

本文引用: 韦昌法, 罗丽琴, 晏峻峰. 中医数字辨证配套医案智能采集与分析系统构建研究[J]. 湖南中医药大学学报, 2020, 40(1): 70-74.

# 中医数字辨证配套医案智能采集与分析系统构建研究

韦昌法<sup>1</sup>, 罗丽琴<sup>2</sup>, 晏峻峰<sup>1\*</sup>

(1. 湖南中医药大学信息科学与工程学院, 湖南 长沙 410208; 2. 湖南大学信息科学与工程学院, 湖南 长沙 410082)

**[摘要]** 中医数字辨证是中医智能辅助诊断的核心, 辨证模型的质量直接影响辨证的准确度和效率。为了建立中医数字辨证模型并验证其准确度, 需要准备大量的配套医案。根据中医数字辨证模型对医案的需求, 设计了中医数字辨证配套医案智能采集与分析系统, 利用人工智能技术和 Python 编程技术实现了医案采集、医案展示、医案检索、数据统计分析及可视化等功能, 有助于快速采集并筛选出符合条件的中医医案, 提高中医数字辨证模型的创建效率和质量。

**[关键词]** 中医数字辨证; 人工智能; 医案采集; 医案分析

**[中图分类号]** R2-03      **[文献标志码]** A      **[文章编号]** doi:10.3969/j.issn.1674-070X.2020.01.016

## Research on Construction of Intelligent Acquisition and Analysis System of Supporting Medical Cases for Traditional Chinese Medicine Digital Syndrome Differentiation

WEI Changfa<sup>1</sup>, LUO Liqin<sup>2</sup>, YAN Junfeng<sup>1\*</sup>

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

2. College of Computer Science and Electronic Engineering, Hunan University, Changsha, Hunan 410082, China)

**[Abstract]** The digital syndrome differentiation of TCM is the core of intelligent assistant diagnosis of TCM. The quality of syndrome differentiation model directly affects the accuracy and efficiency of syndrome differentiation. In order to establish the digital syndrome differentiation model of TCM and verify its accuracy, a large number of supporting medical cases need to be prepared. According to the requirement of TCM digital syndrome differentiation model for medical cases, an intelligent acquisition and analysis system of supporting medical cases for TCM digital syndrome differentiation was designed. Using artificial intelligence technology and Python programming technology, the functions of medical cases collection, medical cases display, medical cases retrieval, data statistical analysis and visualization were realized, which are helpful to quickly collect and screen qualified TCM medical cases, and to improve the efficiency and quality of the establishment of TCM digital syndrome differentiation model.

**[Keywords]** traditional Chinese medicine digital syndrome differentiation; artificial intelligence; medical cases acquisition; medical cases analysis

辨证论治是中医学认识疾病和治疗疾病的基本原则。中医数字辨证指的是在研究中医辨证论治规律的基础上, 建立辨证过程的数字模型, 通过计算机

程序对患者病情资料进行分析判断, 最终得出辨证结果<sup>[1-3]</sup>。国学大师章太炎先生曾说: “中医之成绩, 医案最著, 欲求前人之经验心得, 医案最有线索可

**[收稿日期]** 2019-09-06

**[基金项目]** 教育部科技发展中心“天诚汇智”科研创新基金课题(2018A03011); 湖南省教育厅资助科研项目(18A219); 湖南中医药大学中医学一流学科开放基金资助项目(2018ZYX65); 湖南中医药大学信息科学与工程学院电子科学与技术学科开放基金项目(2018-1)。

**[作者简介]** 韦昌法, 男, 在读博士研究生, 副教授, 主要研究方向为中医数字辨证。

**[通讯作者]** \* 晏峻峰, 女, 博士, 教授, 博士研究生导师, E-mail: teacheryan@qq.com。

寻,循此钻研,事半功倍。”<sup>[4]</sup>为了建立中医数字辨证模型并验证其准确度,往往需要准备大量的配套医案,构建中医数字辨证配套医案智能采集与分析系统,有助于快速采集满足需求的中医医案。

### 1 中医数字辨证模型对医案的需求

在辨证论治中,辨证是论治(立法、处方、用药)的前提,辨证准确与否,是临床疗效的关键<sup>[5]</sup>。中医数字辨证是中医智能辅助诊断的核心,辨证模型在系统中起着关键性的作用,模型的质量直接影响辨证的准确度和效率。因此,辨证模型也成为基于人工智能的中医诊断研究者普遍关注的热点问题。

