

• 理论研究 •

隐结构模型及其在中医证候研究中的应用^{*}

王天芳¹ 张连文² 赵 燕¹ 王 燊² 吴秀艳¹ 袁世宏¹ 王智瑜¹

杜彩凤¹ 徐雯洁³ 于春光¹ 陈 弢² 潘建文² 王庆国¹

(¹北京中医药大学基础医学院 北京 100029; ²香港科技大学; ³北京市中医院)

摘要:用隐结构法研究中医证候的基本思想是运用隐结构模型对临床流行病学调查获取的、未经过医生事先辨证的症状数据进行多维聚类分析,为证候的规范化、标准化研究提供较为客观的依据。在系统阐述隐结构分析及隐结构模型诠释的原理和工具的基础上,提出了隐结构法应用于中医证候研究的意义及步骤,并以抑郁症的隐结构模型为例,示范了隐变量和隐类诠释的方法和步骤,以期今后从病证结合角度进行类似的研究提供参考。

关键词:隐结构模型;证候;抑郁症

中图分类号: R241

Latent structure models and their application in TCM syndrome research^{*}

WANG Tian-fang¹, ZHANG Nevin-lianwen², ZHAO Yan¹, WANG Yi², WU Xiu-yan¹, YUAN Shihong¹,
WANG Zhi-yu¹, DU Cai-feng¹, XU Wen-jie³, YU Chun-guang¹, CHEN Tao², POON Kinman²,
WANG Qing-guo¹

(¹ School of Preclinical Medicine Beijing University of Chinese Medicine Beijing 100029; ² Hong Kong University of Science and Technology; ³ Beijing Hospital of Chinese Medicine)

Abstract: The basic idea to apply latent structure models in TCM syndrome research is to first collect data on the occurrence of symptoms (but without TCM syndrome diagnosis) through epidemiological survey and then perform multidimensional clustering analysis on the data using latent structure models. The objective is to provide an objective and quantitative basis for the standardization of TCM syndrome differentiation. In this paper we systematically explain the conceptual advances on the underlying principles and potentials of the approach and the technical advances on the interpretation of latent structure models. The exposition is in terms of a study on depression so as to provide an example of reference for future research.

Key words: latent structure models; syndrome; depression

证(又称证候)是疾病过程中某一阶段(时点)机体对内、外致病因素做出的综合反映,在宏观上表现为一组特定症状(包括自觉症状和体征),反映了疾病的病因、病位、病性以及邪正相争的状况和趋势,是临床论治的前提和依据。临床中对于证候的辨识和判断,主要是依据医者对于病人症状表现的理解与分析,带有很强的主观性,因此具体到某一证候到底应有哪些表现,以及具备哪些表现可以被判

断为某一证候,也就是被大家认可的证候的规范和诊断标准的建立,一直是中医研究中的一个难点和关键的问题。

症状是病人患病后所表现出来的异常表现,是医生通过望、闻、问、切四诊所获取的疾病信息,它们是可以被直接观察到的,因此称为显变量。而病人所患疾病的具体病理机制及所属证候,则需要医者对病人的症状表现进行一系列分析而得出(该过程

王天芳,女,博士,教授,博士生导师, E-mail: wangtf@bucm.edu.cn

^{*} 国家重点基础研究发展计划(973计划)资助项目(No. 2003CB517106, No. 2003CB517101),香港研究资助局项目(No. 622307, No. 622408)

称为辨证), 相对而言不是直接被观察到的, 而是通过症状(显变量)来间接把握的, 因此称为隐变量。抽象地说, 中医辨证学所描述的是隐变量(证候、病机等)与显变量(症状), 以及隐变量与隐变量之间的关系, 在这种意义下, 可以将中医的辨证学理解为一个用自然语言描述的隐结构模型, 以下称之为中医隐结构模型。

一般来说, 一个变量代表一种对事物的划分。如性别这个变量代表将地球上的人划分为男人和女人两个类。中医中的隐变量代表的是从医学角度对人群进行划分。以肾阳虚为例, 如果谈论有无肾阳虚, 在概念上则是将人群划分为有肾阳虚和无肾阳虚两个类。如果在有肾阳虚的情况下, 进一步谈论其轻、中、重, 在概念上就是将人群划分为无肾阳虚、有轻度肾阳虚、有中度肾阳虚和有重度肾阳虚四个类。中医中有众多隐变量, 每个隐变量都代表一种划分。在这种意义下, 可以将中医辨证学理解为是基于症状、从各种角度对人群进行了多维划分。因此, 隐结构和多维划分是中医辨证学两个不可分割的特征。

