et al. Review on named entity recognition [J]. Journal of Frontiers of Computer Science and Technology, 2022, 16(9): 1954-1968.
- [33] 熊航, 王梦琪, 刘亚军, 等. 沈洪基于温病理论治疗消化系统疾病验案4则[J]. 江苏中医药, 2023, 55(7): 52-55.  
Xiong Hang, Wang Mengqi, Liu Yajun, et al. Shen hong's four clinical cases of digestive system diseases treated based on the theory of warm diseases [J]. Jiangsu Journal of Traditional Chinese Medicine, 2023, 55(7): 52-55.