我国的研究者从上世纪 70 年代中后期就开展了以中医专家系统为代表的中医智能辅助诊疗系统的研发,先后研发出关幼波中医肝病诊断处方专家系统、中医数字辨证机等一大批中医专家系统。40 多年来,许多人工智能方法被应用于构建中医数字辨证模型,不同的辨证模型采用了不同的知识表示和知识推理方法,它们对医案的需求也不尽相同。中医医案是医生对所诊治病证的发生、发展及其转归的真实记录,是对疾病辨证分析、治疗过程的真实记载,对医案进行研究和分析有助于挖掘中医专家的诊疗经验<sup>[6]</sup>。中医数字辨证模型对医案的需求主要集中在两大方面,一方面需要从医案中挖掘出中医专家的辨证经验,将这些经验嵌入到中医数字辨证模型中,使数字辨证模型能集成中医专家的经验来开展辨证;另一方面,在验证数字辨证模型的准确率时,也需要大量的医案来评估数字辨证模型的辨证结果与中医专家的辨证结果的一致程度。

中医数辨证配套医案的来源有很多,其中一个重要的来源是中医名家的医案集,这些医案集中的医案质量较高,在中医数字辨证模型的创建及验证过程中具有重要的作用。因此,笔者开展了面向医案集的中医医案智能采集与分析系统构建研究,以期中医数字辨证研究提供快速获取高质量配套医案的工具。

### 2 中医数字辨证配套医案智能采集与分析系统的设计

中医数字辨证配套医案智能采集与分析系统拟

利用光学字符识别(optical character recognition,OCR)技术将医案集的医案图片转换成医案文本,存储到医案数据库中,进而对医案文本信息进行自动校验和结构化处理,最后对医案数据进行统计分析及可视化处理,并通过 web 页面将医案数据和相应的统计分析及可视化结果予以直观展示,以供用户查看、筛选和导出医案。

中医数字辨证配套医案智能采集与分析系统可分为 4 个主要模块:医案采集模块、医案展示模块、医案检索模块、数据统计分析及可视化模块,系统功能模块图如图 1 所示,系统业务流程图如图 2 所示。

医案采集模块利用 OCR 技术将医案集的医案图片转换成医案文本,并利用人工智能分词技术和模糊匹配技术实现对医案数据进行自动校验,以确保通过 OCR 技术转换出来的医案文本内容与原始的医案图片内容是一致的。医案采集模块利用正则表达式匹配技术对医案文本信息进行结构化处理,得到医案基本信息:患者性别、年龄、初诊信息、辨证结果、治法、方药、其他诊次信息等内容,并存储到医案数据库中。