中医辨证所面临的关键问题是其划分边界缺乏客观的、定量的依据, 难以建立统一的辨证标准。隐结构法^[1-3]的目标就是要解决这个问题。其基本思想是运用隐结构模型对大样本的临床流行病学调查获取的、未经过医生事先辨证的症状数据, 按照统计学原则进行分析, 获得统计学隐结构模型。该模型是基于数据本身的特征、从各种角度对人群进行多维划分的结果, 其反映的是蕴含在数据中的规律, 而不是某一个或几个医者的主观认识, 具有较强的客观性和定量性, 因此可以作为建立中医辨证规范和标准的一个重要依据。

1 隐结构分析的工具

目前, 隐结构法使用的数据分析工具是隐树模型(latent tree model)。隐树模型是一种树状贝叶斯网^[9], 以前称为多层隐类模型(hierarchical latent class model)^[7]。图 1 给出了一个隐树模型的例子, 其中叶节点代表显变量, 其他节点代表隐变量。例子中, 高考学生的数学成绩、理综成绩、语文成绩和文综成绩是显变量, 而智力、分析能力和语言能力则是隐变量。从分析能力到数学成绩有一个箭头, 表示数学成绩直接依赖分析能力, 具体依赖情况通过表 1 中的条件概率而定量刻画。表 1 中第 1 行是说, 分析能力低的学生在数学科有 0.5 的概率不及格, 0.4 的概率及格, 0.1 的概率得良, 而得优的概率则是 0。其他 2 行的含义类推。模型中的其他箭头

代表其他变量之间直接依赖关系, 每个箭头都有相应的条件概率分布。

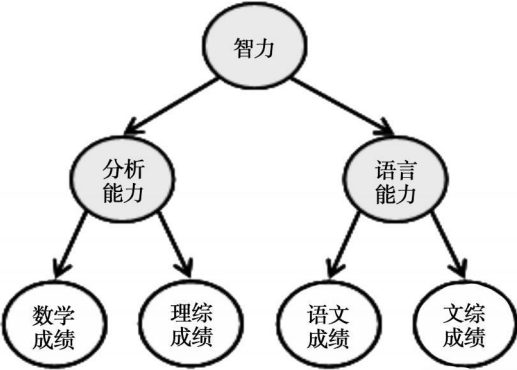


图 1 隐树模型示意图

表 1 条件概率分布[P(数学成绩 | 分析能力)]

分析能力	数学成绩			
	不及格	及格	良	优
低	0.5	0.4	0.1	0
中	0.1	0.4	0.4	0.1
高	0	0.1	0.4	0.5

只包含一个隐变量的隐树模型称为隐类模型(latent class model)。隐类模型用于对数据进行聚类分析, 也就是单维划分。在以往西医的有关研究中, 曾有学者用其建立诸如类风湿关节炎、抑郁症、慢性疲劳综合征等病症的诊断标准^[8-9]。

隐结构法的第一步, 是用隐树模型对无监督数据进行分析, 目前这一步使用的算法是扩调减算法(Expansion Adjustment Simplification until Termination, EAST)。EAST 在一个称为贝叶斯信息准则(Bayes information criterion, BIC)^[10]的统计学原则指导下, 在隐树模型空间中进行搜索。BIC 准则要求模型与数据尽量紧密地拟合, 但其复杂度又不能过高。在搜索过程中, 模型与数据的拟合程度不断改进, 从而 BIC 分逐步增加。当模型变得太复杂时, BIC 不升反降, 这时搜索过程停止。

相对于隐类模型来说, 隐树模型已经做了很大推广, 但是树结构仍然有其局限性, 它有时会使得模型不能很好地拟合数据。为了让模型与数据的拟合程度尽量的高, 在隐树分析结束后, 需要对所得的隐树模型进行加边(即加箭头)。目前使用的加边算法叫 LTM-EA, 它是在 BIC 准则指导下的另一个搜索过程, 每一步加一条边。在高考学生成绩的例子中, 对图 1 所示的模型进行加边, 可能会得到图 2 所示的模型。这个模型不再是树状的, 是一个比隐

