

· 思路与方法学 ·

基于中医证型知识库的辨证模型研究

张 鑫¹ 查青林² 杜建强¹ 聂 斌¹ 刘 博¹ 刘 蕾¹

摘要 目的 研究一种基于中医证型知识库的辨证模型,为计算机辅助辨证提供模型支持。**方法** 在“对象—属性—属性值”三元组(OAV)证型知识规则表示方法的基础上,依据关联规则算法计算症状体征—证型的置信度,以证型为基准进行归一化处理后的置信度作为症状体征—证型的权值因子,证型知识规则因此表示为“对象—属性—属性值—权值”(OAVW);基于 OAVW 知识规则,通过计算激活证型的症状体征的权值和,来辅助判断患者所属的证型;并以文献中的证型描述为验证数据,采用该模型进行计算机模拟辨证,以验证模型的适用性。**结果** 来自文献的肺阴虚证临床表现,经过模拟辨证分析,得到各证型的权值和依次为肺阴虚证(1.0)、心肺阴虚证(0.676)、心肺阴虚血瘀证(0.608),辨证结果与文献相符,并提供了更多相关证型的信息。**结论** 基于中医证型知识库的辨证模型在肺阴虚证上的实验结果表明该模型具有理论正确性和可行性,可以为计算机辅助辨证提供一定模型支持。

关键词 “对象—属性—属性值”三元组;中医证型;置信度;辨证模型;权值;中医药信息

Study on the Model of Syndrome Differentiation Based on the Knowledge Base of Traditional Chinese Medicine Syndrome ZHANG Xin¹, ZHA Qing-lin², DU Jian-qiang¹, NIE Bin¹, LIU Bo¹, and LIU Lei¹ 1 School of Computer, Jiangxi University of Traditional Chinese Medicine, Nanchang (330004); 2 Medical Information Research Center, Jiangxi University of Traditional Chinese Medicine, Nanchang (330004)

ABSTRACT Objective To study a model of syndrome differentiation based on the knowledge base of Traditional Chinese Medicine (TCM) syndrome, thus providing model supports for the computer-aided syndrome differentiation. **Methods** On the foundation of the syndrome knowledge rules representation method of “object-attribute-value” triples (OAV), the confidence of symptoms to syndrome was calculated based on association rules algorithm, and then the confidence was normalized based on syndrome, which is the weight factor (WF) of symptoms to syndrome, so the syndrome knowledge rules can be represented as “object-attribute-value-weight” (OAVW). Based on OAVW knowledge rules, the sum of WF for all activated symptoms of the syndrome was calculated, which could aid to determine the syndrome of a patient. The author took the description of syndrome from literature as validation data, and put it to the model to simulate syndrome differentiation, to validate the applicability of the model. **Results** Clinical manifestation of Fei yin deficiency syndrome was from literature, through simulation analysis of syndrome differentiation, the sum of WF of each syndrome was as follow: Fei yin deficiency syndrome (1.0), heart and lung yin deficiency syndrome (0.676), Xin and Fei yin deficiency and blood stasis syndrome (0.608), the results of syndrome differentiation were consistent with the literature, and provided more information about the related syndromes. **Conclusion** By experiment of Fei yin deficiency syndrome, the result shows that the model of syndrome differentiation based on the knowledge base of TCM syndrome has theoretical correctness and feasibility, and it could provide model supports for the computer-aided syndrome differentiation.

基金项目: 国家自然科学基金资助项目(No. 61363042, No. 61562045); 江西省高校科技落地计划(No. LD12038); 江西省研究生创新基金项目(No. YC2014 - S278)

作者单位: 1. 江西中医药大学计算机学院(南昌 330004); 2. 江西中医药大学医药信息研究中心(南昌 330004)

通讯作者: 杜建强, Tel: 18170069045, E-mail: Jianqiang_du@163.com

DOI: 10. 7661/j. cjim. 20170517. 089

KEYWORDS “object-attribute-value” triples; traditional Chinese medicine syndrome; degree of confidence; syndrome differentiation model; weight factor; traditional Chinese medicine information

目前,中医诊断受主观因素影响比较大,主要是由于中医学理论的模糊性和中医诊断的经验性所导致。随着人工智能技术的发展,基于各类推理机制的专家系统在医疗辅助诊断领域获得广泛研究和发展^[1,2],在中医学领域也有一些学者利用各种方法做过相关研究,如朱咏华等^[3]利用贝叶斯网络建立的中医辨证系统,周恣等^[4]则综合论述了不同机器学习算法用于中医学证候分类的情形,施明辉等^[5]为了解决传统方法的规则激增和推理缓慢等问题提出了利用改进的反向传播(back propagation, BP)神经网络实现中医学辨证的 uncertainty 推理。作为中医治病的核心,辨证过程就是医生通过患者出现的症状信息逐步进行推理判断并最终找出对应证型的过程,这种过程受限于中医学知识的模糊性而具有极强的不确定性。如果计算机能够量化这种模糊性并利用启发式推理模拟专家思维进行推理,那么就可以构建良好的辨证模型,从而有助于辅助诊断。

