

· 理论研究 ·

辨证理论的隐结构思想*

袁世宏¹ 张连文² 杜彩凤¹ 王天芳¹

(1 北京中医药大学基础医学院中医诊断系 北京 100029; 2 香港科技大学)

摘要:将中医辨证理论和证候的获得过程与隐变量和隐结构分析的思想做了简要的对比分析,探索隐结构法对证候研究的实际意义。证候就像隐类分析一样是通过可见的“症状显变量”来探求内在的不可见的“证候隐变量”,并建立证候隐变量和症状显变量以及证候隐变量之间的隐结构,但是传统中医证候的结构是由人脑从经验而习得的,而隐结构模型的构建则是由电脑基于相应的统计学原则,通过对临床流行病学所收集的症状数据分析所构建的。此模型不仅符合中医辨证理论,而且具有定量化的优势,可作为中医证候客观化和定量化的研究方法之一。

关键词:中医辨证;证候;隐变量;隐结构

中图分类号:R241

Concealed structure thought in TCM syndrome differentiation theory*

YUAN Shi-hong¹, ZHANG Lian-wen², DU Cai-feng¹, WANG Tian-fang¹

(1 Department of TCM Diagnostics, School of Fundamental Medicine, Beijing University of Chinese Medicine, Beijing 100029; 2 Hong Kong University of Science and Technology)

Abstract:TCM syndrome differentiation theory, syndrome acquisition process and latent variables, concealed structure thought were compared and analyzed briefly in this paper. The practical value of concealed structure method to syndrome study was discussed. The syndromes, as Latent Class Analysis (LCA), search internal invisible “syndrome latent variables” through visible “symptom significant variables”, and to set up syndrome latent variables, symptom significant variables and a concealed structure among syndrome latent variables. But the syndrome structure of TCM is acquired from the experience of human brain, and concealed structure model is established by computer based on relevant statistical principle through the analysis of symptom data collected from clinical epidemiological survey. The model is accorded with TCM syndrome differentiation theory with quantitative advantage, and can be thought as one of research methods of TCM syndrome objectification and quantification.

Key words:TCM syndrome differentiation; syndromes; latent variables; concealed structure

中医的证候与社会科学、经济学、心理学等学科中的政治倾向性、效益、智力这类概念类似,至今仍然难以直接观察和度量。尽管有人试图用物理和生化的微观指标使“证候”直观可见,但这种尝试尚未成功。越来越多的学者致力于基于流行病学临床调查数据基础上,用数理统计方法研究中医证候,不过由于现有数据分析方法在处理中医临床数据时存在这样或那样的问题^[1],因此我们提出用隐类分析思

想及隐结构方法尝试构建中医证候量化结构。

1 隐结构法与证候的理论关联

1.1 隐变量的含义

隐变量(Latent Variable)是有关隐结构分析方法中一个重要概念,它有多种含义^[2],我们倾向于认为隐变量是一个确实存在但又不能被精确地解释和说明的在语义水平上的理论结构或想象的东西。这种“虚构实体”却有助于更好地了解和把握可观

袁世宏,男,在读博士生,教授

* 国家重点基础研究发展计划(973 计划)资助项目(No. 2003CB517101),香港研究资助局项目(No. 622015)

察世界的固有复杂性及其规律,如思维能力、分析能力、社会效益等。换句话说,“隐变量”指代不能直接度量或观察的“存在实体”,“显变量(Manifest Variable)”指代可直接度量或观察的事物或现象。

带有隐变量的模型一般称为隐结构模型(Latent Structure Model),其共同特征是:基于隐变量及其可观察变量值之间关系的特定假设作为研究的出发点。由于研究领域、探讨内容的不同,使得所研究问题的变量性质千差万别,不可能用一种方法解决所有问题,因此该类方法有不同的具体形式。其中,隐类分析(latent class analysis, LCA)及其推广的多层隐结构形式,适合于隐变量和显变量都取离散值的情况。

