

基于人工智能的 类风湿性关节炎中医辅助诊疗系统

孙明俊¹ 张丹² 郑明智³ 梅舒欢²

摘要 类风湿性关节炎(RA)是一种广泛存在且慢性、难治的全身性免疫风湿病,中医在其治疗中具有副作用较少、价格相对低廉等优势,但是中医师的缺乏限制 RA 中医诊疗方案的推广。因此,文中提出基于人工智能的 RA 中医辅助诊疗系统。通过对患者病历文本和关节影像数据的学习实现对 RA 和 RA 证型的判断,辅助医生诊断,并根据证型智能推荐中医药方。文中还基于 RA 中医药典籍知识构建知识图谱,在医生诊疗过程中提供诊疗知识指导。系统可辅助经验不足的临床医师做出诊疗决策,提高 RA 的治疗水平,有助于对 RA 治疗的研究和推广。

关键词 类风湿性关节炎, 人工智能, 深度学习, 辅助诊断, 药方推荐

引用格式 孙明俊,张丹,郑明智,梅舒欢. 基于人工智能的类风湿性关节炎中医辅助诊疗系统. 模式识别与人工智能, 2021, 34(4): 343-352.

DOI 10.16451/j.cnki.issn1003-6059.202104006

中图法分类号 TP 391

Traditional Chinese Medicine Aided Diagnosis and Treatment System for Rheumatoid Arthritis Based on Artificial Intelligence

SUN Mingjun¹, ZHANG Dan², ZHENG Mingzhi³, MEI Shuhuan²

ABSTRACT Rheumatoid arthritis (RA) is a widespread, chronic and refractory systemic immune rheumatism. Traditional Chinese medicine (TCM) presents the advantages of less side effects and low price. However, the spread of RA TCM diagnosis and treatment scheme with curative effect advantages is limited due to the lack of experienced TCM practitioners, especially in primary medical institutions. In this paper, a traditional Chinese medicine aided diagnosis and treatment system for RA based on artificial intelligence is proposed. RA and pattern of syndrome in RA can be determined after learning patient medical records and medical imaging of joints, and then TCM prescription is recommended intelligently according to the pattern. Next, the information is exploited to assist doctors in diagnosis. Based on RA TCM knowledge, the knowledge graph is built. It provides doctors with knowledge guidance in the process of diagnosis and treatment. The system can assist less experienced doctors in making treatment decisions, improving the treatment level of RA, and studying and promoting RA treatment.

Key Words Rheumatoid Arthritis, Artificial Intelligence, Deep Learning, Aided Diagnosis, Prescription Recommendation

收稿日期:2020-06-15;录用日期:2021-02-01

Manuscript received June 15, 2020;

accepted February 1, 2021

国家重点研发计划项目(No. 2018YFC1705504)资助

Supported by National Key R&D Program of China(No. 2018YF C1705504)

本文责任编辑 杨勤

Recommended by Associate Editor YANG Meng

1. 中国信息通信研究院 云计算与大数据研究所 北京 100191

2. 南京新一代人工智能研究院 技术部 南京 210046

3. 中国信息通信研究院 人工智能(杭州)研究中心 杭州 311121

1. Cloud Computing and Big Data Research Institute, China Academy of Information and Communications Technology, Beijing 100191

2. Technical Department, Nanjing Research Institute of Next-Generation Artificial Intelligence, Nanjing 210046

3. Hangzhou Research Center of Artificial Intelligence, China Academy of Information and Communications Technology, Hangzhou 311121

Citation SUN M J, ZHANG D, ZHENG M Z, MEI S H. A Traditional Chinese Medicine Aided Diagnosis and Treatment System for Rheumatoid Arthritis Based on Artificial Intelligence. *Pattern Recognition and Artificial Intelligence*, 2021, 34(4): 343–352.

类风湿关节炎 (Rheumatoid Arthritis, RA) 是一种以手、足小关节的多关节、对称性、侵蚀性关节炎为主要表现的全身性免疫风湿疾病。RA 发病率高达 0.3% ~ 1%，是一种难治性疾病，病程反复、缠绵不愈、致残率高，可累及全身多个系统，严重危害人类健康。

近年来现代医学虽然在 RA 发病机制和病理变化等方面取得明显的研究进展，但是病因至今仍然不明确。中医药治疗 RA 具有较长的历史，大量的临床研究已表明中医药治疗 RA 的有效性。采用不同的治疗方案，中医药治疗 RA 的治愈率多数超过 90%^[1-2]。然而中医诊治方案存在治疗不规范、没有统一标准的问题，例如：中药种类繁多、中医师用药习惯不尽相同、千人千方等。而且，经验尚浅的中医师或基层医生由于对 RA 的诊治经验较少，开出的药方及治疗方案缺乏循证医学证据，因此无法保证对患者有效。

基于人工智能的辅助诊疗系统是基于人工智能相关技术构建的智能化诊疗决策支持系统，可辅助医生进行诊断决策，为医生提供诊疗建议。目前，一般通过医学影像和患者病历文本两类数据学习构建医疗领域的辅助诊疗系统。