医案展示模块将已采集的医案数据展示出来,

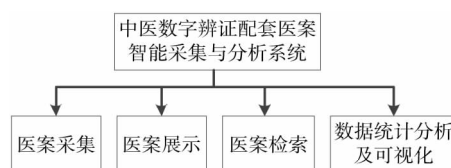


图 1 系统功能模块图

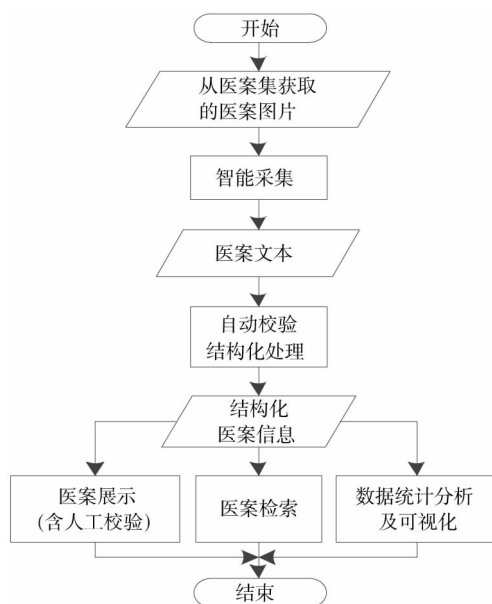


图 2 系统业务流程图

编号	姓名	年龄	性别	初诊	辨证结果	治法信息	方药信息	医生
824	秦某	18	女	患者于数年前因受精神刺激,	情志不遂;肝郁抑脾;	养心安神定志;滋阴清热;	炙甘草,9g;浮小麦,30g;	五谢海洲
825	王某	42	女	患者因恐惧和精神高度紧张,	惊恐气乱;焦虑伤神;	健脾补肾;理气安神;	熟地,20g;阿胶(烊化),20g;	王文彦
826	关某	24	女	3年前因失恋;精神受刺,后渐	气郁血虚;	养心安神;	甘麦大枣汤加味;小麦60g;	王文彦
827	李某	65	女	2年前患神经强迫症经张老治	五志过极;肝郁化火;	清泻肝火;涤痰安神;	川芎,15g;苍术,15g;香附,	张琪
828	张某	40	男	患者自去年12月起时感胸闷,	情志失畅;肝失条达;	疏肝理气;健脾宁心;	川连,3g;川厚朴,12g;郁金	杨继芬
829	肖某	40	男	患者受精神刺激后失眠10余	痰湿阻滞;肝气郁结;	理气化痰解郁;	方用温胆汤加味。竹茹、	邓铁涛
830	贺某	56	女	近一年来经常出现畏寒微风,	冲任失调;	调补冲任;	二仙汤加减。仙茅,20g;	陈可冀
831	杨某	52	女	患者既往有抑郁病史,在胡老	心气不足;肝气不舒;	疏肝解郁;化痰定志;养心	自拟加味甘麦大枣汤、	胡建华
832	方某	52	女	胃脘觉冷5月余,因情志不畅,	肝郁不达;胃阳不振;	开郁行气;宁心安神;调和	解郁合欢汤化裁。合欢皮	徐景藩
833	张某	30	女	患者2年前患皮肤湿疹,因病	至肝气郁结;	疏肝解郁;清心安神;	炙甘草,18g;桂枝,18g;麦	袁沛然

图3 系统采集到的郁病医案信息

用户在浏览医案数据时,可以对医案数据进行人工校验,比对医案数据与原始医案图片是否一致。因为医案采集模块已提供了医案数据自动校验功能,故人工校验功能不是必须的,但是它可以增强用户对系统的认可度和信任度。

医案检索模块负责根据用户的需要对医案数据进行检索,用户可以在性别、年龄段、辨证结果、治法信息和方药信息等字段中输入要检索的关键词,系统根据用户输入的条件对数据库中的医案数据进行精确或模糊检索,并将检索结果直观展示出来。

数据统计分析及可视化模块负责对已采集的医案数据进行统计分析,生成相应的统计图表和词云图,使用户可以更直观地了解数据库中的医案数据特征,为他们的科研和临床工作提供参考。系统根据用户的统计需求,从医案数据库中读取相应的数据,进行数据可视化相关的参数设置,最终生成相应的统计图表。系统还可以在数据统计分析的基础上,对辨证结果、治法信息和方药信息等内容生成词云图,让用户更直观地了解统计分析结果。

### 3 中医数字辨证配套医案智能采集与分析系统的实现

课题组研究团队采用 Python 编程语言开发实现了中医数字辨证配套医案智能采集与分析系统,并以郁病医案为例,开展了医案智能采集与分析工作,成功采集了一大批郁病医案的基本信息:患者性别和年龄信息、初诊信息、辨证结果、治法信息、方药信息、其他诊次信息等内容,并从辨证结果、治法信息、方药信息、性别、年龄、医生和医案来源七个角度对所采集的医案数据进行统计分析,得出相应的统计分析图表和词云图,使用户可以更直观地了解医

案数据特征,以便筛选出符合要求的医案来开展中医数字辨证研究。

#### 3.1 医案采集模块的实现

医案智能采集与分析系统所采集的医案可能来源于不同的医案集,这些医案集对医案信息的组织结构不尽相同,为了支持从不同的医案集中采集医案,系统引入了医案采集模板的概念,用户可以为不同的医案集建立相应的医案采集模板。进行医案采集时,系统根据用户指定的医案采集模板将医案集的医案图片转换成医案文本,进而对医案数据进行自动校验和结构化处理,并将医案信息存储到医案数据库中。系统采集到的医案信息如图3所示,医案信息中的辨证结果、治法信息和方药信息如图4所示。