树模型更复杂的贝叶斯网。

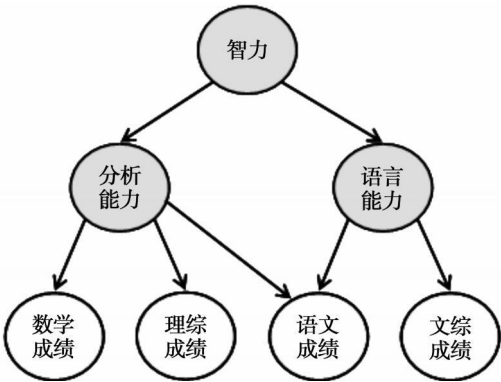


图 2 隐树模型加边示意图

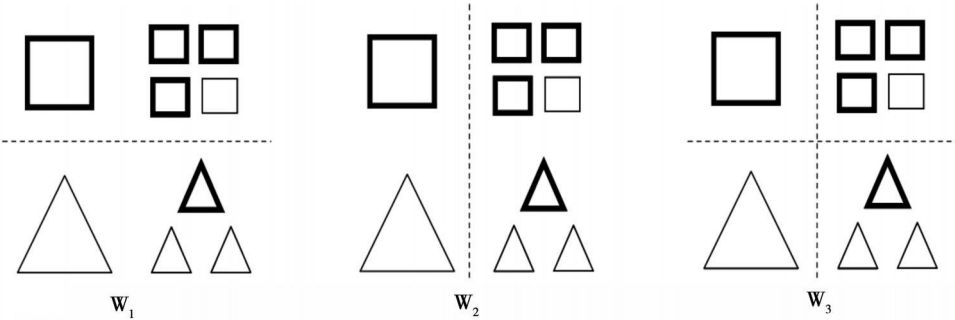
2 隐结构模型诠释的原理和工具

隐结构分析的结果是一个隐结构模型，其中一般包含多个隐变量。隐结构模型中的每个隐变量代表一个对数据样本的划分，隐变量的每个取值对应划分中的一个类，称为隐类。隐结构模型诠释就是

要基于应用领域知识，把握这些划分以及各个隐类的含义。

2.1 把握划分的含义

把握一个划分的含义，就是要看看划分使用的主要标准是什么？为此，需要分析其中各个类之间的区别主要在哪些方面。作为例子，假设有智力游戏机给出如图 3 所示的、关于几何图形的三个划分，并且问：每个划分所使用的标准是什么？第一个划分 (W_1) 由两个类组成，直观上它们的区别主要在图形的形状，上面的类中全部是正方形，下面的类全部是三角形，所以 W_1 是一个基于图形形状的划分。第二个划分 (W_2) 也是由两个类组成的，直观上它们的区别主要在图形的大小，左边类中的图形大，右边类中的图形小，所以 W_2 是一个基于图形大小的划分。第三个划分 (W_3) 由四个类组成，它们在图形的形状和大小两方面都有区别，直观上 W_3 是一个基于图形形状和大小这两个因素的划分。



W_1 为基于图形形状的划分 W_2 为基于图形大小的划分 W_3 为基于图形形状和大小的划分

图 3 几何图形的划分示意图

在图 3 中对 W_1 划分的诠释应注意的第一个基本问题是：划分 W_1 中的两个类，除了在图形形状方面有区别以外，在线条粗细方面也有区别，上面类中的线条多数较粗，而下面类中的线条多数较细。那么，为什么在图形形状方面的区别是主要的，而在线条粗细方面的区别是次要的呢？一般地讲，在一个划分中，不同类之间在多个方面可能都有区别，哪些方面的区别是主要的，哪些方面的区别是次要的？怎样判断？这个问题可以利用互信息来回答。设 Y 是一个隐变量， X 是一个显变量， Y 和 X 的互信息 $I(Y; X)$ 度量的是这两个变量之间关系的紧密程度^[11]。从另外一个角度看，变量 Y 代表一个划分，划分中的各个类之间有区别，而且区别是多方面的， $I(Y; X)$ 可以用来度量各个类在变量 X 这方面的区别，互信息越高，在 X 这方面的区别就越大。如果在所有显变量中， X 与 Y 的互信息最大，那么， Y 划分中各个类在 X 这方面的区别就是主要的，而在其

他方面的区别就是次要的。

在上述几何图形划分的例子中，把变量分别简记为图形形状 (XZ)、图形大小 (DX) 和线条粗细 (CX)。通过计算可以得到： $I(W_1; XZ) = 0.69$ ， $I(W_1; DX) = 0$ ， $I(W_1; CX) = 0.16$ 。这表明， W_1 与 XZ 的互信息大于它与 CX 的互信息。所以划分 W_1 中的两个类在“图形形状”方面的区别确实是主要的，而在“线条粗细”方面的区别确实是次要的。