1.2 隐结构

隐变量与显变量之间以及隐变量与显变量之间存在的关联关系构成了一个隐结构(Latent Structure)。如果用一个树状图来形象地表示这个结构的话,树状图的叶端代表显变量,根端代表隐变量。另外在叶端和根端之间还存在着中间联系的隐变量(子隐变量)。自然界乃至我们的周围存在着大量的这种结构。如中学生的数学、物理、化学、语文、历史、英文等各科成绩是可以直接度量和考察的,可认为是显变量,而影响不同学科成绩的分析能力和文字能力以及最后的智力水平则是不能直接考察和度量的,可称之为隐变量。这样智力水平、分析能力、文字能力和中学生的数学、物理、化学、语文、英语等课程的成绩之间形成一个隐结构^[3]。

1.3 辨证的探“隐”过程

中医理论和中国传统思想一样以“关系”作为认识整个世界的主要视角,其理论体系的内容在很大程度上是对生命复杂关系的系统表述^[4],其中辨证理论就是对各种病理关系进行表述所形成的证候与症状的隐结构理论。

第一,就中医对证的认识过程而言,它是通过对可见现象的长期观察发现规律而得到的。如《灵枢·刺节真邪》曰:“下有渐洳,上生苇蒲,此所以知形气之多少也。”讲的就是通过对地面上可见的芦苇、葛蒲丛生繁茂与否的现象长期观察,发现规律、推测并引入地下水湿润之气的隐藏因素,即用土质的肥瘠与水湿多寡来解释地表的植物生长现象。同理由人体外在形体强弱表现状态也可以推知内在气血的盛衰。在长期的临床实践中,医学家发现了大量的症状具有相对稳定的伴随或关联关系,为解释这些关联规律,引用“证”这个不可观察的概念。证候就是通过对症状显现规律的总结和抽象而构建的。

第二,辨证是对四诊所获取的全部临床资料进行分析,发现和找出最能反映疾病状态的一组症候群之间的关联规律,进一步推测和确认与此关联规律相对应的病理本质(病因、病位、病性、邪正对比等)。一组症状所呈现的关联性变化是由于背后更为本质的“证”所引起的,一个证候肯定会产生一组相关联的症状群。

第三,证是古人在中国古代特定的社会人文背景影响下,根据“以表知里”“以外测内”的认识论原则,基于外现的相关生命现象,从整体上把握人体内外各部的联系,对收集的各种生理病理现象进行逻辑推理而形成的一种病证思维模型^[5]。这种思维模型是从病因、病性、病位、邪正关系等几个方面对疾病发展过程中某一阶段病理本质的概括,是由抽象而归纳出来的一种“不能直接度量”的理念上的“实体”。辨证就是在中医理论指导下,通过对病人由四诊直接获取的各种“可见的”临床资料进行分析、综合,从而对疾病当前的本质做出判断,并概括为一个完整“证”名的诊断思维过程。

综上所述,中医辨证实际上就是通过对外在可见的症状之间关联性的分析,去推寻导致这种关联性的内在不可见的“证”,建立中医证候的结构和体系。

1.4 证的隐结构形式

就隐结构而言,“证候”是隐性因子(隐变量),是通过对临床表现的一组症状(显变量)的分析来确认和把握。就隐结构树图而言,叶端代表症状,根端代表证型,而中间隐藏的其他节点代表更多中间层的其他病理因子。见图 1。

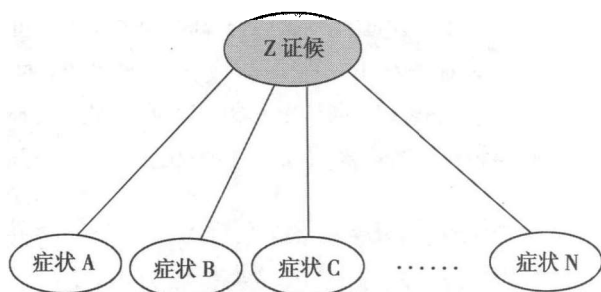


图 1 中医证候的隐结构基本模型

实际上证的隐结构模型要比这复杂的多。例如关于肾阴虚,中医认为:“肾阴亏虚,脑髓、官窍、骨骼失养,则见腰膝酸痛,眩晕耳鸣,健忘,齿松发脱;……虚火上扰心神,故烦热少寐;肾阴不足,失于滋润,虚火蕴蒸,故见口燥咽干,形体消瘦,潮热盗汗,或骨蒸发热,颧红,尿黄少。舌红少苔或无苔,脉细数,为阴虚内热之象。”这里,诸如“腰膝酸痛”“舌红少苔”、