编号	辨证结果	编号	治法信息	编号	方药信息
222	情志不遂	202	养心安神定志	816	炙甘草
223	肝郁抑脾	203	滋阴清热除烦	817	浮小麦
224	心气耗伤	204	健脾补肾	818	五味子
225	阴虚内热	205	理气安神	819	枳龙骨
226	惊恐气乱	206	养心安神	820	生牡蛎
227	焦虑伤神	207	清泻肝火	821	首乌藤
228	气郁血虚	208	涤痰安神	822	莲子心
229	五志过极	209	补肝理气	823	百合
230	肝郁化火	210	健脾宁心	824	生地黄
231	情志失畅	211	理气化痰解郁	825	知母

图4 医案信息中的辨证结果、治法信息和方药信息

#### 3.2 医案展示模块的实现

医案展示模块将医案数据库中的医案数据展示出来,用户可以分页查看所有的医案,医案信息包括患者基本信息、初诊信息、辨证结果、治法信息、方药信息、医生和医案来源等内容,医案展示界面如图5所示。在浏览医案信息时,用户还可以对医案数据进行人工校验,校验界面如图6所示,该界面的左半部分显示了医案图片,右半部分显示了系统采集并结构化处理后得到的医案信息,如果用户在校验过程

中发现医案信息与图片中的信息不一致,可以即时修正。如前所述,因为医案采集模块已提供了医案数据自动校验功能,故人工校验功能不是必须的,但是它提供了一个备用的校验渠道,使得系统更加完善。

### 3.3 医案检索模块的实现

医案检索模块根据患者性别、辨证结果、治法信息、方药信息、医生信息等检索条件对医案数据进行检索,将用户所需的医案筛选出来予以直观展示。例如,可以在医案数据中检索辨证结果包含“肝气郁结”的医案,其结果如图 7 所示。

### 3.4 医案数据统计分析及可视化模块的实现

数据统计分析及可视化模块根据用户的需求对医案数据进行统计分析,生成相应的统计图表和词云图,使用户可以更直观地了解医案数据特征,以便

筛选出符合要求的医案来开展中医数字辨证研究。该模块可以细分为辨证结果统计及可视化模块、治法信息统计及可视化模块、方药信息统计及可视化模块、性别和年龄统计及可视化模块、医生和医案来源统计及可视化模块,部分实现效果如图 8、图 9 和图 10 所示。

## 4 结论

为了建立中医数字辨证模型并验证其准确度,需要准备大量的配套医案。本文分析了中医数字辨证模型对医案的需求,设计并实现了中医数字辨证配套医案智能采集与分析系统,并以郁病医案为例开展了医案智能采集与分析工作,从辨证结果、治法信息、方药信息、性别、年龄、医生和医案来源七个角

序号	个人信息	年龄	性别	初诊	辨证结果	治法信息	方药信息	医生	医案来源	是否校验
1	秦某	18	女	患者于数年前因受精神刺激,一直精神恍惚,影响学习和工作,良以为苦。曾服用西药安定剂多种,不愈。1987年9月25日诊查:精神恍惚,纳差,心烦不安,手心出汗,二便尚可,月经错后,时而淋漓不断,舌红胖嫩,苔薄黄,脉沉弦。	情志不遂; 肝郁抑脾; 心气耗伤; 阴虚内热	养心安神 定志; 滋阴 清热 除烦	炙甘草, 9g; 浮小麦, 30g; 五味子, 9g; 煨龙骨, 24g; 煨牡蛎, 24g; 首乌藤, 18g; 莲子心, 3g; 百合, 12g; 生地黄, 15g; 知母, 9g; 桂枝, 10g; 白芍, 18g。水煎服, 10剂。	谢海洲	专病专科 中医古今 证治通览 丛书-郁证	已校验
2	王某	42	女	患者因恐惧和精神高度紧张7昼夜未眠,其后即不能入睡。1999年10月20日诊查:两目干涩,周身酸楚,腰酸痛,乏力倦怠,精神不能集中,健忘,心烦,不思饮食,舌淡红,苔白,脉弱。	惊恐气乱; 焦虑伤神	健脾 补肾; 理气 安神	熟地黄, 20g; 阿胶(烊化), 20g; 香附, 15g; 陈皮, 15g; 远志, 15g; 合欢花, 15g; 半夏, 15g; 茯苓, 20g; 甘草, 15g; 黄连, 7.5g; 鸡蛋黄, 2个; 枳壳, 15g; 竹茹, 15g。3剂, 日1剂, 水煎, 分2次口服。	王文彦	专病专科 中医古今 证治通览 丛书-郁证	未校验
3	关某	24	女	3年前因失恋,精神受刺激,后渐觉头胀失眠,历2年不解,近2个月加剧。1982年4月9日诊查:头胀欲裂,昼夜不眠,自觉神志较前模糊,伴面苍白食少,间有干呕,气短乏力,懒于言语,验其舌质淡,苔薄,候其脉弦细而寸弱甚。	气郁血虚	养心安神	甘麦大枣汤加味:小麦, 60g, 甘草, 10g; 大枣, 5枚; 百合(先煎), 30g; 炒白芍, 20g; 当归, 20g; 珍珠母, 20g; 薄荷, 10g; 陈皮, 15g; 竹茹, 15g; 炒酸枣仁, 30g; 柏子仁, 20g。3剂, 水煎服。	王文彦	专病专科 中医古今 证治通览 丛书-郁证	已校验

