

· 论坛 ·

文章编号 1000-2790(2007)04-0289-03

# 基于模块化思维的模糊 ISODATA 中医证候诊断方法

许晓娟,王庆国 (北京中医药大学基础医学院,北京 100029)

**【摘要】**中医学面临在继承发扬传统的基础上探求现代化转变,中医诊断的科学化、规范化、标准化、定量化和精确性等问题变得更为迫切。提出一种基于模块化思维并以模糊熵作为判断识别依据的模糊 ISODATA 中医证候诊断方法,以期在系统整理临床资料、古今资料及专家经验的基础上,模拟辨证诊治过程建立中医证候诊断方法。该方法通过对 IgA 肾病的辨证分型为例,识别结果与中医辨证诊断的思维大致吻合。该方法可用于典型疾病的中医辨证诊断,对于中医专家的经验积累和知识共享具有实际应用价值。

**【关键词】**模糊 ISODATA 方法;模块化思维;模糊熵;中医诊断;证候

**【中图分类号】**R2-03 **【文献标识码】**A

## 0 引言

中医辨证诊断是对疾病进行多层次、多角度的系统思考。虽然历代医家给作为中医辨证论治体系重要组成部分的“证候”赋予了丰富的内涵,但在中医诊断中也出现许多表述模糊抽象的不规范之处,病、证、症、病种与病类、证名与证候等概念混淆,称谓繁杂,名实异同,内容交错,理解各异,临床运用不知其所而随意取。随着中西医学理论的交贯渗透、相互沟通,中医学面临在继承发扬传统的基础上探求现代化的转变<sup>[1]</sup>,中医诊断的科学化、规范化、标准化、定量化和精确性等问题变得更为迫切。然而,中医诊断研究是一项系统工程和复杂性科学<sup>[2]</sup>,其开放性、模糊性和高维性等复杂性特征为中医现代化转变带来一定的难度,若对疾病和症状建立诊治识别关系将非常复杂,而且由于中医“四诊”信息的模糊特征,使中医证候诊断标准的准确性和规范性难以统一。我们根据中医疾病诊断的模糊特性和模块化思维,探索应用集成中医学、西医学、模糊数学、计算机等学科的理论、经验和方法,提出一种基于模块化思维并以模糊熵作为判断识别依据的模糊 ISODATA 中医证候诊断方法,为典型疾病的中医辨证诊断的规范化提供可行方法。

收稿日期 2006-09-08 接受日期 2006-11-23

基金项目:973“国家重点基础研究发展计划(2003CB517101)”

作者简介:许晓娟,北京中医药大学医学博士生。Tel:13911172804

Email:graceXXJ@126.com

## 1 中医的模糊概念及模块化思维

**1.1 中医的模糊概念** 中医的辨证实际上是一个模糊推论过程。中医疾病诊断过程中望、闻、问、切四诊方法获取患者的症状与体征,获取的这些症状与体征或多或少带有患者和医生的主观因素,在症状判断和辨证分型过程中融入医生的临床经验,这些都带有不确定性和无法精确量化。在中医学经典古籍中也有模糊定量及半定量的记载,模糊性思维贯穿于中医辨证论治的整个过程中。仲景伤寒论在辨证论治过程中处处表现出典型与非典型汤证的鉴别诊断与治疗。《伤寒论》六经病证可分为典型汤证和不典型汤证两大类。对于典型汤证,可采用二值逻辑思维,判断其是何方之证,而对于不典型汤证,则采用与典型汤证加以比较的方法,从而使其治疗达到近似于精确程度。清朝名医徐灵胎将伤寒论诸方“细分为十二类,每类先定主方(典型汤证方),即以同类诸方(较典型汤证方,即加减方)附焉”……使读者于病情药性一目了然(达到精确性断定)。不论从何经来,从何经去,而见症施治,与仲景之意,无不吻合。这是模糊思维在中医辨证论治运用中的典型例子<sup>[3-4]</sup>。

