



# 确定证素辨证权值的“双层频权剪叉”算法

朱文锋, 何军锋, 晏峻峰, 黄碧群

湖南中医药大学中医诊断研究所, 湖南 长沙 410007

**摘要:** 根据证候, 辨别证素, 组成证名, 是中医辨证的规律与过程。中医辨证是非线性复杂巨系统。为解决证素辨证研究中诊断权值的确定这个关键问题, 在频数统计基础上, 实行“双层频权剪叉”算法, 可获得准确的辨证参数, 在非线性和多变量分析上探索出一种新的运算方式。

**关键词:** 辨证; 证候; 数据收集; 计算机辅助诊断; 算法

**中图分类号:** R2-05; **文献标识码:** A; **文章编号:** 1672-1977(2007)06-0607-05

## Algorithm of double levels of frequency and weight to determine the weight value for syndrome factor differentiation

Wen-feng ZHU, Jun-feng HE, Jun-feng YAN, Bi-qun HUANG

Institute of Traditional Chinese Medical Diagnosis, Hunan University of Traditional Chinese Medicine, Changsha, Hunan Province 410007, China

**Abstract:** Differentiating syndrome factor and forming syndrome type according to symptoms and signs are the rules and processes of syndrome differentiation in traditional Chinese medicine (TCM). TCM syndrome differentiation is a nonlinear complex giant system. In order to solve the key problem of determination of diagnosing weight value for syndrome factor differentiation, a new algorithm of double levels of frequency and weight based on the analysis of frequency statistics was applied, and the accurate syndrome differentiation parameters were acquired. Therefore, based on the nonlinear and multivariate analysis, a new algorithm of calculating diagnostics for syndrome factor differentiation was established.

**Keywords:** syndrome differentiation; syndrome; data collection; computer-assisted diagnosis; algorithms

Zhu WF, He JF, Yan JF, Huang BQ. *J Chin Integr Med/Zhong Xi Yi Jie He Xue Bao*. 2007; 5(6): 607-611. Received April 12, 2007; published online November 15, 2007. Free full text (PDF) is available at [www.jcimjournal.com](http://www.jcimjournal.com)

中医辨证是非线性的复杂巨系统, 证素辨证研究应该建立在数据挖掘的理念上, 以证素为信息分析的核心, 运用复杂系统分析方法, 探索从证候确认证素的有效方法。

### 1 辨证思维的规律

在中医学中, “证”相对于“病”来说, 是对疾病中机体整体反应状态的阶段性病理本质概括。患者的症状、体征, 如发热、咳嗽、舌苔黄、脉滑数等各种病理信息, 中医称为证候; 中医辨证的目的是为了明确病位与病性等辨证要素(简称证素), 主要有心、肺、脾、肝、肾、气虚、寒、痰、血瘀等约 50 项; 各证素的相互组合, 可概括成完整的证名诊断, 如肝胆湿热证、

肺肾阴虚证等。“辨证”就是根据中医学理论, 通过对证候进行分析, 确定其病理本质——证素, 并做出证名诊断的思维认识过程<sup>[1]</sup>。

“根据证候, 辨别证素, 组成证名”, 既是辨证的原则、规律, 也是辨证思维过程中的三个台阶、三个步骤。通过“四诊”而获取各种病理信息, 其目的是为了辨别证素, 任何规范的证名, 都是由证素的相互组合而形成。证候与证素之间、证素与证名之间, 存在极其复杂的网络关系, 构成了以证素为核心的辨证体系。因此, 中医辨证是非线性的复杂巨系统。

### 2 辨证的数据处理方法

中医辨证研究应该建立在数据挖掘的理念上,

相应数理模型和算法的建立是实现从定性描述到定量分析的必要方法。在中医辨证研究中,其数据处理方法主要有非条件 Logistic 多元逐步回归分析<sup>[2]</sup>、基于贝叶斯网络的辨证模型<sup>[3]</sup>、基于聚类分析的神经网络模式<sup>[4]</sup>、隐变量分析法中的结构方程模型<sup>[5]</sup>以及粗糙集和支持向量机等<sup>[6]</sup>。数据挖掘技术很多,不同的方法解决问题的能力不同,适应范围也不同。由于中医辨证是非线性复杂系统,现有的数据处理方法尚不能完全反映中医辨证的规律,仍难以解决在证的多变量信息分析中诸如有效因素组合爆炸之类的问题。因此,找到一种适合要求的数据挖掘方法,对于证素辨证的研究至关重要。