医学影像分析与处理已成为医学信息中发展最快的领域之一，而物体分类、检测、分割和生成等作为基于人工智能 (Artificial Intelligence, AI) 的图像处理的关键问题，在医学应用中具有重要意义。目前，深度学习模型在多种影像模态相关任务中已接近甚至超越人类专家的水平，俞益洲等^[3]使用 YOLO 网络对乳腺肿块进行自动检测，定位准确率较高，并能区分良性和恶性病变。对于 RA 涉及到的手部等数字 X 线摄影 (Digital Radiography, DR) 关节影像的识别，目前基于人工智能的图像处理技术也同样具备良好的效果。

基于病历文本的辅助诊断通过对病历关键信息的提取、语义分析、相似数据匹配等自然语言处理技术，实现疾病预测、诊疗建议等任务。余帅帅等^[4]提出自身免疫疾病专家系统的设计思路和实现方案，运用专家系统的设计原理与方法，模拟医学专家诊断、治疗疾病的思维，并基于统计的方法采用机器学习或深度学习模型通过大量的标注数据形成训练方案。浦东旭^[5]设计和实现基于病历文本语义分析的

智能肝病辅助诊疗系统，提供基于词向量和深度学习技术的疾病预测、关键信息提取、智能语音转换、相似数据匹配、用户和病历数据管理等功能。因此，可通过基于规则或统计的方法实现基于文本的 RA 诊疗系统。

实践中结合医学的影像分析与处理和病历文本的分析，可建立多因子综合分析的系统，实现临床诊断与影像检查相结合的多模态辅助诊断。

另外，疾病的诊疗指南一般通过知识图谱的形式展现，同时，医学知识图谱还能为人工智能技术在医疗领域的应用提供知识基础。知识图谱作为知识的载体，能以图形化的方式将不易于理解的信息形象地表示出来^[6]。贾李蓉等^[7]提出中医药知识图谱构建的方法，以中医药学语言系统为骨架，将中医药领域主要的词表、术语资源和单表型数据库的内容导入中医药知识图谱之中，从海量中医药文献中发现新颖的实体和语义关系的方法，构建交互式的中医药文献知识获取工具，以半自动的方式扩展中医药知识图谱的内容。知识图谱构建的一项基础任务是实体识别。尤其在医疗病历中，实体和实体关系从不同角度反映电子病历中的医疗知识和患者的健康情况^[8]。目前，国内公开的医疗命名实体识别数据集相对较少，并无针对 RA 病历的实体数据集。

在辅助诊疗系统中，另一项重要的功能就是根据患者诊断结果、相关信息等自动推荐处方。Li 等^[9]提出端到端的中医药方自动生成的方法，基于具有注意力机制的 seq2seq 模型基础上，使用覆盖机制和遮盖机制 (Masking) 减少重复问题，通过症状的文本描述，更好地自动生成对应的中药处方。由于同种疾病可使用具备同种功效的不同药物，在药方推荐上很难采用绝对的方式评价，目前采用的评价方式一般是医生主观评价的方法。

目前，人工智能在医疗领域的应用研究主要是基于疾病诊疗过程中的某个模块。由于 RA 的中医诊疗需要结合患者的检验检查信息、病历文本信息和关节影像数据进行决策，并基于诊断结果提供中医药方用法、用量的推荐，而其它疾病的智能诊疗系统架构不能直接利用在 RA 中医诊疗中，同时由于 RA 相对小众、数据集匮乏等原因，行业内还缺失针对 RA 中医诊疗全过程的辅助诊疗系统。因此，本文

提出基于人工智能的 RA 中医辅助诊疗系统,基于人工智能,从诊疗指南、患者病历文本和关节影像数据及资深专家的处理和药方学习诊疗经验,实现对 RA 及 RA 证型的判断,并根据 RA 证型智能推荐中医药方.该系统可辅助经验不足的基层医生做出科学的判断,帮助医生开出有循证依据的药方,减少不同医生带来的差异,提高诊疗水平.同时,由于能够基于 RA 中医药典籍知识构建知识图谱,在诊疗过程中为医生提供诊疗知识指导,因此,在使用 RA 中医辅助诊疗系统时,医生可将知识图谱模块作为查询工具,查询症状、药方、药物等信息.引入人工智能

技术,构建 RA 中医辅助诊疗系统,可规范 RA 中医诊疗过程,减少对专家的依赖程度,提高 RA 诊疗水平,同时为 RA 治疗相关研究提供支撑.

1 类风湿性关节炎中医辅助诊疗系统

RA 中医辅助诊疗系统是辅助医生进行 RA 诊疗的系统,核心功能框图如图 1 所示,对于其它非核心功能,本文不进行介绍.