本文以知识表示为出发点,引入权值因子(weight factor, WF)来度量中医学证型知识的模糊性,同时利用关联分析中的置信度来表示权值因子以对传统的“对象—属性—属性值”(object attribute value, OAV)模型进行改进,基于此模型构造中医证型知识库。在此基础上,构建基于知识库的辨证模型,并利用文献中的证型案例充分验证了模型的有效性和可行性。

1 基于改进型 OAV 的中医学证型知识表示

1.1 OAV 理论基础 形如 <对象,属性,属性值>知识表示方法称为三元组表示法,也称 OAV 法,是知识工程中的一种知识表达方法,可用来描述任何一种知识单元或事实。在 OAV 表示法中,O 表示对象,这种对象是一种概念类实体;A 表示属性,是指与对象相关的特征性质;V 表示属性值,是指该对象对应属性的值。如某条知识“阴虚证的脉象是脉细数”的 OAV 表示方式为 <阴虚证 脉象 脉细数>。这种灵活的知识表示方法应用极其广泛,任何一种知识或事实都可以用 OAV 法进行表达,并且较易在此基础上抽取规则,形成可供推理用的规则库^[6]。

1.2 基于置信度的 OAV 模型改进 OAV 模型可以用于对中医学知识进行表示,然而在表示症状体征和各证型之间关系时,仍然存在不足。各证型之间的某些症状互相重复与交错,关系复杂且具有强烈的模糊性,很难找到确切的量化标准来衡量关系的主次^[7]。传统的 OAV 模型并不能较为完整的表示这种

模糊关系,譬如用 OAV 模型表示知识“五心烦热和口干是阴虚证的症状”见图 1。

<阴虚证	症状	五心烦热>
<阴虚证	症状	口干>

图 1 OAV 表示中医学证型知识

从图 1 中可见五心烦热和口干都是阴虚证的症状,至于哪个症状是主要症状哪个是次要症状却很难得知。中医学专家一般都认为“五心烦热”是阴虚证的主要症状,“口干”是阴虚证的次要症状。对此情况,OAV 模型却未表示出同一证型与不同症状之间的模糊对应关系,针对这些不足,本文引入权值因子,借助权值因子对症状和证型之间的模糊关系进行量化,将传统的 OAV 模型改进为“对象—属性—属性值—权值”(object attribute value weight, OAVW)模型。将上文的两条知识用 OAVW 进行表示,如图 2 所示。

<阴虚证	症状	五心烦热	WF>
<阴虚证	症状	口干	WF>

图 2 OAVW 表示中医学证型知识

面对证型与症状之间复杂的关系结构,如何客观计算这些模糊关系的 WF 值是一个需要考量的问题。如图 3 所示,ZX1、ZX2、ZX3 表示三个不同证型,ZZ1、ZZ2、……、ZZk 表示 ZX1 的 k 个症状,其中 ZZ1 是 ZX1、ZX2 和 ZX3 的公共症状,ZZ2 是 ZX1 和 ZX2 的公共症状,ZZ3、……、ZZk 为 ZX1 独有的症状,通过图 3 可以发现当某个症状所在证型越多时,相比较此证型中的其他症状而言,此症状对应该证型的权值应该较小;当某个证型的症状总数越少时,相比较此证型中某个症状在其他证型中的权值而言,该症状在此证型中的权值应该较大。

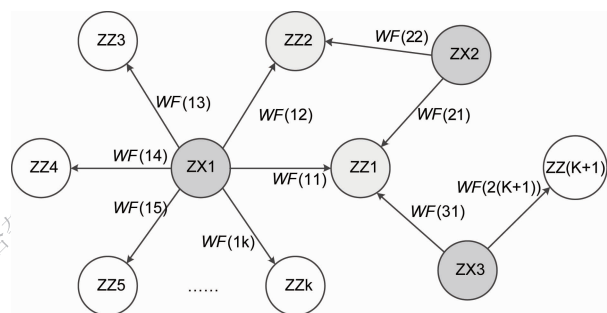


图 3 证型和症状加权关系图

基于此考虑,本文采用关联分析方法中的置信度 (confidence, 简记为 Conf) 来表示证型与症状的权值。数据挖掘中的关联分析方法是通过对事务数据库 D 中两个事务对象的置信度和支持度来确定这两个事务的密切程度,然后设置最小的置信度和最小的支持度阈值来筛选关联程度最大的两个事务^[8]。其中置信度的计算方法为:

$$\text{Conf}(x = > y) = \frac{x \text{ 与 } y \text{ 同时出现的记录数}}{\text{出现 } x \text{ 的记录数}}$$