### 3 中医辨证数据库的建立

开展辨证的流行病学调查,建立中医辨证数据库,是获取证素辨证相关参数的有效方法。该数据库是为建立全病域的辨证系统提供信息依据,因此要求样本量大,病种全面,涉及证广,证候规范,辨证准确。

数据库资料来源于住院病例的首诊病历,门诊病历,名老中医诊疗验案及已颁发和制定的证候诊断标准等,也纳入了中医教材和辨证学论著中的辨证诊断资料。该数据库资料总计 5 139 例,证候条目近 700 个,涉及内、外、妇、儿以及皮肤科、眼科、耳鼻喉科等数百种疾病,各证素所见证候有较全面合理的分布。

数据库的内容,除一般资料外,主要包括每例病历的全部病情,现在症的主次轻重,临床辨证诊断,证名中涉用的证素,可能的病名诊断。数据库中的

资料(证候、证名、证素等)在入库时首先需经过规范化处理,保证原始资料的准确性。

病历数据的管理、统计、挖掘等,见图 1。

### 4 双层频权剪叉算法原理

中医临床辨证的实践提示,证候辨证具有多维复杂性,每个证候对各证素判断的贡献度,并不是简单的以出现频数的多少为依据,有些证候临床出现的频数虽然很高,但其对任何证素的判断能力均不强,与之相反,某些证候临床发生的频数虽然不高,但其对证素的诊断具有很强的特征性。为避免一些变量频数范围过大,而另一些变量频数范围过小所形成的局部优化和判别偏移,必须将频数转化成相应的权值,即根据证候的不同属性,拟定各证候及各证素的标准化权值。为使每个证候和每项证素纳入判断的机遇相等,应当依据高频数变量的权值轻,低频数变量的权值重的原则,这就是“频权剪叉”。据“频权剪叉”原理,对证素所见证候的权值进行分配,将各证候对各证素和证型的贡献度进行分配,形成证候标准化权值及证素标准化权值,故为“双层”。见图 2。