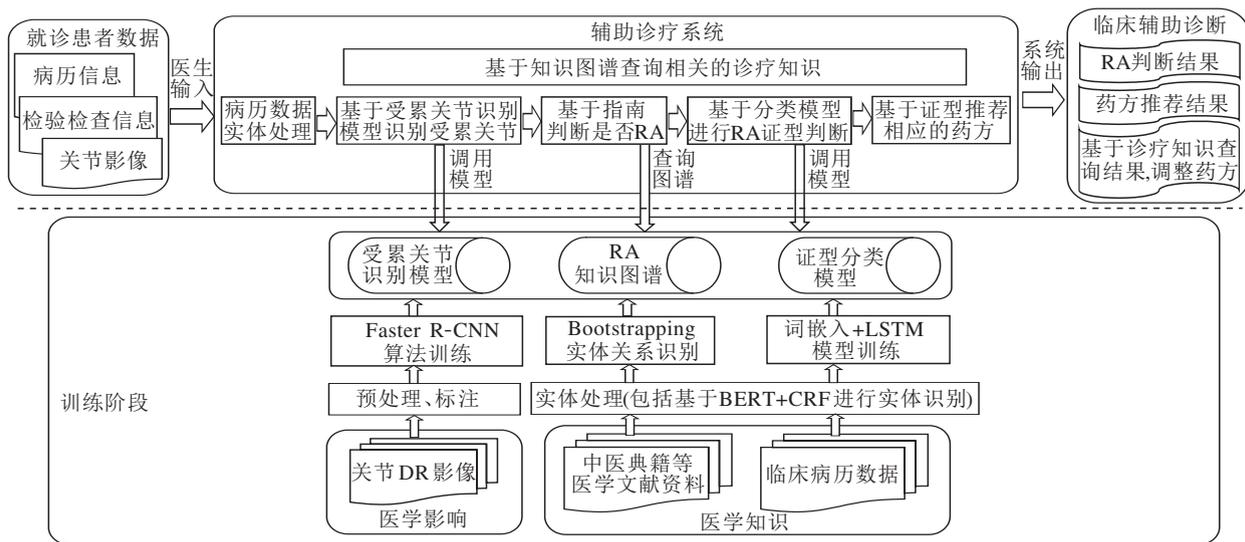


图 1 RA 中医辅助诊疗系统核心功能框图

Fig. 1 Framework of core function of TCM aided diagnosis and treatment system for RA

在诊疗过程中,医生在 RA 诊疗系统中输入就诊患者的检验检查信息、病历文本信息和关节影像数据,诊疗系统将对患者病历数据进行实体处理,根据关节影像判断患者的受累关节,结合影像和病历中的患者信息,基于 RA 的诊疗指南判断患者是否患有 RA.在确诊 RA 后,系统基于分类模型进行 RA 中医证型的判断,并依据证型向医生推荐相应的中医药方.在整个诊疗过程中,RA 知识图谱依据病历中的症状、疾病、检查等信息为医生提供诊疗建议.RA 中医辅助诊疗系统侧重辅助功能,能为医生提供判断依据,提升诊疗效率,方便医生操作.医生复核 RA 诊疗系统的判断是否准确,之后根据病人的其它症状和经验,在给出的推荐药方上进行药品和剂量的调整.

RA 中医诊疗辅助核心步骤如下. 1) 病历实体处理. 2) 基于 Faster R-CNN 识别受累关节. 3) 基于

指南判断患者是否确诊 RA. 4) 基于分类模型判断 RA 证型. 5) 基于证型推荐中医药方. 6) 构建知识图谱.

2 病历实体处理

为了后续的基于分类模型 RA 证型判断、知识图谱构建,首先对病历文本进行实体处理,包括 RA 实体识别和 RA 实体同义词归一化.

为了进行实体识别,首先定义 RA 的实体类型,然后基于迁移学习对实体进行识别.

中医在 2000 多年的发展历史中,形成许多特有的名词术语.例如,中医具有独有的病名(胸痹、关格等)和治法名词(寒者热之、培土生金等^[10]).这些名词术语包含患者大量的信息.经过对 RA 病历

的分析,本文针对 RA 病历规定 3 种命名实体,如表 1 所示.命名实体涵盖从疾病、症状到治法、药方的命名实体,为病历文本的理解起到重要的支持作用.

表 1 RA 实体类型
Table 1 Entity types of RA

类型	说明	举例
疾病	疾病名称	类风湿性关节炎、RA
症状	病人因疾病所引起的主观感觉或客观病态	晨僵、胸闷、全身性畏风寒
检查项	患者的医学检查项及结果	尿酸:147 B2 微球蛋白:0.58

由于 RA 相关领域数据相对较少,针对语料库不充足的问题,本文采用迁移学习^[11]技术,在预训练的中文语言模型基础上,针对 RA 治疗领域,进行微调.其中,预训练的语言模型选用谷歌开源的 BERT (Bidirectional Encoder Representations from Transformers)^[12]中文模型,并将 BERT 的输出向量输入条件随机场(Conditional Random Field, CRF)^[13]分类器,根据标签的属性进行进一步判断.实体提取识别模块框图如图 2 所示.采用这种方法可大幅降低对于 RA 语料库数据量的要求,节约开发时间和成本.

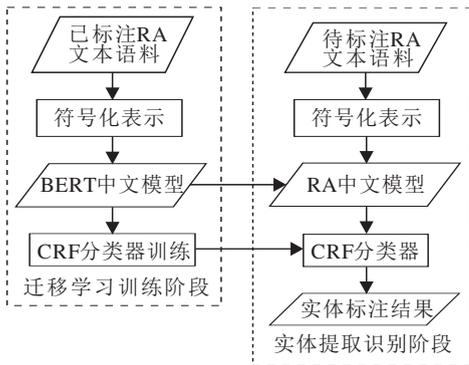


图 2 RA 实体识别流程图

Fig. 2 Flow chart of RA entity recognition

在病历中,每个医生对于实体的描述方式不一定完全相同,但表达的含义一致,例如:症状“苔白”和“舌苔白”描述方式不同,但表达含义相同.对于机器来讲,在数据量有限的情况下,很难根据上下文理解“苔白”和“舌苔白”表达同一含义,需要加大后续机器对于病历的理解能力.因此,针对此类同义词,需要归一化到统一的表达方式,增强机器对病历的理解能力.