本文将证型知识库作为事务数据库 D, 症状和证型分别作为 D 中的事务, 则某症状 *i* 和证型 *j* 之间的权值 *WF* (证型 *j*, 症状 *i*) 的计算过程如下所示:

(1) 首先计算症状 *i* 和证型 *j* 的置信度 Conf (症状 *i* = > 证型 *j*)。

$$\text{Conf}(\text{症状 } i = > \text{证型 } j) = \frac{N(\text{证型 } j, \text{症状 } i)}{N(\text{症状 } i)}$$

(2) 对证型 *j* 中所有症状的置信度进行归一化处理, 即可计算证型 *j* 和其中任意症状 *i* 的 *WF* (证型 *j*, 症状 *i*)。

$$WF(\text{证型 } j, \text{症状 } i) = \frac{\text{Conf}(\text{症状 } i = > \text{证型 } j)}{\sum_{i=1}^n \text{Conf}(\text{症状 } i = > \text{证型 } j)}$$

此公式中 *n* 表示证型 *j* 中总共含有 *n* 个症状。

2 中医辨证模型框架

2.1 辨证推理过程 在辨证过程中, 以知识库中单条知识为诊断规则, 采用启发式推理机制, 以患者症状信息激活知识库中的证型作为候选证型, 再计算每个候选证型激活症状的所有权值的和 Sum, 并将 Sum 作为该证型的概率 P, 最后将候选证型及其概率作为结果予以输出。假设知识库存在证型 ZXx, 且 ZXx 存在 7 个症状为 ZZ1、ZZ2、ZZ3、ZZ4、ZZ5、ZZ6、ZZ7, 对应的权值分别为 WF (x1)、WF (x2)、……、WF (x7), 现从某患者的病历中抽取的症状信息包含症状 ZZ1、ZZ2、ZZ3、ZZ5、ZZ7, 则该患者为证型 ZXx 的概率 P (ZXx) 为:

$$P(ZXx) = \text{Sum}(ZXx) = WF(x1) + WF(x2) + WF(x3) + WF(x5) + WF(x7)$$

2.2 模型框架设计 本文提出的中医辨证模型总框架为三层式, 包含中医证型知识层、辨证推理层和用户层。底层的证型知识层是由专家以中医文献和古籍为基础提取中医证型知识, 以 OAVW 模型进行存储和表示; 在辨证推理层, 则运用基于证型知识库的启发式推理机制模拟中医专家的诊断思维, 逐步确定对证型的判断并将得到的权值和作为该证型的概率, 达到辨证的目的。模型框架如图 4 所示。

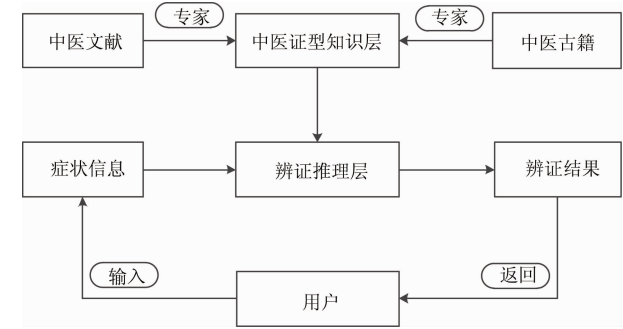


图 4 基于中医证型知识库的辨证模型框架

3 模拟辨证举例

本文以江西中医药大学医药信息研究中心提供的 1 317 个证型为基础构建中医学证型知识库, 该证型知识库来源于国家标准规范^[9]中对证的描述, 对证型和症状都进行规范化处理。以《邓铁涛医学文集·中医证候规范》^[10]中关于肺阴虚证的描述为例, 利用 MATLAB 程序进行模拟辨证, 辨证结果见表 1。肺阴虚证临床表现如下, 主症: 干咳无痰或痰少而黏、或痰中带血, 声音嘶哑, 午后潮热, 口咽干燥, 五心烦热, 盗汗。主舌: 舌红少津, 苔薄或苔少而干。主脉: 脉细数。

表 1 肺阴虚证症状模拟辨证结果

序号	证型	激活症状数	权值和 (Sum)
1	肺阴虚证	15	1.000
2	心肺阴虚证	5	0.676
3	心肺阴虚血瘀证	6	0.608
4	肺热津伤证	3	0.522
5	心阴虚火旺证	5	0.454
6	阴液亏虚证	6	0.329
7	脾肺气阴两虚证	2	0.285
8	肝肾阴虚证	4	0.246
9	心肺气阴两虚证	2	0.239
10	阴虚痰热证	5	0.220
11	心气阴两虚证	2	0.215
12	阴虚咽喉失濡证	5	0.196
13	心阴血虚证	1	0.183
14	阴虚鼻窍失濡证	4	0.180
15	心阴虚证	5	0.177
16	肝阴虚血瘀证	3	0.177
17	肝脾气阴两虚证	2	0.172
18	胃热阴虚证	4	0.166
19	燥热伤肺证	1	0.165
20	阴虚耳窍失濡证	4	0.151