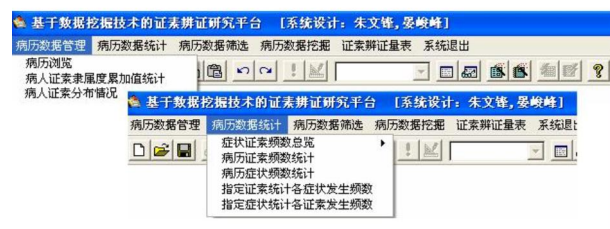


图 1 病历数据管理统计示意图

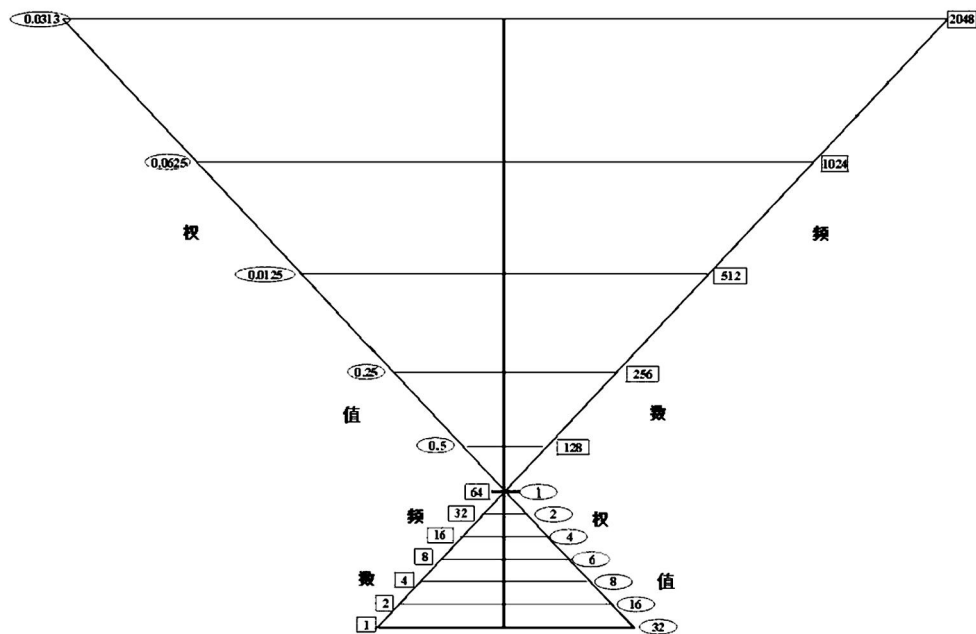


图 2 频权剪叉示意图

权值拟定的方法是:(1)从“中医辨证数据库”中将每个证候和每项证素出现的频数(证候总频数、证素总频数;各证候发生相关证素的频数、各证素发生相关证候的频数)进行累计;(2)按高频变量权轻,低频变量权重的原理,根据各证候的总频数计算出各自的权重,根据各证素的总频数合理分配出各自的权重;(3)某证候在某证素中出现的频数乘以该证候的权值和该证素的权值,即为该证候对该证素的判别系数。

根据双层频权剪叉算法,生成证候辨别证素的

权值,见图 3。

5 应用举例

在所建“中医辨证数据库”的基础上,采用“双层频权剪叉”算法,能够提取证素、证型的特征证候,获得准确的辨证系数。

5.1 证素的特征证候提取 从“中医辨证数据库”中提取指定证素的证候频数。证候频数×证候权值×证素权值=证素的证候系数。如证素“脾”的主要证候系数及排序见表 1。



图 3 证候辨别证素权值示意图

表 1 证素“脾”的主要证候系数及排序

症状代码	证候	辨证系数	排序	症状代码	证候	辨证系数	排序
1007	经常便溏	34.449 0	1	1505	面睑水肿	20.392 3	17
1002	经常腹泄	28.797 6	2	0637	气下坠感	19.801 6	18
0626	倦怠乏力	27.166 4	3	1640	内脏下垂	19.718 4	19
0910	食欲减退	27.144 0	4	0717	懒言	19.522 9	20
0633	腹胀	27.089 4	5	1013	黏液便	19.510 4	21
1623	腹水	24.733 8	6	1702	形瘦	19.325 3	22
0701	神疲	23.315 8	7	1803	舌淡	19.094 4	23
1622	腹部膨隆	22.880 0	8	1831	舌苔薄白	19.006 0	24
1706	经常浮肿	22.855 0	9	1110	少尿	18.634 7	25
0911	纳少	21.905 5	10	0823	气短	18.617 0	26
1422	面色萎黄	21.506 2	11	0706	嗜睡	18.558 8	27
0913	食后痞胀	21.385 0	12	0429	腹痛	18.512 0	28
1909	脉弱	20.992 4	13	1832	舌苔白	18.528 1	29
1804	舌淡胖	20.975 8	14	0216	经常畏冷	18.353 4	30
1028	肛门坠胀	20.748 0	15	.....			
1020	完谷不化	20.592 0	16				

采用双层频权剪叉算法,提取出病位证素“脾”的特征证候,与新世纪全国高等中医药院校规划教材《中医诊断学》<sup>[7]</sup>“临床以腹胀腹痛、便溏、浮肿、困重、内脏下垂、慢性出血为脾病常见症状”的论述完全相同。

5.2 证型的特征证候提取 从“中医辨证数据库”中提取指定常见证型所含证素的证候频数。证候频数×证候权值×证素权值=指定证型主要证候系数。如肝阳上亢证主要证候系数及排序见表 2。

采用双层频权剪叉算法,提取出肝阳上亢证的证候特征,与肝火亢盛证<sup>[8]</sup>“主症:眩晕、头痛、急躁易怒;次症:面红、目赤、口干、口苦、便秘、溲赤、舌红苔黄、脉弦数。”的论述非常近似。

