

DOI: 10.13288/j.11-2166/r.2020.03.006

思路与方法

基于人工智能算法的中医状态辨识规则

徐佳君^{1,2}, 罗志明³, 赵文^{1,2}, 雷黄伟^{1,2}, 周常恩^{1,2}, 李绍滋³, 李灿东^{1,2*}

1. 福建中医药大学中医学院, 福建省福州市闽侯县上街镇华佗路 1 号, 350122; 2. 福建省中医健康状态辨识重点实验室; 3. 厦门大学信息学院人工智能系

[摘要] 中医诊断方法技术的客观化研究是实现中医药现代化、智能化发展的重要途径和必经阶段。证素辨证学中病位、病性、“证候-证素”积分等概念的提出, 促进了中医诊断学走向数字化发展。状态辨识理论的出现, 进一步完善了对人体状态要素的客观、全面、动态的中医评估体系, 使得中医诊断学逐步走向客观化。提出围绕要素阈值、一(少)元论、伴随表征的共同指向、三因制宜、重视“辨病”、重视“主症”“重症”、重视“辨病机”七条规则, 并借助人工智能算法加以转化实现, 从而提升中医诊断辨证的客观性和准确性。

[关键词] 证素辨证; 状态要素; 中医诊断学; 人工智能

构建于证素辨证基础上的中医状态学^[1]是进行中医健康管理的重要理论和技术指导, 将之与人工智能技术结合不仅能更好地为健康管理事业服务, 也有利于成果在临床医生和科研工作人员之间加以转化应用。从定性描述到定量分析是科学发展的必由之路, 作为开启中医临床定量诊断的证素辨证^[2]是基于中医学理论, 同时又结合计算机数据挖掘分析处理技术而建立起的科学定量的辨证模型。该方法通过对所采集的症状赋值计算积分, 判断总积分是否达到或超过诊断阈值而得出诊断结果。这一模型对症状采集的全面性要求较高, 而继承该思想的中医状态学所提出的状态辨识原理也同样如此。然而由于临床工作者的水平参差不齐, 实际操作中表征参数(疾病状态下为症状)采集遗漏的情况普遍存在, 致使状态要素(疾病状态下为证素)的积分受到影响, 易使得出的诊断结果不尽理想。因此, 在发展中医诊断客观化以全面准确采集表征参数的同时, 有必要设计一套诊断规则来应对参数采集不全的情况, 以提升诊断结果的精确性。为解决这一问题, 本文将结合中医理论基础及中医临床诊疗实际需求加以阐述。

1 七条规则理论阐述

1.1 围绕要素阈值

状态要素和表征是多对多的关系, 临床情况复杂多变, 用一个状态要素来解释所有采集来的表征较难实现, 需要多个状态要素的组合。组合的数量可能是数千乃至数万, 但并非所有组合里的状态要素都是正确的。表征参数(症状)有对应状态要素(证素)的计量权值, 只有加权求和达到阈值的状态要素才是明确成立的^[3]。因此, 要选择包含有明确成立的状态要素的组合, 并且该组合内状态要素所映射表征的并集必须能解释所有采集而来的表征。所有此类组合形成一个集合, 我们暂且称其为入选组合集。若所有状态要素均未达到明确成立的阈值, 则围绕加权求和后权值最高的状态要素形成入选组合集。

1.2 一(少)元论

颜德馨^[4]认为, 再复杂的病情也有一个起决定性作用的症状, 其他症状都是据此产生和转变的, 临床思维渐进的踪迹, 都贯穿着“一元论”思想。该观点与现代医学的诊断“一元论”观点^[5]不谋而合, 用尽量少的、精准的、切中机要的诊断去解释繁复病情, 是临床医师的基本思维, 而非“一症一病”或“一症一证”。尽管临床实际

基金项目: 国家自然科学基金(U170520134)