注: 本表仅列出排名前 20 位的证型

以辨证结果中排名前 3 位的证型为例说明模拟辨证过程, 具体结果如表 2 所示。表 2 显示了该证型案例激活知识库中肺阴虚证、心肺阴虚证和心肺阴虚血瘀证对应症状知识的情况, 所得结果也较为符合实际的辨证结论, 证明模型具有一定的可行性。

表 2 模型验证结果表

肺阴虚证			心肺阴虚证			心肺阴虚血瘀证		
症状知识	激活状态	WF	症状知识	激活状态	WF	症状知识	激活状态	WF
咽干	YES	0.018	心悸	NO	0.000	胸痛	NO	0.000
形体消瘦	YES	0.038	五心烦热	YES	0.155	心悸	NO	0.000
午后潮热	YES	0.180	舌红	YES	0.028	五心烦热	YES	0.055
五心烦热	YES	0.018	少苔	NO	0.000	痰中带血	YES	0.367
痰中带血	YES	0.120	颧红	YES	0.232	舌红	YES	0.010
痰黏	YES	0.077	脉细数	YES	0.072	少苔	NO	0.000
痰少	YES	0.077	咳嗽	NO	0.000	颧红	YES	0.083
声音嘶哑	YES	0.180	盗汗	YES	0.189	脉细数	YES	0.026
舌少津	YES	0.025	—	—	—	脉涩	NO	0.000
舌红	YES	0.003	—	—	—	咳嗽	NO	0.000
颧红	YES	0.027	—	—	—	盗汗	YES	0.067
脉细数	YES	0.008	—	—	—	刺痛	NO	0.000
口燥	YES	0.027	—	—	—	—	—	—
干咳无痰	YES	0.180	—	—	—	—	—	—
盗汗	YES	0.022	—	—	—	—	—	—
激活症状数	15		激活症状数	5		激活症状数	6	
权值和(Sum)	1.000		权值和(Sum)	0.676		权值和(Sum)	0.608	
证型概率(P)	100%		证型概率(P)	67.6%		证型概率(P)	60.8%	

4 结束语

本文研究了中医学领域知识表示和辨证推理等相关知识,利用权值因子作为量化中医学证型与症状之间模糊关系的指标来改进传统 OAV 模型,并利用关联规则算法计算症状体征—证型的置信度来表示权值因子,以此模型构建中医证型知识库能够有效表征证型与症状的模糊关系。而在辨证推理阶段,本文采用基于中医证型知识库的启发式推理过程,模拟辨证的结果也与实际较为一致,表明基于中医证型知识库的辨证模型具有理论正确性和可行性,可以为计算机辅助辨证提供一定的模型支持。但是本文提出的模型是基于证型知识库,因此对知识库提出了较高要求,而部分证型的冗余症状信息对模型的准确性造成了一定影响,如何有效优化知识库以提高模型准确性是需要进一步考虑的问题。

(作者对文章的贡献:张鑫、查青林的工作对本文具有同样的贡献。)

参 考 文 献

[1] 林和平,管仁初,王艳.基于面向对象方法的医疗辅助诊断系统[J].计算机工程,2007,33(16):272-273.

[2] 张惠康,钱宗才,屈景辉,等.骨肿瘤辅助诊断专家系统模糊推理模型[J].医学争鸣,2003,24(2):182-185.

[3] 朱咏华,朱文锋.基于贝叶斯网络的中医辨证系统[J].湖南大学学报(自科版),2006,33(4):123-125.

[4] 周忞,褚娜,李洁.中医证候研究中的分类算法方法学研究[J].结合医学学报(英文),2010,8(10):911-916.

[5] 施明辉,周昌乐.一种基于 ANN 的中医辨证不确定性推理模型研究[J].计算机工程与应用,2007,43(7):10-13.

[6] 周和玉.科技文献的知识表达及应用研究[J].情报学报,1996,15(6):431-435.

[7] 江一平,刘明芝,姜灿文.模糊数学在中医四虚证鉴别诊断中的初步应用[J].中医药学报,1984,12(5):13-19.

[8] 李园白,崔蒙,杨阳,等.中医处方的关联分析与频次统计方法异同探析[J].中医学报,2013,28(12):1940-1943.

[9] 国家技术监督局.中华人民共和国国家标准中医临床诊疗术语:证候部分[M].北京:中国标准出版社,1997.

[10] 邓铁涛主编.邓铁涛医学文集[M].人民卫生出版社,2001:1786.

(收稿:2015-12-16 修回:2017-05-15)
责任编辑:汤 静