**1.2 中医的模块化思维** 在中医领域自古也在不自觉地运用模块化思维。中医理论框架和思维模式是:以人体为中心与自然界息息相关的“天人合一”观,把天、地、人、时的统一关系作为研究对象,以五脏为中心、经络为联系的整体和系统观念,医生在获取患者体征症状和辨证诊断时,分别从头面、心胸、四肢、五官等“模块”进行四诊,获取患者的各部位症状后再进行综合分析和评判。模块化思维在仲景《伤寒论》中有充分体现。如“胃中不和,心下痞硬,干噎食臭,肋下有水气,腹中雷鸣下利者”为生姜泻心汤证的症候群“模块”。临床上医生见患者有此主症候群者,便可“生姜泻心汤主之”。再如少阳主症“口苦,咽干,目眩,往来寒热,胸胁苦满,默默不欲饮食,心烦喜呕,脉弦细”之“但见一证便是,不必悉具”,便是模块化思维在《伤寒论》中的典型代表。

## 2 模块化思维在中西医结合诊断的应用

假设人体常见的疾病集  $U$  中的疾病种类有  $m$

个,每种疾病的类型有  $M_1, M_2, \dots, M_m$  个;疾病的中医症状和西医病理集  $V$  的状态变量有  $n$  个,每个状态变量取值分别有  $N_1, N_2, \dots, N_n$  个. 首先,对中医症状和西医病理进行频率分析、因子分析、回归分析等初步统计处理,将大量中医症状和检查指标进行归类、相关分析和主次分析,并判断疾病的病理、症状以及表现部位之间的初步关系. 根据症状性质、检查结果和初步分析,将中医症状和西医病理集划分  $n'$  ( $n' < n$ ) 个症状子集  $\{S_1, S_2, \dots, S_{n'}\}$ ,每个中医症状和西医病理子集包括  $n$  个变量中的若干个. 然后,根据初步统计分析结果结合专家经验,确定疾病证型分类标准以及特征症状的选取和隶属度的确定,将疾病集  $U$  表示为一系列证候的特征症状向量,特征症状向量包括  $m$  ( $m < m'$ ) 个特征参数  $x_1 - x_m'$ . 因此,  $m$  个疾病种类和对应类型  $M_1, M_2, \dots, M_m'$  可以表示为证型模糊特征向量  $\tilde{A}_i$ , 即:

$$\tilde{A}_i = [\mu_{A_i}(x_1), \mu_{A_i}(x_2), \dots, \mu_{A_i}(x_m)] \quad i = 1, 2, 3, \dots, \sum_{j=1}^m M_j$$

经过上述模拟疾病诊断过程处理,将疾病和症状进行子集划分和模块化处理后,疾病模糊诊断规模较直接识别大为减小,使之转变为  $n'$  个症状子集与  $m'$  个疾病特征症状间的回归关系,以及疾病特征症状向量与疾病证候间的模糊评判或识别问题. 在进行模拟中西医结合诊断时,可以采取模糊 ISODATA 聚类方法进行中西医结合疾病诊断,包括无经验的学习诊断和有经验的判别诊断. 模糊评判或识别主要包括以下 3 个步骤:从识别对象中提取与识别有关的特征并进行度量;建立标准类型的隶属函数或评判矩阵;建立识别判决准则和确定归属原则,以判定识别对象属于哪一个标准类型.

### 3 模糊 ISODATA 中西医结合诊断方法应用

**3.1 模糊 ISODATA 识别诊断方法** 设有某种疾病的  $n$  个临床调查病例的样本集  $X = \{X_1, X_2, \dots, X_n\}^T$ ,其中每一个样本有  $m$  个中医症状和西医病理指标,其特性指标矩阵为:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix}$$

将临床调查病例样本集  $X$  分成  $c$  类,设  $c$  个隶属度标准矩阵为  $V = (V_1, V_2, \dots, V_c)^T$ ,分类识别结果隶属度矩阵为  $R = (R_1, R_2, \dots, R_c)^T$ ,分别如下:

$$V = \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_c \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} & \dots & v_{1m} \\ v_{21} & v_{22} & \dots & v_{2m} \\ \dots & \dots & \dots & \dots \\ v_{c1} & v_{c2} & \dots & v_{cm} \end{bmatrix}$$

$$R = \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_c \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ \dots & \dots & \dots & \dots \\ r_{c1} & r_{c2} & \dots & r_{cn} \end{bmatrix}$$