* 通讯作者: fjzylcd@126.com

中几乎不可能仅用一个状态要素便囊括所有表征参数,但仍需尽量以少为佳,因此,要在入选组合集中优先选择状态要素少的组合。

1.3 伴随表征的共同指向

“司外揣内”是中医诊断思维之一,所谓“有诸内者,必形诸外”。作为外候,每个表征参数均对应着诸多状态要素,如“头晕”对应肝、半表半里、痰、阳亢、气虚、血虚、阴虚、动风等状态要素,要精准地辨识状态,则有赖于对其伴随表征的综合判断。若伴头眼胀痛、阵发烘热等,基本可以判断为阳亢;若伴少气懒言、易外感等,则可判断为气虚。因此,正确地辨识往往是表征群共同指向的一个状态要素,可将之称为综合归纳判断推理思维^[6]。在入选组合集中,优先选择所采集的表征群共同指向率高的组合,是该思维在诊断算法规则中的具体体现。

1.4 三因制宜

许多研究^[7-10]发现,在不同疾病状态下,体质、季节节气、地域等均对证素的分布有影响,如气虚体质与气虚、阳虚、阴虚显著相关;广东易出现脾湿,新疆易出现肾阴虚;春分时,肝气郁滞产生的可能性大;而昆明的秋冬季节,燥产生的可能性大。另外,老年人多见包含虚的证素,儿童多见食积证素。中医状态的影响因素也包括自然因素、社会因素和个人因素^[1],如嗜食辛辣之人易见热、津亏状态要素。因此,要在入选的组合集中优先选择包含人、地、时易出现的状态要素的组合。

1.5 重视辨病

辨病论治是中医亘古未老的诊治模式^[11],与辨证论治有机结合能更好地提高临床疗效^[12]。病是对疾病发展全过程的概括,对证有提示作用,如痹的基本病机为风、寒、湿三气杂至,提示其基本状态要素为“(外)风”“寒”“湿”。重视“辨病”规则,即在入选组合集中,应优先选择包含疾病所关联的状态要素且平均关联率最高的组合(与疾病关联的状态要素个数/总状态要素个数),如当病名诊断为“痹”时,优先选择包含“(外)风”“寒”“湿”的状态要素组合。

1.6 重视“主症”“重症”

“主症”即主诉的表征,“重症”即程度重的表征。因此,在同等条件下,应优先选择包含主症或(及)重症所映射状态要素且平均映射率最高的组合[与主症或(及)重症映射的状态要素个数/总状态要素个数]。如某患者主诉为阵发性呛

咳3天,伴随症状为痰中带血、咽喉痛、鼻唇干燥、舌质红、干燥、苔黄干、脉数,此时的状态要素组合可能为“肺,阴虚”或“肺,热”,但由于主诉为阵发性呛咳,对应应有状态要素“热”而无“阴虚”,故应优先选择组合“肺,热”。

1.7 重视“辨病机”

状态要素之间并非孤立的,病位要素之间、病性要素之间、病位要素与病性要素之间存在着大量的关系,如脾病可能导致肺病(土不生金),(外)风可能侵袭肺,(外)风可能兼夹寒等。随着中医学的发展,目前学术界的共识认为病机即疾病发生、发展与变化的机理^[13],因此,状态要素之间的关系即为病机,如“见肝之病,知肝传脾,当先实脾”。重视“辨病机”规则,即在入选组合集中,优先选择状态要素两两之间关系最多的组合。

2 计算机算法的相关实现流程

假设一个病例的所有表征组成的集合为 $A = \{a_1, a_2, \dots, a_n\}$ 其中“主症、重症”组成的子集合为 $B = \{b_1, b_2, \dots, b_n\}$ 。基于前一部分介绍的规则,相应构建的算法的具体计算流程主要分为两个部分:1) 基于规则1,在证素辨证积分的基础上通过优先队列算法计算得到可能诊断的候选集合;2) 基于规则2~7,对候选集合中的可能诊断进行相应的投票筛选,并根据投票得分确定最终诊断结果。

2.1 构建可能诊断的候选集合

基于朱文锋编著的《证素辨证学》^[2]中的“证候辨证素量表”,首先计算表征集合 A 相对应的状态要素积分,并按照积分从高到低进行排序,得到候选状态要素集合 $C = \{c_1, c_2, \dots, c_n\}$ 。同时将 C 中已达到阈值能够明确诊断的状态要素作为最基础的状态要素集合 $C_{basic} = \{c_1, c_2, \dots, c_a\}$,并在集合 A 中提取所有与 C_{basic} 具有关联的表征集合 A_{basic} 。通常 C_{basic} 对应的表征集合 A_{basic} 并不能包含所有的表征 A ,接下来将通过优先队列算法通过集合 $\{c_{a+1}, c_{a+2}, \dots, c_m\}$ 来补充状态要素到 C_{basic} 集合中,使之可以涵盖所有的症状 A 。