“脉细数”等症状是通过四诊直接观测到的,类似于中学生的学习成绩,可称之为显变量;而诸如“肾阴虚”“肾阴虚失养”“肾阴虚失滋润”“阴虚内热”等病理因子,迄今没有办法对它们进行直接观测,类似于人的分析能力、智力水平,可称之为隐变量。中医辨证讲述的就是证候(隐变量)与症状(显变量)之间的关系,据此关系可以构建中医证候的隐结构模型。这与隐类分析通过一组相关联的显变量的概率分布来刻画隐变量的目的相似。

1.5 辨证理论隐结构的局限与发展

辨证理论虽然具有类似隐结构模型思想和结构,但由于众所周知的背景,在今天来看具有难以克服的局限性。这种局限性主要表现为缺乏客观性和定量性。而缺乏客观性的根源主要有:①辨证理论依据的资料来源于病人的主观感受和或医生对这些资料的主观感觉,缺少客观的度量标准;②这些理论大多是个体经验的积累和升华,不同的实践经验、专业知识、认识水平对证的辨识也不尽统一;③临床辨证结论的评价也缺乏客观的标准。

缺乏定量性的根源主要因为:①证的获取带有很浓的主观思辨色彩,而主观感觉具有不确定性和模糊性,对症状的轻重程度、症状的主次关系、症状和证候的关联关系、证候与病的关系等均缺乏严格的量化表述;②证候描述的是一个十分复杂的生命巨系统,具有复杂性、动态性、不确定性等特点,也使得很难进行量化研究。近几十年来,虽有人对证候做了大量的量化研究,但由于研究方法的局限性,尚未取得理想的结果^[6]。

中医辨证理论虽然也是通过可观察的症状来构建证候结构,但它是无数的医学家人脑通过对众多事例的经验性观察,发现规律来构造证候结构,并用此结构来解释症状表现的规律指导辨证,其缺乏客观性、难以量化就在所难免。如果改用电脑来构造证候隐结构并指导辨证,客观性就会大大提高,量化问题也迎刃而解。由此启发我们用计算机,对多中心大样本的临床流行病学调查所收集的经过规范化处理的症状数据进行分析,用隐结构的思想和方法寻找临床症状数据中的规律,构造证与症状之间的隐结构,为辨证客观化、量化提供新的研究思路和方法。

2 隐结构模型对“证”研究的实际意义

2.1 目标的一致性

中医辨证通过一组相互关联的可见症状变量来探寻内在“证候”的存在以及性质。回归分析、相关

分析、关联模型分析和相关模型分析等,也能展示显变量之间关系/影响的存在,但这些方法只是说明2个可观察变量之间有关系,或彼此有影响,且这种关系或影响有统计学意义。但只有隐类模型才可以检视这种有统计学意义的可见关系或可见影响是否是真实的^[7]。通过隐类分析才知道,症状显变量之间的关联只是隐变量“证”的“赝品”。如临床上常见畏寒和肢冷具有明显的关联关系,但当确定“阳失温煦”隐变量后,发现阳失温煦才是解释畏寒和肢冷之间关联性的惟一因素,原来存在于畏寒和肢冷的关联只是阳失温煦隐变量的“赝品”。

聚类分析也可以通过分析相互关联的一组症状变量的相似性,而将他们聚为一类。但是聚类方法只是简单地把密切相关的显变量聚在一起,并没有引入新的变量,也就是说只是简单地展示显变量的相关性以便看起来直观容易而已(没有解释为什么)。在隐类模型中,引入新的隐变量来解释显变量之间的相关性,或者说找出显变量相关的背后原因。另外,聚类的类所决定的仅仅是基于两两相关而已,隐类分析中隐变量解释了全部的相关,既包括两两变量的相关,也包括多对多的相关。