为了得到最优的分类,需使如下目标函数取得极小值:

$$J(R, V) = \sum_{k=1}^n \sum_{l=1}^c (r_{lk})^q \|X_k - V_l\|^2 \quad \|X_k - V_l\| = \left[ \sum_{j=1}^m |x_{kj} - v_{lj}|^p \right]^{1/p} \quad (p > 0)$$

$\|X_k - V_l\|$  表示样本为第  $k$  个样本  $X_k$  与第  $l$  类聚类中心  $V_l$  的闵氏权距离. 当  $p=1$  时,为 Hamming 距离;  $p=2$  时,为欧氏权距离. Bezdek 已经证明:当  $q \geq 1$ ,  $X_k \neq V_l$  时,上述目标函数存在极小值<sup>[5]</sup>. Bezdek 通过对参数  $q$  的确定进行了模拟试验研究,试验结果表明参数  $q=2$  时相对较优<sup>[6]</sup>.

**3.2 模糊 ISODATA 方法求解过程** ① 初始化:对样本集  $X$  进行数据预处理,使其中元素值归一化至  $[0, 1]$  区间,选取分类组数  $c$  ( $2 \leq c \leq n-1$ ),选取迭代初始变量  $s=0$ ,给定初始聚类标准矩阵  $R^{(0)}$ ,进行逐步迭代计算. ② 第  $s$  步迭代:对于聚类标准矩阵计算模糊分类结果矩阵  $V^{(s)}$ ,然后根据计算得到的  $V^{(s)}$  修正聚类标准矩阵  $R^{(s+1)} = (R_1, R_2, \dots, R_c)^T$ .

$$V_l^{(s)} = \frac{\sum_{k=1}^n [(r_{lk}^{(s)})^q \cdot X_k]}{\sum_{k=1}^n (r_{lk}^{(s)})^q}$$

$$r_{lk}^{(s+1)} = \left[ \sum_{i=1}^c \left[ \frac{\|X_k - V_i\|}{\|X_k - V_{i1}\|} \right]^{q-1} \right]^{-1}$$

③ 迭代判别:比较  $R^{(s)}$  和  $R^{(s+1)}$ ,若对确定误差标准有  $\max\{|r_{lk}^{(s+1)} - r_{lk}^{(s)}|\} < \epsilon$  ( $\epsilon > 0$ ),则  $V^{(s)}$  和  $R^{(s+1)}$  即为所求的识别标准矩阵和模糊分类结果矩阵,迭代结束. 否则  $s = s + 1$ ,返回步骤②,继续迭代. ④ 聚类判定:对于确定分类组数  $c$ ,求得模糊分类结果矩阵,可借助分类系数和平均模糊熵判定聚类效果. 分类系数  $F$  越接近 1,平均模糊熵  $H$  越接近 0,聚类识别效果越好. 分类系数和平均模糊熵示下:

$$F(R) = \frac{1}{n} \sum_{k=1}^n \sum_{l=1}^c r_{lk}^2 \quad H(R) = \frac{1}{n} \sum_{k=1}^n \sum_{l=1}^c r_{lk} \ln(r_{lk})$$

**3.3 模糊 ISODATA 方法的应用** 设计的临床调查信息表内容包括研究对象、研究因素和效应指标三个部分. 根据临床采集到的症状数据进行模块化分解,

将患者症状分解为全身症状、头面症状、脉象舌象、男妇科症状、西医实验指标与病理指标等子模块。经过 2005-11/2006-04 在浙江省和湖南省医院肾内科的临床调查,采集到诊断为原发性 IgA 肾病的 117 个合格病例。以其中 20 个病例数据处理过程及结果为例:

① 临床调查数据预处理:对采集数据进行模糊化和标准化处理,其中如脉象中的弦脉“是/否”类型逻辑变量以“1/0”代替,神疲乏力等症状等级变量的“无、轻度、中度、重度”分别以“0, 1/3, 2/3, 1”的[0-1]范围内数值代替,对采集数据进行模糊化和标准化处理。② 辨证分型计算识别:参照中国中医药学会内科学会肾病专业委员会 1996 年通过的《慢性肾炎中医辨证分型标准》和结合 1986 年第二届全国中医肾病专题学术讨论会《慢性肾炎中医辨证分型试行方案》将 IgA 肾病划分为脾肾气虚、肝肾阴虚、气阴两