由于 RA 病历数据相对较少,本文采用基于规则的方法进行同义词归一化.首先,基于标注数据抽取所有命名实体;然后,对同一类别的实体进行聚类;最后,将聚类结果反馈给医生,医生标注,建立同义词库,并规定每类同义词的标准词.此方法需要在遇到未在同义词库中的新词时,人工加入到同义词库中.另外,当数据量扩充到一定程度时,在采用人工方式工作量太大时,可采用序列到序列的同义词归一化方法,将实体的原词作为源序列,将标准词作为目标序列.

3 基于 Faster R-CNN 识别受累关节

为了减轻医生的工作量,诊疗系统还设置受累关节 DR 影像识别功能:如果医生上传手部、肘部、膝部等关节 DR 影像后,诊疗系统将自动识别受累关节的位置及数量,并以可视化形式将受累关节部位展示在系统界面上,供医生确认.医生可在图形界面上点选,对识别结果进行调整.

受累关节识别模块的处理流程如图 3 所示.图中以区别于其它关节的颜色表示识别的受累关节.

要实现受累关节的识别,首先需要训练能识别受累关节的模型.医生使用标注软件,标注图像中的正常关节和病变关节,使用目标识别算法进行离线学习训练,得出能识别病变关节的模型.RA 诊疗系统使用 Faster R-CNN 训练受累关节识别模型.

为了有效提取关节的视觉特征,系统采用基于 VOC2007 中 20 类常见的物体预训练得到的 Faster R-CNN,并使用标注的关节数据对 Faster R-CNN 进行微调.标注的关节数据必须包含关节在图像中的区域,以及该区域关节是否为病变关节.微调后的模型可识别图像中关节的位置并判断是否病变.

受累关节识别模型的识别效果如图 4 所示.从一幅手部关节 DR 影像中识别关节及预测关节是否为受累关节,0 表示非受累关节,1 表示受累关节,之后的小数表示概率(可信度).

在医生使用系统识别 DR 影像时,辅助诊断系统将影像输入给受累关节识别模型,然后根据模型返回的结果,在界面上的人体关节示意图中相应的关节处展示识别的受累关节.目前还达不到 100% 的准确率,需要医生人工确认,对于识别错误的关节,医生可在界面上进行修改调整.

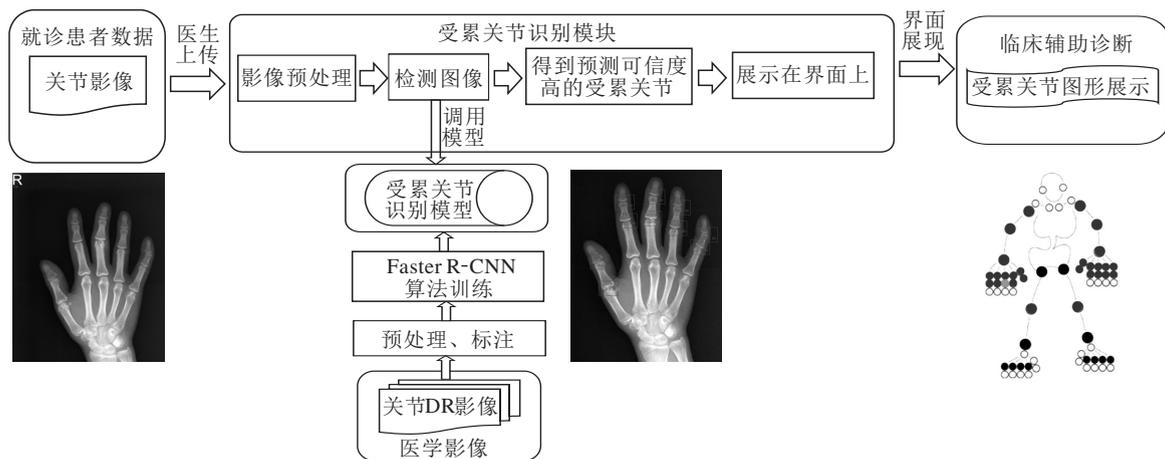


图 3 受累关节识别模块的处理流程图

Fig. 3 Processing flow chart of involved joints recognition module

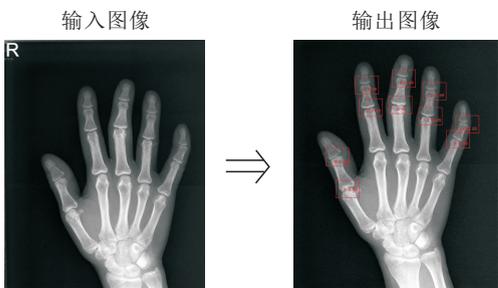


图 4 受累关节识别模型输入输出效果图

Fig. 4 Input and output samples of involved joints recognition model

3)其它与病历文本有关的分数项. 系统从病历文本抽取的实体及相关信息进行判断并计算. 如果识别信息有误差, 医生可在界面上对结果进行调整.