5.3 辨证检验举例

临床辨证时,将所出现的证候,按照“双层频权剪叉”算法原理,先确定各证候映射相关证素的系数,再计算证素累加系数,系数达到或超过阈值以上的证素诊断成立。如 916 号病例的辨证计算,见表 3。

表 2 肝阳上亢证主要证候系数及排序

证候		频数和权值			
症状代码	症状	证候频数	×证候权值	×证素权值	排序
0710	急躁易怒	74	24.42	19.54	1
0603	眼胀及痛	22	24.20	19.36	2
1926	血压高	103	21.63	17.30	3
1514	眼球突出	8	21.60	17.28	4
1517	目赤	18	19.62	15.70	5
0636	头重足轻	14	19.04	15.57	6
0602	头胀及痛	38	21.66	15.50	7
0233	阵发烘热	15	18.45	14.76	8
1430	面红	58	15.66	12.53	9
0642	项背拘急	13	14.30	11.44	10
0615	新起耳聋	5	14.00	11.20	11
0613	突发性耳鸣	12	13.68	10.94	12
.....					

表 3 916 号病例辨证系数统计表

证候	阴虚	心	肝	血虚	气虚	气滞	热	痰	湿	心神	...
心悸(+)	14.8	38.1	10.2	20.0	15.1	7.1	5.5	12.5	3.9	7.8	
失眠	12.7	10.6	9.5	10.5	7.3	9.3	4.8	7.0	6.4	10.2	
眩晕	8.1	7.1	9.9	11.8	7.7	4.4	3.9	7.6	4.9	5.0	
神疲	6.5	6.7	5.9	8.8	10.8	5.3	4.2	5.2	6.9	4.4	
口苦	10.1	5.9	12.7	3.3	4.5	10.3	10.1	7.4	7.7	4.4	
大便干结	13.7	5.9	8.1	6.3	5.3	7.4	7.6	7.1	5.6	5.3	
尿短黄	9.3	5.0	7.1	2.9	3.6	7.1	12.8	5.6	9.5	4.2	
形瘦	11.3	5.1	8.2	0.8	9.4	7.8	4.4	5.5	5.2	3.2	
舌苔黄(一)	5.2	3.4	4.9	2.5	2.5	5.3	8.3	4.9	5.7	3.1	
舌红(一)	7.5	3.0	5.3	2.0	2.6	4.6	8.2	4.3	5.5	3.2	
脉细	8.7	6.3	6.3	8.2	7.2	5.3	3.6	5.2	4.6	4.0	
合计	107.9	97.1	88.1	77.1	76.0	73.9	73.4	72.3	65.9	54.8	

据上述计算,取系数 100 左右的证素——阴虚、心,从而诊断为心阴虚证,与临床辨证完全相符。

6 结 论

数据挖掘只有原始资料丰富完整,才能从中挖掘出有意义的信息。从大量的数据中进行辨证特征的规则提取,能够抽提出比较全面的内在规律。“中医辨证数据库”中的资料(证候、证素等)已经经过规范化处理,保证了原始资料的准确性,从而为辨证参数的获取奠定了基础。

在中医辨证体系的框架下,以临床信息为对象,

采用双层频权剪叉算法,从杂乱无章的数据中,找出中医辨证的规律,合理度量变量间的相关性,能明确证素与常见证型的特征证候以及证候对相关证素、证型的诊断贡献度。证候与证素间非线性映射函数的建立,对认识证候与证素及证型间复杂的非线性关系具有重要意义。“双层频权剪叉算法”克服了神经网络、贝叶斯网络等算法的某些不足,为解决中医辨证研究中诊断权值的确定这个关键问题,找到了一种简便准确的方法,在非线性、多变量分析上探索出了一种新的运算方式。