图 5 医案展示界面

<p><b>医案图片:</b></p> <p>病例一: 秦某, 女, 18岁。</p> <p>初诊: 患者于数年前因受精神刺激, 一直精神恍惚, 影响学习和工作, 良以为苦。曾服用西药安定剂多种, 不愈。1987年9月25日诊查: 精神恍惚, 纳差, 心烦不安, 手心出汗, 二便尚可, 月经错后, 时而淋漓不断, 舌红胖嫩, 苔薄黄, 脉沉弦。</p> <p>[辨证] 情志不遂, 肝郁抑脾, 心气耗伤, 阴虚内热。</p> <p>[治法] 养心安神定志, 滋阴清热除烦。</p> <p>[方药] 炙甘草 9g, 浮小麦 30g, 五味子 9g, 煨龙骨 24g, 煨牡蛎 24g, 首乌藤 18g, 莲子心 3g, 百合 12g, 生地黄 15g, 知母 9g, 桂枝 10g, 白芍 18g。水煎服, 10剂。</p>	<p><b>医案编号:</b></p> <p>101</p> <p><b>个人信息:</b></p> <p>秦某</p> <p><b>性别信息:</b></p> <p>女</p> <p><b>年龄信息:</b></p> <p>18</p> <p><b>初诊信息:</b></p> <p>患者于数年前因受精神刺激, 一直精神恍惚, 影响学习和工作, 良以为苦。曾服用西药安定剂多种, 不愈。1987年9月25日诊查: 精神恍惚, 纳差, 心烦不安, 手心出汗, 二便尚可, 月经错后, 时而淋漓不断, 舌红胖嫩, 苔薄黄, 脉沉弦。</p> <p><b>辨证结果:</b></p> <p>情志不遂; 肝郁抑脾; 心气耗伤; 阴虚内热</p> <p><b>治法信息:</b></p> <p>养心安神定志; 滋阴清热除烦</p> <p><b>方药信息:</b></p> <p>炙甘草, 9g; 浮小麦, 30g; 五味子, 9g; 煨龙骨, 24g; 煨牡蛎, 24g; 首乌藤, 18g; 莲子心, 3g; 百合, 12g; 生地黄, 15g; 知母, 9g; 桂枝, 10g; 白芍, 18g。水煎服, 10剂。</p>
--	---