优先队列是一种队列节点具有不同优先级的队列,为实现本文中的诊断计算,我们将队列中的每一个节点用于记录当前新补充的状态要素,并把其与 C_{basic} 一起能对应的表征个数作为相应的优先级。后续优先队列算法的具体计算流程是:首先按照状态要素积分从高到低将 $\{c_{a+1}, c_{a+2}, \dots, c_m\}$ 中

的每一个状态要素都单独创建对应补充状态要素节点添加到队列中；然后从队列中取出优先级最高的节点，判断是否已对应涵盖 A 中所有的表征，如果是就将该结果作为一个可能诊断加入到候选集合 D 中，否则就在该节点的基础上继续补充一个新的状态要素，并作为新的节点添加到优先队列中。之后重复前面的计算直到整个队列为空，得到所有的可能诊断。

2.2 诊断结果的投票筛选

在得到候选集合 D 之后，通过规则 2~7 对 $D = \{d_1, d_2, \dots, d_i\}$ 中的每一个元素进行相应的投票积分，具体的计算分为以下 6 个步骤。1) 基于“一（少）元论”规则，对补充状态要素最少的 d_i 的投票得分加 1；2) 基于“伴随表征的共同指向”规则，计算每一个 d_i 的补充状态要素相对应表征数量平均数，并对平均数最大的 d_i 的投票得分加 1；3) 基于“三因制宜”规则，对指向优先规则最多的 d_i 的投票得分加 1，对指向否定规则的 d_i 的投票得分减 6（否定规则指如男性出现胞宫、女性出现精室等绝不可能有的状态要素）；4) 基于重视“辨病”规则，计算补充状态要素与“病”的平均关联率，对最大的 d_i 的投票得分加 1；5) 基于重视“主症”“重症”规则，计算补充状态要素与“主症”“重症”的平均映射率，对最大的 d_i 的投票得分加 1；6) 基于重视“辨病机”规则，计算每一个补充状态要素与基础状态要素集合 $C_{basic} = \{c_1, c_2, \dots, c_a\}$ 之间相关关系的平均关联度，对最大的 d_i 的投票得分加 1。

通过上面 6 步的计算之后，按照投票得分从高到低进行排序，其中得分最高的诊断将作为最终的诊断结果。

3 实际应用验证分析

为保证所分析的表征参数齐同，所选医案应当一致。通过对比运用本文所拟规则与基于证素辨证的原有“中医（辅助）诊疗系统”分别对医案加以判别，从而分析说明加入计算规则的必要性。

医案选自《中国名老中医药专家学术经验集·第五卷》罗元恺名案评析：寒凝血瘀痛经案^[14]。基本信息：患者，女，34 岁，职业教师。1989 年 2 月 8 日初诊。主诉：经行腹痛 19 年，以经行首日为甚。该患者从 15 岁月经初潮始即痛经，尤以第 1 日为甚，需长期口服或注射止痛药。两年前诊为子宫内膜异位症，手术切除左侧卵巢巧克力囊

肿。术后痛经减轻，近期又逐渐加重，经量多，色暗，有血块，持续 7~8 天，周期尚准。末次月经 1989 年 1 月 18 日，伴胃纳欠佳，形体消瘦，舌淡暗、苔白，脉沉细迟涩。已婚 4 年，性生活正常，未孕。诊断：痛经；不孕。辨证：寒凝血瘀，胞脉阻滞。治法：温经散寒，活血化瘀。若将病情输入“中医（辅助）诊疗系统”，其诊断结果为：1) 辨证要素为肝（80），气虚（73），血瘀（140），达到阈值以上的证素为血瘀（140）。2) 证名提示为瘀阻胞宫证（60），肾精亏虚证（40），故诊断为瘀阻胞宫证。据辨识结果来看，得出的诊断结果并不理想，未体现出证素“寒”，辨识所得之证素也未能解释脉沉和脉迟，因此，仅仅依据积分阈值较难凭借有限的表征参数得出精确的诊断结果。