表 2 ACR 1987 年 RA 计算标准^[14]

Table 2 ACR1987 classification criteria for RA^[14]

条件	定义	分数
晨僵	至少 1 小时(≥6 周)	1
3 个(或以上) 关节炎	持续(≥6 周)	1
手关节炎	腕或掌指或近端指间关节肿胀(≥6 周)	1
对称性关节炎	持续(≥6 周)	1
类风湿结节	出现	1
影像学改变	手和腕关节的 X 线有骨质侵蚀或骨质疏松	1
类风湿因子	阳性	1

4 基于指南的类风湿性关节炎判断

基于 1987 年^[14]和 2010 年^[15-16]美国风湿病学会(American College of Rheumatology, ACR)的 RA 诊疗标准判断,将分数项分为 3 类:受累关节数量项、血清学和急性时相反应物等检查相关分数项、其它与病历文本有关的分数项. 对医生输入的患者检验检查信息进行分数计算. ACR 1987 年 RA 计算标准如表 2 所示. ACR 2010 年 RA 计算标准如图 5 所示.

对于这 3 种类别的分数项,系统的处理方式各有不同.

1)受累关节数量项. 根据上传的关节 DR 影像识别的受累关节数量及部位计算. 如果识别有误差, 医生可在界面上对结果进行调整.

2)血清学和急性时相反应物等检查相关分数项. 根据医生在系统中输入的检查项结果判断并计算.

关节受累情况			
受累关节情况	受累关节数	分数	
中大关节	1	0	
	2~10	1	
小关节	1~3	2	
	4~10	3	
至少 1 个为小关节	>10	5	
血清学			
RF或抗CCP抗体均阴性		0	
RF或抗CCP抗体至少 1 项低滴度阳性		2	
RF或抗CCP抗体至少 1 项高滴度(>正常上限 3 倍)阳性		3	
滑膜炎持续时间			
<6 周		0	
≥6 周		1	
急性时相反应物			
CRP或ESR均正常		0	
CRP或ESR增高		1	

图 5 ACR 2010 年 RA 计算标准^[15-16]

Table 5 ACR2010 classification criteria for RA^[15-16]

将上述分数线相加,计算 1987 年及 2010 年标准的得分,如果 2010 年标准计算分数超过 5,或 1987 年标准计算分数超过 3,即可确诊患者患有 RA.

5 基于分类模型的 RA 证型判断

根据 RA 诊疗经验,将 RA 分为如下 9 种中医证型:肺气不利,痰饮停肺;肝肾阴虚,瘀毒阻络;肝郁脾虚,湿毒阻络;脾肾两虚,湿毒阻络;脾肾阳虚,瘀毒内阻;脾虚湿胜,湿毒阻络;脾虚湿胜,痰饮内停;阴阳两虚,湿毒瘀滞;营卫失和,寒湿阻络.

RA 证型诊断的目标是根据医生输入的患者病历文本,对于确诊 RA 的患者,智能判断其属于哪种证型. RA 证型判断模块的处理流程如图 6 所示.

本文采用基于长短期记忆网络 (Long-Short Term Memory, LSTM)^[17] 的分类模型. LSTM 能避免 RNN 的梯度消失的问题,通过门机制利用长时序的信息. 在训练阶段,使用词嵌入技术,将病历文本转化为数字 (转化后的文本将是一个同长度的向量形式 [2,62,31,...]), 形成模型输入. 数据处理后,输入

与标签均转化为模型所能接受的向量输入形式,放入模型进行训练,提前规定好迭代次数,并打印训练准确率,存储模型. 在使用阶段,加载训练好的模型,将新的文本病历通过同样的数据处理流程,输入模型,得到输出结果.

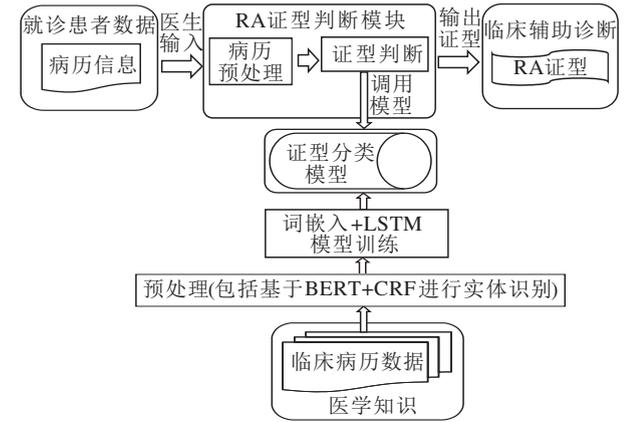


图 6 RA 证型判断模块的处理流程图

Fig. 6 Processing flow chart of classification of syndrome in RA

部分病历文本如图 3 所示.