## REFERENCES

- 1 Zhu WF. Establishing a new system of syndrome factor differentiation. *Ke Xue Yan Jiu Yue Kan*. 2004; 2(3): 24-25. Chinese.  
朱文锋. 论中医证素辨证新体系. 科学研究月刊. 2004; 2(3): 24-25.
- 2 Zhang QM, Li KJ. Traditional Chinese medical statistics and diagnosis. Beijing: China Press of Traditional Chinese Medicine. 2004; Preface. Chinese.  
张启明, 李可建. 中医统计诊断. 北京: 中国中医药出版社. 2004; 前言.
- 3 Zhu WF, Yan JF, Huang BQ. Application of Bayesian network in syndrome differentiation system of traditional Chinese medicine. *Zhong Xi Yi Jie He Xue Bao*. 2006; 4(6): 566-571. Chinese with abstract in English.  
朱文锋, 晏峻峰, 黄碧群. 贝叶斯网络在中医证素辨证体系中的应用. 中西医结合学报. 2006; 4(6): 566-571.
- 4 Li JS, Hu JL, Yu XQ, *et al*. A research on RBF neural network based on clustering analysis in the application of syndrome diagnosis. *Zhongguo Zhong Yi Ji Chu Yi Xue Za Zhi*. 2005; 11(9): 685-687. Chinese with abstract in English.  
李建生, 胡金亮, 余学庆, 等. 基于聚类分析的径向基神经网络用于证候诊断的研究. 中国中医基础医学杂志. 2005; 11(9): 685-687.
- 5 Chen QG, Shen CT, Zhang HQ, *et al*. Application of structural equation model in standardization of syndrome differentiation. *Zhongguo Wei Sheng Tong Ji*. 2005; 22(1): 2-4. Chinese with abstract in English.  
陈启光, 申春梯, 张华强, 等. 结构方程模型在中医证候规范标准研究中的应用. 中国卫生统计. 2005; 22(1): 2-4.
- 6 Yan JF, Zhu WF. Apply rough sets theory in TCM syndrome factor diagnosis research. *Zhongguo Zhong Yi Ji Chu Yi Xue Za Zhi*. 2006; 12(2): 90-93. Chinese with abstract in English.  
晏峻峰, 朱文锋. 粗糙集理论在中医证素辨证研究中的应用. 中国中医基础医学杂志. 2006; 12(2): 90-93.
- 7 Zhu WF. Diagnostics of traditional Chinese medicine. Beijing: China Press of Traditional Chinese Medicine. 2002; 185-186. Chinese.  
朱文锋. 中医诊断学. 北京: 中国中医药出版社. 2002; 185-186.
- 8 Zheng XY. Guide principle of clinical research for the new drug of traditional Chinese herbs (trial implementation). Beijing: China Press of Traditional Chinese Medicine. 2002; 74. Chinese.  
郑筱萸. 中药新药临床研究指导原则(试行). 北京: 中国中医药出版社. 2002; 74.

## 《中西医结合学报》在 PubMed 实现全文链接

《中西医结合学报》是一本开放式访问期刊(Open Access Journal)。《中西医结合学报》杂志社致力于向世界展示我国中西医结合研究的成果, 向我国读者介绍世界结合医学的进展, 为全球读者提供一个即时和自由交流的学术平台, 倾情打造医学精品期刊、品牌期刊。

2007 年 1 月 6 日,《中西医结合学报》在 PubMed 实现了全文链接。全球读者不仅可以在 PubMed 检索到《中西医结合学报》发表的任何一篇论文的摘要和题录, 而且可以在 PubMed 网站通过其全文链接(Full Text LinkOut)功能免费下载阅读所检索到的论文的全文。

读者只要登陆 PubMed 网站 <http://www.ncbi.nlm.nih.gov/entrez>, 在检索框内输入《中西医结合学报》刊名的汉语拼音“Zhong Xi Yi Jie He Xue Bao”或英文全名“Journal of Chinese Integrative Medicine”, 或 ISSN 号“1672-1977”, 即可检索到《中西医结合学报》创刊以来发表的全部论文。点击每篇论文的全文链接图标, 即可免费阅读和下载全文。

国内外读者也可直接登陆本刊中英文双语网站 [www.jcimjournal.com](http://www.jcimjournal.com), 免费检索、阅读和下载全部论文。

中西医结合学报杂志社