若将上述医案的表征参数经新添加的规则重新诊断，其结果为：依据规则 1，围绕明确达到阈值的状态要素“胞宫、血瘀、气滞”进行组合，组合映射表征的并集必须能对应所有采集而来的表征，得到排名前 5 的状态要素组合：1) 状态要素组合 1：胞宫（243）、血瘀（172）、气滞（131）、气虚（92）、阳虚（74），状态名为胞宫气阳两虚气滞血瘀证。2) 状态要素组合 2：胞宫（243）、血瘀（172）、气滞（131）、气虚（92）、寒（66），状态名为胞宫气虚寒凝气滞血瘀证。3) 状态要素组合 3：胞宫（243）、血瘀（172）、气滞（131）、气虚（92）、血寒（52.5），状态名为胞宫寒凝血瘀证兼气虚气滞证。4) 状态要素组合 4：胞宫（243）、血瘀（172）、气滞（131）、血虚（74）、阳虚（74）、湿（20），状态名为阳虚湿阻气滞证兼胞宫血虚瘀证。5) 状态要素组合 5：胞宫（243）、血瘀（172）、气滞（131）、血虚（74）、阳虚（74）、外风（20），状态名为胞宫阳虚气滞血瘀证兼血虚风袭证。

根据规则 2 和规则 3，优先选择所采集的表征群共同指向率高的组合，详见表 1。表征“痛经”即为“主症”也为“病名”，故根据规则 5 及规则 6，选择关联/映射率高的组合，详见表 2。根据规则 7，每个状态要素均与其余所有状态要素有关系详见图 1。最终结果组合序号 1~5 得票数分别为 3、4、3、0、0。故而最终诊断的状态要素组合应为：胞宫（243），血瘀（172），气滞（131），气虚（92），寒（66），状态名称为胞宫气虚寒凝气滞血瘀证。进一步提升了“中医（辅助）诊疗系统”的合理性。

因该案为既有病例，无法采集患者所处地域、气候节气以及个人体质等表征参数进行“三因制宜”规则的综合判断。

表 1 基于中医状态辨识规则的痛经医案表征共同指向情况

组合序号	组合独有状态要素 (个数)	所映射表征个数	共同指向率 (%)
组合 1	气虚、阳虚 (2)	8 + 5 = 13	6.5
组合 2	气虚、寒 (2)	8 + 4 = 12	6
组合 3	气虚、血寒 (2)	8 + 4 = 12	6
组合 4	血虚、阳虚、湿 (3)	5 + 5 + 2 = 12	4
组合 5	血虚、阳虚、外风 (3)	5 + 5 + 2 = 11	3.7

表 2 基于中医状态辨识规则的痛经医案
关联/映射“痛经”情况

组合序号	组合独有状态要素 (个数)	关联/映射“痛经”的状态要素	关联/映射率 (%)
组合 1	气虚、阳虚 (2)	无	无
组合 2	气虚、寒 (2)	寒 (1)	50
组合 3	气虚、血寒 (2)	血寒 (1)	50
组合 4	血虚、阳虚、湿 (3)	血虚 (1)	33
组合 5	血虚、阳虚、外风 (3)	血虚 (1)	33

4 讨论

基于人工智能算法的中医状态辨识规则能够解

决表征过少导致状态要素积分未达阈值而无法诊断的问题。表征参数少可能由于采集缺漏也可能本来如此，但表征不会无由存在，存在则必有因，因此，诊断规则极为重要。不仅如此，通过规则算法亦可加入专家经验，如“五辨”思维^[15]在本规则算法中已得到体现。临床医师必备的鉴别诊断思维亦可通过规则实现，具体为在入选组合集中，以并列得票最高的 3 个状态要素组合，组成状态名称后进行鉴别诊断。将各状态名称中状态要素的表征集交集与采集而来的表征进行对比，并检验组合中所有状态要素的必有表征、主要表征、特异性表征和否定表征是否存在。对未采集到的表征补充采集，剔除含有否定表征以及不含有必有表征的状态要素，并重复上述规则步骤，重新计分和状态辨识，明确最终状态。

人工智能技术在中医药的应用是近年来研究的热点之一，其中基于数据驱动的新一代人工智能算法更是当前研究主流，同时也对数据的完整性、规范化提出了更高的要求。此外，由于医学中存在一些不宜违背的基本原理，如脏腑的生理功能、病理