表 3 部分病历示例

Table 3 Some medical record samples

病历	处方	证型
类风湿关节炎:掌指关节隐痛,项背牵强,肩周酸痛,肢麻,畏风寒,舌暗苔腻,脉细.拟:运脾解毒疏风祛湿治疗.检查:2018-06-19查:抗 ccp:1820, RF:96, ESR:28	厚朴 12g、炒苍术 12g、蜜麻黄 10g、葛根 30g、川芎 20g、威灵仙 15g、蜈蚣 2 条、防己 10g、茯苓 12g、炒银花 15g、细辛 4g、蜜甘草 6g	脾虚湿胜, 湿毒阻络
类风湿关节炎:三月前严重腹泻后出现多关节肿痛,下肢浮肿,胃脘痞满,反酸,有慢性萎缩性胃炎,胃溃疡,外伤致股骨颈股骨折病史,舌暗苔根腻,脉细.拟:运脾解毒通络祛湿治疗.	蜜麻黄 10g、炒苍术 12g、厚朴 12g、金银花 15g、蒲公英 12g、蜈蚣 2 条、炒鸡内金 18g、延胡索 15g、炒薏苡仁 18g、土茯苓 20g、蜜甘草 6g、防己 10g	脾虚湿胜, 湿毒阻络
RA,伴腰椎间盘突出膨出,慢性胃炎,腰骶酸痛,畏风寒,膝关节僵滞,舌质暗红,苔腻,脉沉.拟:运脾益肾,通络祛湿治疗.	延胡索 30g、甘草 6g、炒白术 15g、炮姜 8g、黄芪 20g、牛膝 12g、炒银花 15g、蒲公英 12g、地肤子 15g、徐长卿(后下)12g、酒乌梢蛇 6g、杜仲 12g	脾肾两虚, 湿毒阻络
类风湿性关节炎:确诊近 5 年,多关节疼痛,曾以激素、甲氨蝶呤、艾拉莫德治疗.抗 ccp:467, ESR69, CRP:29,口苦,现关节症状改善,舌质红绛脉细.拟:运脾养阴解毒通络治疗.	当归 10g、生地黄 15g、大青叶 15g、贯众 10g、延胡索 30g、炒白术 20g、丹参 15g、麦冬 20g、玉竹 20g、芡实 20g、赤小豆 20g、蜜甘草 6g	肝肾阴虚, 瘀毒阻络
类风湿性关节炎:病 20 余年,先掌指关节肿大变形,腕关节活动障碍,膝关节反复肿痛,RF:137.9, ESR:30,近日外感咳嗽,舌暗红苔薄,脉沉细.拟:益肾解毒通络祛湿治疗.检查:2019-06-11查:RF:136, CRP:12.9, ESR43.	金银花 20g、蜜甘草 6g、蜈蚣 2 条、炒白术 15g、防己 10g、虎杖 20g、薏苡仁 20g、透骨草 20g、徐长卿(后下)12g、葶草 15g、炒防风 6g、桂枝 10g	脾肾两虚, 湿毒阻络

系统对病历文本进行命名实体识别,除了实体以外的词语采用字. 每个词经过词嵌入之后,进入 LSTM 层,这里 LSTM 是标准的 LSTM. 然后经过一个时间序列得到 t 个隐藏 LSTM 神经单元的向量. 这些向量经过均值池化 (Mean Pooling) 层后,得到向量 h . 紧接着是一个简单的 logistic 回归层或一个

softmax 层,得到一个类别分布向量,即判断患者的证型.

6 基于证型的中医药方推荐

基于 RA 中医专家通过多年临床经验给出的建

议及整理,通过如下步骤得到适合患者证型的药方,在系统中推荐给诊断医生.获得推荐药方的步骤如下.

1) 根据患者证型,先得到该证型对应的基本药方(药物及剂量、煎煮方法),在其基础上,根据后面步骤的条件进行相应调整.

2) 根据从第 2 节中病历数据实体处理后获得的患者症状实体,将症状分别组合,得到组合后的症状列表,判断是否在症状组合药物库(由医生整理的不同症状组合情况对应的常用药物库)中,并将药物和相应剂量加入 1) 的基本药方中.症状组合药物库部分数据示例如表 4 所示.

表 4 症状组合药物库部分数据示例

Table 4 Samples of symptom combination and drug library

症状组合	治疗药物
全身性发热、不恶寒	生石膏 30g(先煎)、知母 20g
全身性发热、恶寒、胸闷、咳嗽	炙麻黄 9g、杏仁 10g、生石 20g(先煎)
五心烦热	龟板 15g(先煎)、熟地黄 15g、枸杞子 12g
全身性畏风寒、肢冷	淡附片 8g(先煎)、桂枝 12g、细辛 3g
局部红肿、热痛	炒黄柏 10g、忍冬藤 30g
动则汗出、乏力	生黄芪 15g、炒党参 15g
夜寐汗出、潮热	熟地 18g、五味子 10g
头晕、神倦、乏力	生黄芪 18g、当归 15g
头晕、肢麻	葛根 30g、桃仁 10g、红花 10g
头顶痛、干呕、吐涎沫	吴茱萸 3g、炒党参 15g、生姜 9g、红枣 10g

3) 第 1 节中病历数据实体处理后获得患者检查结果实体.当检查结果异常时,从检查药物库(由医生整理)中获取对应药物和相应剂量,加入药方.检查药物库部分数据示例如表 5 所示.

表 5 检查药物库部分数据示例

Table 5 Samples of medical examination and drug library

检查项异常	治疗药物
尿蛋白阳性	芡实 15g、黄芪 15g、泽兰 12g
尿白细胞阳性	瞿麦 12g、车前草 10g
尿红细胞阳性	小蓟 15g、墨旱莲 15g、仙鹤草 15g
肝功能异常(AST、ALT)	垂盆草 15g
血压偏高	天麻 9g、钩藤 15g(后下)、石决明 15g

4) 对于仍没有匹配到药物的症状,根据症状药物表(由医学典籍抽出整理而成的数据)给出相应

的药物和剂量.如果症状对应多个药物,优先选择能治疗患者当前表现多个症状的药物.