因此,中医是通过相互关联的一组可见症状来探求内在的不可见的“证候”,隐类分析是通过外在可见的相互关联的显变量,来寻找内在的引起这种关联的“隐变量”,两者的目标是一致的。这是隐结构分析方法适用于中医证候研究的最重要的理论优势。

2.2 变量性质的适用性

现存用于对隐结构分析的工具,按隐变量和显变量的特性分为4类:第一类主要有因子分析(factor analysis)、主成分分析(principal component analysis)和独立成分分析(independent component analysis),其特点是隐变量和显变量都取实数值,主要用于实数值数据的降维。而且这3个工具均假设隐变量之间相互独立,而中医的证候之间存在广泛联系。第二类包括隐概况分析(latent profile analysis),其特点是显变量取实数值(连续变量),隐变量取离散值(离散变量),目前其应用并不多见,中医症状变量目前也不太可能取连续值。第三类包括隐特性分析(latent trait analysis),多用于确定诸如智力(IQ)这样的隐特性,其特点是显变量取离散值,隐变量取实数值。第四类主要指隐类分析(latent class analysis, LCA),其特点是隐变量和显变量都取离散值。

就证候研究的数据性质而言,证候作为变量显然是离散的而非连续的。尽管类似肝气郁结、气郁

化火、肝火上炎这样的证候之间有连续发展演变过程,但肯定不能用实数性质的连续变量来表示,更不能视为正态分布。症状变量定量化规范,一般都只是根据“有/无”,或“无”“轻”“中”“重”等,而记为“0/1”,或者 0、1、2、3 分值等半定量形式。针对中医症状和证候半定量化取值特点,其变量性质基本属于离散变量,因此隐类分析工具比较适合中医证候的研究。

2.3 局部独立规则的意义

隐类分析的基本思想是通过分析对显变量的分析,探索显变量所显示出的关联规律,并用隐变量来解释这种规律。当隐变量被确定后,原来存在于显变量之间的关系则消失,即当隐变量确定后,对于每一个隐类而言,显变量之间相互独立,它们之间的关系惟一由隐变量来解释^[8]。这种假定被称为局部独立(Local independent)规则,是隐类分析的一个重要规则。

局部独立规则可以确定在一组可观察变量之间的关系是否因某些不可度量的变量所决定和影响的。当发现一组变量的关联由于“另外”一个变量的出现而表现出局部独立的话,可以认为,这个“另外”的变量“解释”了那些可观察变量之间的关系,也代表了变量的真实关系。这种规则可以帮助确认一组相关症状之间的伴随出现和关联关系,是否存在一个不可观察的“病机/证候”因子。

在临床实践中,经常会注意到畏寒怕冷、四肢冷、腰背冷这几个症状有密切的关联关系,当出现腰背冷症状时,患者也常会有不同程度的四肢冷和畏寒怕冷,反之亦然。同样,便秘和完谷不化也常显示出密切的相关性。当依据中医理论辨证后,发现便秘和完谷不化都可以用肾阳虚衰不温脾土这个证候(隐变量)而解释,即当确定肾阳虚衰不温脾土证候后,原来便秘和完谷不化之间的相互关联关系消失了,并不是便秘引起完谷不化,也不是完谷不化导致便秘,两者的出现均是由于肾阳虚衰所引起。同样,肾阳虚失温煦才是引起畏寒、肢冷、腰背怕冷这 3 个症状不同程度出现的内在原因。

图 2 是我们研究肾虚证候模型的一个局部。从所示模型局部结构可见,Y1 代表肾阳虚失温煦的隐变量,而 Y2 代表肾阳虚衰不温脾土的隐变量。这两个隐变量可用来解释畏寒怕冷(X2)、四肢冷(X3)、腰背冷(X4),以及完谷不化(X1)、便秘(X0)之间存在的内在联系^[9]。