虚、脾肾阳虚、阴阳两虚 5 种证候及主次症状进行识别。③ 结果分析:经过 20 例病例计算识别,脾肾气虚、肝肾阴虚、气阴两虚、脾肾阳虚、阴阳两虚 5 种证候分别有 1, 4, 4, 6 和 5 例,分类系数 F 和平均模糊熵 H 分别为 0.987 和 0.036,识别分类效果较好。部分特征症状的证型标准隶属度见表 1,5 种证候识别分类结果矩阵隶属见表 2。通过表 1 列出的部分特征症状对 5 种证型的隶属度数值可以看出:畏寒、水肿在脾肾阳虚中隶属度最大,午后低热、五心烦热在肝肾阴虚中隶属度最大,面色少华在脾肾气虚中隶属度最大,腰膝酸痛、神疲、畏寒、五心烦热在阴阳两虚中隶属度最大。这几组隶属度计算结果表明:采用模块化思维和模糊识别的方法,进行的算法所识别判断的结果,与中医辩证诊断的思维大致吻合。

表 1 IgA 肾病特征症状对应 5 种证型的标准隶属度

证型	症状											
	易感冒	午后低热	畏寒	神疲	乏力	面色少华	咽干	气短	腰痛	五心烦热	血尿	水肿
脾肾气虚	0.0	0.0	0.0	0.3	0.3	1.0	0.0	0.0	0.3	0.0	0.7	0.0
肝肾阴虚	0.0	0.2	0.0	0.2	0.3	0.5	0.2	0.0	0.2	0.3	0.6	0.0
气阴两虚	0.0	0.0	0.3	0.2	0.4	0.5	0.5	0.2	0.3	0.2	0.4	0.2
脾肾阳虚	0.2	0.0	0.5	0.2	0.2	0.8	0.4	0.2	0.3	0.1	0.6	0.3
阴阳两虚	0.8	0.2	0.7	0.4	0.7	0.8	0.5	0.3	0.4	0.3	0.8	0.2

表 2 IgA 肾病 10 例对应 5 种证型的分类隶属度

(%)

证型	病例号									
	1	2	3	4	5	6	7	8	9	10
脾肾气虚	0	0	0	100	0	0	0	0	0	0
肝肾阴虚	100	1	0	0	0	0	0	0	0	0
气阴两虚	0	0	0	0	0	100	100	0	100	100
脾肾阳虚	0	2	2	0	100	0	0	0	0	0
阴阳两虚	0	97	98	0	0	0	0	100	0	0

## 4 讨论

基于模块化思维并以模糊熵作为判断识别依据的模糊 ISODATA 中医证候诊断方法,以 IgA 肾病的辨证分型为例,初步探讨了该方法的可行性。通过临床信息的标准化和模块化处理,根据设定标准对临床症状和证候进行处理,同时得出证候分类结果和特征症状隶属度,提高了疾病辩证诊断的效率和准确度,是一种比较切合实际的中医证候诊断方法。我们参照专家委员会关于慢性肾炎的辨证分型标准,对中医四诊信息进行中医证候识别判断。下一步将直接根据数据统计提取特征症状和建立识别规则,并且考虑中医证候与西医理化指标之间的相互关系,从而在中医证候的客观化和规范化研究中进行深入探索。

## 【参考文献】

- [1] 白云静, 申洪波等. 中医证候复杂性特征及证候研究思路探析 [J]. 中国中医药信息杂志 2004, 11(9): 754-756.
- [2] 包含飞. 初议中医学是复杂性科学 [J]. 上海中医药大学学报, 2003, 17(2): 3-6.
- [3] 王阶, 姚魁武. 中医学证候量化诊断的发展现状和思考 [J]. 世界科学技术-中医药现代化 2003, 5(5): 10-14.
- [4] 马斌荣. 中医专家系统与中医知识库-中医领域计算机软件的开发与应用 [M]. 北京: 北京出版社, 1997: 86-98.
- [5] 钱夕元, 邵志清. 模糊 ISODATA 聚类分析算法的实现及其应用研究 [J]. 计算机工程与应用 2004, 15: 70-71.
- [6] 黄健元. 模糊 ISODATA 聚类分析方法的改进 [J]. 南京航空航天大学学报 2000, 32(2): 179-183.