5) 对 4) 得出的药物与现有药方进行药物冲突检测(“十八反十九畏”规则),若存在冲突,使用具备相同效用的其它药物替换冲突的药物.

6) 得到药方.对于还没有明确剂量的药物,根据药物常用剂量(由医学典籍抽出整理而成的数据),给出推荐药物的常规剂量,形成推荐药方推荐给诊断医生.

7 知识图谱的构建方法

知识图谱(Knowledge Graph)是显示知识发展进程与结构关系的一系列各种不同的图形,用可视化技术描述知识资源及其载体,挖掘、分析、构建、绘制和显示知识及它们之间的相互联系.

知识图谱功能与辅助诊疗结合使用,医生在诊治时可使用它查找信息,了解证型和药方、药物的信息.将 RA 疾病病情、治疗信息、典型病历等形成知识图谱,供基层医生在日常学习和实际诊断中参考,有助于医生提高诊断水平,减少误诊.知识图谱能增强中医药知识资源的联通性,支持中医用户在概念层次上浏览领域知识资源,发现中医药概念或知识资源之间的潜在联系^[18].

本文的 RA 中医知识图谱的构建主要包括知识图谱设计、实体和实体关系抽取及知识图谱存储 3 个步骤,具体流程图如图 7 所示.

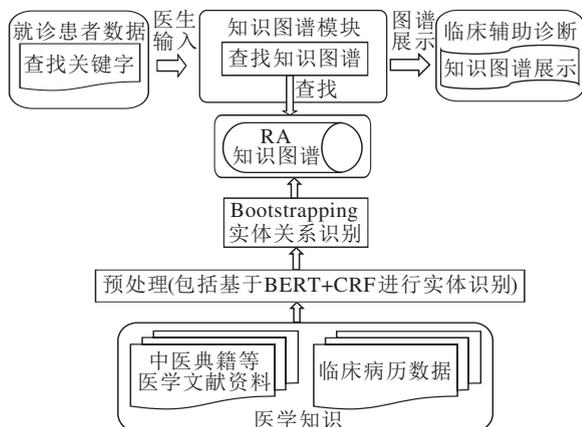


图 7 知识图谱的构建及查询流程图

Fig. 7 Construction and query flow of knowledge graph

7.1 RA 知识图谱设计

RA 知识图谱设计需要明确知识图谱的数据来

源,定义与 RA 相关的实体及实体之间的关系. 本文收集 RA 中西医的诊疗指南、专家病案、中国药典等资料作为知识图谱的数据来源,并定义如下的实体和实体关系.

实体包括与疾病有关的实体和治疗相关的实体.

疾病相关的实体有:1) 疾病;2) 疾病属性,包括特点、原因、部位、检查方法及指标值、临床表现等;3) 证型;4) 症状,包括主症、次症等.

治疗相关的实体有:1) 治法;2) 药方及中成药;3) 用药注意,包括用药频率、用药方法、禁忌、不良反应等;4) 药方属性,包括效果、优点、缺点、推荐等级等;5) 药方组成;6) 药方剂量.

实体关系有如下 4 种.

1) 具有关系. 疾病具有疾病属性(包括特点、原因、部位、检查方法、临床表现等)、药方(及中成药)具有用药注意事项(包括用药频率、用药方法、禁忌、不良反应等)、药方(及中成药)具有药方属性(包括效果、优点、缺点、推荐等级等).

2) 包含关系. 疾病包含证型,证型包含症状(包括主症、次症等),药方(及中成药)包含药方组成.

3) 治疗关系. 治法治疗证型、药方(及中成药)治疗证型.

4) 对应关系. 药方组成对应药方剂量.

7.2 RA 知识图谱实体及实体关系抽取

本文中 RA 知识图谱的实体抽取方法与 RA 实体识别的方法一致,采用 BERT-CRF 对实体进行抽取.

实体关系抽取采用 Bootstrapping 方法,可减轻标注的工作量,适应于小规模数据的关系抽取. Bootstrapping 是一种再抽样的统计方法,根据 Efron 提出的统计理论,从已知数据中得到的分布是对总体分布最准确的预测.因此在标注样本较少、非标注样本较多时,通过反复抽样迭代扩充标注样本,数据分布会不断接近真实分布.

使用 Bootstrapping 抽取实体间的关系后,根据预设计的实体关系分类,将其归类,得到实体关系三元组.

7.3 RA 知识图谱存储

本文构建的 RA 知识图谱存储在图数据库 Neo4j 中. Neo4j 通过 Cypher 提供多种方式的搜索及排序,如查询疾病的症状等,同时提供多种图算法,为应用知识图谱提供便利.

将抽取的实体-关系-实体三元组保存到 csv 等格式的文件,通过 Neo4j 的 Cypher load 命令导入 Neo4j 数据库中.

通过 Match 语句查询存储在 Neo4j 中的数据,Neo4j 将以图形展示数据.

8 系统展示

应用第 2 节~第 6 节技术,初步实现 RA 的中医辅助诊疗系统,实现第 1 节提到的功能.目前正在由合作医生内部试用,根据反馈意见进行改进.病历输入界面如图 8 所示,病历智能诊断界面如图 9 所示.