2.4 模型概率对证的量化解

隐类分析和传统聚类分析的最大区别在于,前

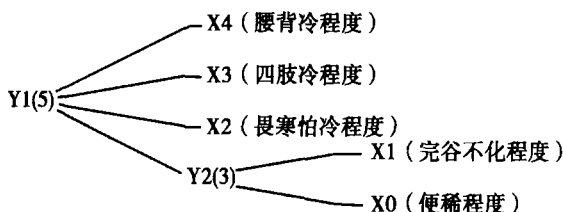


图 2 肾阳虚失温煦和不温脾土隐变量与相关显变量

者是基于模型的多维聚类,后者是基于相似性/距离的一维聚类。基于模型的聚类有 2 个基本参数,一是隐类概率(Latent Class Probabilities),一是条件概率(Conditional Probabilities)。模型通过这 2 个概率指标来实现中医证候解释的量化。见图 3。

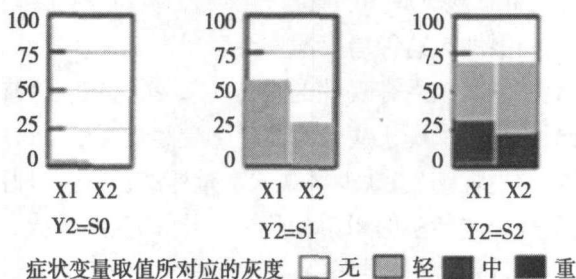


图 3 肾阳虚衰不温脾土隐变量的概率图

隐类概率描述隐变量不同类(水平)的分布,在这一类中可观察值是相互独立的,其有 2 个重要的方面:即隐类的个数和类的相对大小。类的个数表示由隐类模型定义的隐类(水平)有多少,从图 2 和图 3 可见,该模型就隐变量 Y2 而言,将整个老年人群(原始数据来源于老年人群)分成 3 类,即 Y2 = S0、Y2 = S1 和 Y2 = S2。由于一个隐变量的每一类中某一症状不同程度表现的概率之和等于 1,在 Y2 的 3 个类中,便秘程度和完谷不化程度 2 个变量的无、轻、中、重的不同表现程度的概率总和均为 1。这一参数可以帮助找到一组症状背后的病机/证候,以及其可能表现的不同类型。

条件概率是度量一个观察变量和一个隐类关联程度,表示在隐变量的某个类(水平)显变量取某特别值的概率大小,它可以刻画和认识每一隐类的特性及其隐变量的本质属性。比如通过图 3 和表 1 可以看到,在 Y2 = S0 这一类,便秘和完谷不化这 2 个症状不出现的条件概率高达 0.947 2 和 0.970 4,其表现为中度和重度的概率几乎为 0,因此如果从肾阳虚衰不温脾土这个维度来划分老年人群的话,可以认为这一人群的隐类特征是没有肾阳虚衰不温脾土的病理变化。当 Y2 的取值从 S0 变成 S1 和 S2 时,2 个症状出现的概率以及程度逐步增加,当 Y2 = S2 时,便秘症状中度发生概率为 0.293 1,重度发

表 1 肾阳虚衰不温脾土与便秘和完谷不化之间的概率(定量)关系

症状程度	$P(X0 Y2=S0)$	$P(X1 Y2=S0)$	$P(X0 Y2=S1)$	$P(X1 Y2=S1)$	$P(X0 Y2=S2)$	$P(X1 Y2=S2)$
无	0.947 2	0.970 4	0.422 1	0.696 0	0.422 1	0.309 0
轻	0.043 5	0.027 6	0.569 0	0.304 0	0.380 6	0.461 8
中	0.009 3	0.002 0	0.005 1	0.000 0	0.293 1	0.210 5
重	0.000 0	0.000 0	0.003 8	0.000 0	0.021 9	0.018 7

注: P 代表概率; $Y2$ 代表肾阳虚衰不温脾土; $X0$ 代表便秘程度; $X1$ 代表完谷不化程度; $S0$ 、 $S1$ 、 $S2$ 分别代表肾阳虚衰不温脾土证候的无、轻、中、重 3 类人群