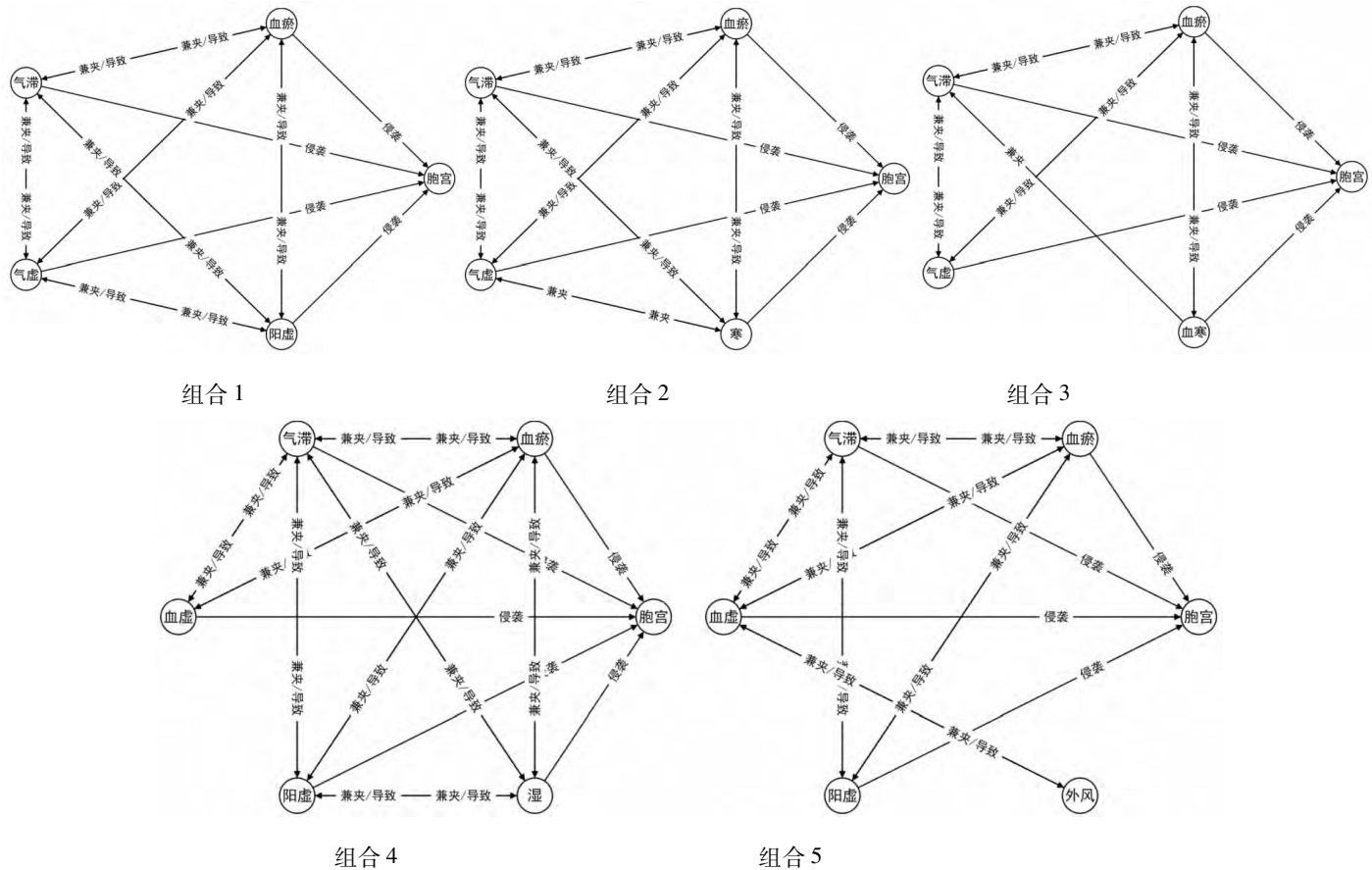


图 1 基于中医状态辨识规则的痛经医案状态要素关系图

状态等,在构建证素辨证体系基础上的人工智能辨识算法时,亦需将规则驱动与数据驱动相结合以提高其准确率。本文设计的相关规则符合中医思维,并通过编程实现,符合中医现代化发展需要,可更好地为健康管理和中医临床服务。