病人信息	病历	RA 计算	病历分析	智能诊断	新增病例	当前病历: 2020-05-22	状态: 未诊断																				
现病史	类风湿关节炎, 甲状腺癌术后: 右膝关节肿痛, 右肘关节酸痛, 左踝偶发疼痛, 胃纳一般, 恶心, 眠可, 腹泻呈蛋花样, 小便可, 肢冷, 月经延期, 量少, 舌质暗红苔腻, 脉沉细, 全身性发热, 不恶寒。																										
既往史																											
过敏史																											
家庭史																											
体格检查	输入内容可以进行匹配	治则治法																									
病程		中医四诊																									
辨证依据		建议																									
辅助检查	<table border="1"> <tr> <td>检查项</td> <td>尿蛋白</td> <td>200</td> <td>mg/24 Hours</td> <td>移除</td> </tr> <tr> <td>检查项</td> <td></td> <td>0</td> <td>g/t</td> <td>移除</td> </tr> <tr> <td>检查项</td> <td>血红蛋白</td> <td colspan="3">添加</td> </tr> <tr> <td>检查项</td> <td>血糖</td> <td colspan="3"></td> </tr> </table>	检查项	尿蛋白	200	mg/24 Hours	移除	检查项		0	g/t	移除	检查项	血红蛋白	添加			检查项	血糖									
检查项	尿蛋白	200	mg/24 Hours	移除																							
检查项		0	g/t	移除																							
检查项	血红蛋白	添加																									
检查项	血糖																										

图 8 RA 中医辅助诊疗系统病历输入界面

- cine Prescriptions from Symptoms with an End-to-End Approach // Proc of the CCF International Conference on Natural Language Processing and Chinese Computing. Berlin, Germany: Springer, 2019: 486–498.
- [10] 赵移畛, 尚东挺. 中医电子病历系统应具备的中医特点. 中医药管理杂志, 2007, 15(12): 894–896.
(ZHAO Y Z, SHAO D T. Characteristics of TCM That Should Be Possessed by TCM Electronic Medical Record System. Chinese Journal of Management in Chinese Medicine, 2007, 15(12): 894–896.)
- [11] PAN S J, YANG Q. A Survey on Transfer Learning. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345–1359.
- [12] DEVLIN J, CHANG M W, LEE K, *et al.* BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // Proc of NAACL-HLT 2019. Stroudsburg, USA: ACL, 2018: 4171–4186.
- [13] LAFFERTY J, MCCALLUM A, PEREIRA F C N. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data // Proc of the 18th International Conference on Machine Learning. New York, USA: ACM, 2001: 282–289.
- [14] ARNETT F C, EDWORTHY S M, BLOCH D A, *et al.* The American Rheumatism Association 1987 Revised Criteria for the Classification of Rheumatoid Arthritis. Arthritis and Rheumatism, 1988, 31(3): 315–324.
- [15] ALETAHA D, NEOGI T, SILMAN A J, *et al.* 2010 Rheumatoid Arthritis Classification Criteria: An American College of Rheumatology/European League Against Rheumatism Collaborative Initiative. Annals of the Rheumatic Diseases, 2010, 69(9): 1580–1588.
- [16] 中华医学会风湿病学分会. 类风湿关节炎诊断及治疗指南. 中华风湿病学杂志, 2010, 14(4): 265–270.
(Rheumatology Branch of Chinese Medical Association. Guidelines for the Diagnosis and Treatment of Rheumatoid Arthritis. Chinese Journal of Rheumatology, 2010, 14(4): 265–270.)
- [17] HOCHREITER S, SCHMIDHUBER J. Long Short-Term Memory. Neural Computation, 1997, 9(8): 1735–1780.

- [18] 于彤, 刘静, 贾李蓉, 等. 大型中医药知识图谱构建研究. 中国数字医学, 2015, 10(3): 80–82.
(YU T, LIU J, JIA L R, *et al.* Research on the Construction of Big Knowledge Graph for Traditional Chinese Medicine. China Digital Medicine, 2015, 10(3): 80–82.)

作者简介



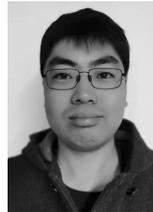
孙明俊(通信作者), 硕士, 高级工程师, 主要研究方向为人工智能、多媒体. E-mail: sunmingjun@caict.ac.cn.

(**SUN Mingjun** (Corresponding author), master, senior engineer. Her research interests include artificial intelligence and multimedia.)



张丹, 硕士, 工程师, 主要研究方向为自然语言处理. E-mail: zhangdan@ngai.ac.cn.

(**ZHANG Dan**, master, engineer. Her research interests include natural language processing.)



郑明智, 学士, 工程师, 主要研究方向为自然语言处理、智慧医疗. E-mail: zhengmingzhi@ngai.ac.cn.

(**ZHENG Mingzhi**, bachelor, engineer. His research interests include natural language processing and wise medical.)



梅舒欢, 硕士, 工程师, 主要研究方向为图像检索、图像识别. E-mail: meishuhuan@ngai.ac.cn.

(**MEI Shuhuan**, master, engineer. His research interests include image retrieval and image recognition.)