生概率 0.021 9;完谷不化症状中度发生概率 0.210 5,重度发生概率 0.018 7。这 2 个类的特征则分别是中度和重度肾阳虚衰不温脾土。因此,依据 2 个症状显变量在每一类的条件概率,将 $Y2$ 隐变量 3 类特性分别定义为“无肾阳虚衰不温脾土”“轻度肾阳虚衰不温脾土”和“重度肾阳虚衰不温脾土”。另外,从模型结构可见,老年人群还可从肾阳虚失温煦维度分成 5 类,等等。隐类分析是同时对数据进行多维聚类,这是该方法的优势也是与传统聚类分析最大的差异之一。

这种概率分布的定量化解释不仅与中医理论的诠释和临床实践体验基本相似,而且与专家辨证也非常吻合^[10-11]。

可见这种概率模型利用这些定量化的数据,不仅对证候轻重程度等级有明确的量化说明,而且对每一类别均给予详细的量化性的定义和解释。对于定量地认识证候,以及证候与症状的关系均具有显著的意义。然而,隐类分析的隐类模型只包含一个隐变量,这对中医证候研究显然是不够的,因为中医的证候大都涉及众多隐变量。要应用隐类模型就必须对其进行必要的推广。张氏^[12]专门对此提出多层隐类分析模型(hierarchical latent class model, HLC 模型),并用推广后的模型对中医肾虚数据进行辨证试用,效果良好。

3 小结

从中医辨证理论形成的文化背景和临床证候的经验习得,都显示出中医辨证和隐类分析在思想上具有非常相似之处。证候这个“实体”虽然无法从生物物理学的“物质”角度去把握,但可通过对大样本临床流行病学收集的症状数据的挖掘和分析,来获得对证候规律性的量化认识。隐类分析的思想以及在此基础上的隐结构模型在挖掘这些数据,探索中医证候的规范化和定量化研究中具有非常良好的适用性。

参考文献:

[1] 袁世宏,王天芳. 多元统计方法在建立证候诊断模型研

究中存在问题的思考[J]. 北京中医药大学学报,2004, 27(4):9-11.

[2] ALEXANDER VON EYE, CLIFFORD C CLOGG. Latent Variables Analysis: Applications for Developmental Research [M]. USA: Sage Publications Inc,1994: 20-22.

[3] 张连文,袁世宏. 隐结构模型与中医辨证研究(I)——隐结构法的基本思想以及隐结构分析工具[J]. 北京中医药大学学报,2006,29(6):365-369.

[4] 王小平. 关系——证本质研究的新视点[J]. 医学与哲学, 2001,22(12):51-53.

[5] 李翠娟,烟建华,巩振东. 对中医证本质研究的反思[J]. 中国中医基础医学杂志, 2005,11(4):271-272.

[6] 袁世宏,张连文,王米渠. 数理统计思想及方法在辨证规范化研究中的应用及思考[J]. 中国中医基础医学杂志,2003,9(4):15-18.

[7] GOODMAN L A. Latent Class Analysis: The Empirical Study of latent Types latent variables and latent structures [M]// HAGENAARS J A, MCCUTCHEON A L. Applied Latent Class Analysis. United Kingdom: Cambridge University Press, 2002:3-11.

[8] WILLIAM R DILLON, MATTHEW GOLDSTEIN. Multivariate Analysis: Methods and Applications [M]. Canada: John Wiley & Sons Inc, 1984:24-25.

[9] 袁世宏. 肾虚证候的隐结构模型研究[D]. 北京:北京中医药大学,2006.

[10] 张连文,袁世宏,陈 弢,等. 隐结构模型与中医辨证研究(II)——肾虚数据分析[J]. 北京中医药大学学报,2008,31(9):584-587.

[11] 袁世宏,张连文,陈 弢,等. 隐结构模型与中医辨证研究(III)——模型辨证与专家辨证[J]. 北京中医药大学学报,2008,31(10):659-663.

[12] ZHANG N L, YUAN SHI HONG, CHEN TAO, et al. Latent Tree Models and Diagnosis in Traditional Chinese Medicine [J]. Artificial Intelligence in Medicine, 2008, 42: 229-245.

(收稿日期:2008-07-10)