参考文献

- [1] 李灿东. 中医状态学 [M]. 北京: 中国中医药出版社, 2016: 46-50.
- [2] 朱文锋. 证素辨证学 [M]. 北京: 人民卫生出版社, 2008: 71-72, 90-161.
- [3] 朱文锋, 樊新荣, 姜瑞雪, 等. 用评定量表法进行中医辨证研究 [J]. 湖南中医学院学报, 2006, 26(2): 17-18, 21.
- [4] 颜德馨. 中医辨证思维与临床诊疗决策之优化 [J]. 上海中医药杂志, 2000, 44(5): 4-7.
- [5] 张鹏鸽, 陈新蕊. 浅谈对中医一元论的认识 [J]. 中国卫生标准管理, 2017, 8(11): 106-107.
- [6] 王天芳, 杨维益. 中医辨证推理思维方式的探讨 [J]. 北京中医药大学学报, 1994, 17(4): 2-4.
- [7] 田松, 祁若可, 程月招. 514 例亚健康人群中医体质与证素特点及其关系初步研究 [J]. 中华中医药杂志, 2015,

30(1): 243-245.

- [8] 姜枫, 符林春, 马建萍, 等. 艾滋病中医病因病机中的地域因素 [J]. 中国中西医结合杂志, 2012, 32(6): 748-750.
- [9] 孙玉琴. 小儿咳嗽变异性哮喘中医证素与环境因素相关性研究 [D]. 昆明: 云南中医学院, 2016.
- [10] 韩木龙. 慢性胃炎的证素特征及其与节气、气象因素的相关性研究 [D]. 福州: 福建中医药大学, 2014.
- [11] 马冠军. 关于辨病论治的认识 [J]. 中医杂志, 2014, 55(7): 626-627.
- [12] 马继征, 姚乃礼. 论辨证论治与辨病论治相结合 [J]. 中华中医药杂志, 2015, 30(12): 4251-4253.
- [13] 王维广, 陈子杰, 王慧如, 等. 当代中医病机概念的演变 [J]. 中医杂志, 2017, 58(17): 1441-1443, 1457.
- [14] 邱德文, 沙凤桐, 熊兴平. 中国名老中医药专家学术经验集: 第五卷 [M]. 贵阳: 贵州科技出版社, 1999: 600.
- [15] 李书楠, 李思汉, 赵文, 等. “五辨”与中医临床诊断思维的综合运用 [J]. 中华中医药杂志, 2019, 34(1): 18-22.

Rules of Traditional Chinese Medicine State Identification Based on Artificial Intelligence Algorithm

XU Jiajun^{1,2}, LUO Zhiming³, ZHAO Wen^{1,2}, LEI Huangwei^{1,2}, ZHOU Chang'en^{1,2}, LI Shaozi³, LI Cande^{1,2}

1. School of Chinese Medicine, Fujian University of Traditional Chinese Medicine, Fuzhou, 350122; 2. Fujian Key Laboratory of Traditional Chinese Medicine Health Status Identification; 3. Department of Artificial Intelligence, School of Information Science and Technology, Xiamen University

ABSTRACT The research on objectification traditional Chinese medicine (TCM) diagnostic method technology is an important way and necessary stage to realize the modernization and intelligent development of TCM. The concept of disease position, disease nature, and “syndrome-syndrome elements” in syndrome element differentiation has promoted the digitalized development of TCM diagnostics. The emergence of state identification theory has further improved the objective, comprehensive and dynamic TCM evaluation system for human body factors, making TCM diagnostics gradually objective development. The author put forward seven rules including around the threshold of factors, meta-ism, common direction of accompanying representation, adjusting measures to three factors, attached importance to differentiating disease, strengthening primary syndromes, and severe disease, and strengthening differentiating pathogenesis. With the aid of artificial intelligence algorithms to achieve transformation of these rules, the accuracy and objectification of TCM diagnostic and syndrome will be improved.

Keywords syndrome element differentiation; state factor; diagnosis in traditional Chinese medicine; artificial intelligence

(收稿日期: 2019-07-05; 修回日期: 2019-09-05)

[编辑: 黄健]