图 6 医案数据人工校验界面

序号	个人信息	年龄	性别	初诊	辨证结果	治法信息	方药信息	医生	医案来源	是否校验
1	肖某	40	男	患者受精神刺激后失眠10余年,长期服用中西药治疗,效果不佳。1999年4月2日诊查:失眠,不能入睡,伴头晕,胸闷,记忆力差,四肢疲乏,纳食一般,舌淡红,苔黄稍浊,脉弦滑。各项理化检查无异常发现,血压正常,既往有“精神分裂症”病史。	痰湿阻滞; 肝气郁结	理气 化痰 解郁	方用温胆汤加味。竹茹、半夏、胆南星、素馨花各10g,枳壳、橘红、甘草各6g,茯苓、白术各15g;杜仲12g。14剂,每天1剂,水煎服,复渣再煎晚上服。	邓铁涛	专病专科 中医古今 证治通览 丛书-郁证	未校验
2	张某	30	女	患者2年前患皮肤湿疹,因病致精神紧张、失眠,当地医生诊为抑郁症,服药3个月好转(具体用药不详)。停药后6个月复发,继服抗抑郁药6个月。2005年12月又出现抑郁症状。2006年1月26日诊查:心悸,胸闷,精神易紧张,情绪低落,失眠严重,经中医治疗有所好转,但也只能入睡4~5小时,并时有眩晕,乏力,纳食不馨,月经愆期,40~50天经行1次,大便正常。舌边尖红,苔薄,脉细少力。	肝气郁结	疏肝 解郁; 清心 安神	炙甘草,18g;桂枝,18g;麦冬,18g;西红花(后下),1g;黄连,9g;生地,30g;龙骨(先煎),30g;龙齿(先煎),30g;生牡蛎(先煎),30g;常山,9g;茯苓,12g;茯神,12g;郁金,15g;党参,18g;生姜,6g;大枣,7枚。7剂,水煎服,日1剂。	袁沛然	专病专科 中医古今 证治通览 丛书-郁证	未校验

图7 在医案数据中检索出辨证结果包含“肝气郁结”的医案



图8 郁病医案辨证结果词云图

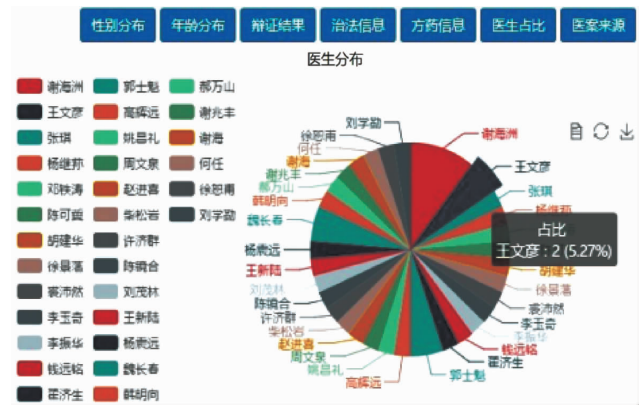


图10 郁病医案所属医生分布图

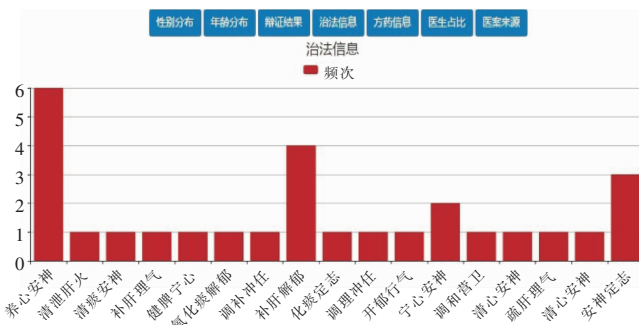


图9 郁病医案治法信息统计柱状图

度对所采集的医案数据进行了统计分析,得出相应的统计分析图表和词云图,使用户可以更直观地了解医案数据特征,有助于快速采集并筛选出符合要求的医案来开展中医数字辨证研究,提高中医数字

辨证模型的创建效率和质量。

参考文献

- [1] 朱文锋.《中医数字辨证机》研究技术报告(医理部分)[J].湖南中医药大学学报,1980(1):1-11.
- [2] 韦昌法,晏峻峰.从知识表示与推理方法探讨中医数字辨证发展[J].中华中医药杂志,2019,34(10):4471-4473.
- [3] 黄惠勇.证素辨证与数字中医药[J].湖南中医药大学学报,2012,32(11):3-6.
- [4] 盛增秀,陈永灿.借古鉴今写好医案[J].浙江中医杂志,2018,53(7):469-471.
- [5] 朱文锋.证素辨证学[M].北京:人民卫生出版社,2008:1.
- [6] 张家玮,鲁兆麟.也谈中医医案研究的价值与意义[J].国际中医药杂志,2011,33(1):63-65.

(本文编辑 李路丹)