在图 3 中对 W_1 划分的诠释应注意的第二个基本问题是：虽然划分 W_1 中的两个类在“图形形状”方面的区别相对为大，但是它们在“线条粗细”方面也有明显区别。那么，在诠释 W_1 时，除了考虑“图形形状”，是否需要同时考虑“线条粗细”呢？为了回答这个问题，计算 W_1 与 XZ 和 CX 这两个变量间的互信息，结果是： $I(W_1; XZ, CX) = 0.69$ ，这与 $I(W_1; XZ)$ 相同。这就是说，划分 W_1 中的两个类在 XZ 这一个方面的区别与它们在 XZ 和 CX 这两个方

面的区别是等量的,所以只需要考虑在 XZ 这方面的区别。实际上, W_1 与 XZ, CX 和 DX 这三个变量间的互信息 $I(W_1; XZ, CX, DX)$ 也是 0.69。这表明“图形形状”这一个因素已经反映出了划分 W_1 中两个类的所有区别,因此在诠释 W_1 时,只需考虑“图形形状”,而不需要同时考虑其他因素。于是,可以将 W_1 诠释基于“图形形状”这个因素的划分。同样,在诠释划分 W_2 时,由于 $I(W_2; DX) = 0.53 = I(W_2; DX, CX, XZ)$,只需要考虑“图形大小”这个因素。于是,可以将 W_2 诠释基于“图形大小”这个因素的划分。

在诠释划分 W_3 时,由于 $I(W_3; XZ) = 0.69 < 1.21 = I(W_3; DX, CX, XZ)$ 和 $I(W_3; DX) = 0.53 < 1.21 = I(W_3; DX, CX, XZ)$,只考虑“图形形状”或者只考虑“图形大小”是不够的,需要同时考虑多个因素。由于 $I(W_3; XZ, DX) = 1.21 = I(W_3; DX, CX,$

XZ),同时考虑“图形形状”和“图形大小”这两个因素就足够了。于是,可以将 W_3 诠释基于“图形形状”和“图形大小”这两个因素的划分。

2.2 信息曲线与信息覆盖度

在对一个隐变量 Y 所代表的划分进行诠释时,需要找出与 Y 关系最密切的显变量,从而确定划分中的各个类在哪些方面的区别是主要的。为了方便找到与 Y 密切相关的显变量,引入两两互信息曲线的概念。计算 Y 与每一个显变量的互信息,并且按照其大小把显变量排序为 X_1, X_2, \dots, X_n ,其中 Y 与 X_1 的互信息最大,与 X_2 的互信息次之,等等。将 Y 与各个显变量的互信息 $I(Y; X_1), I(Y; X_2), \dots, I(Y; X_n)$ 依次描记在一个坐标里,并用一条线把它们连接起来,就得到 Y 的两两互信息曲线。图 4 给出了划分 W_1, W_2 和 W_3 的两两互信息曲线。

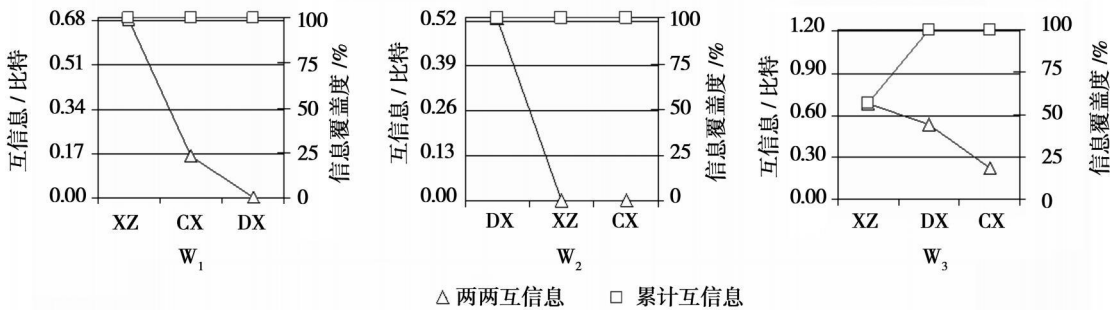


图 4 划分 W_1, W_2 和 W_3 的两两互信息曲线和累计互信息曲线

在对一个隐变量 Y 所代表的划分进行诠释时,需要考虑显变量是否反映出隐类之间的全部区别,从而决定诠释需要几个显变量。因此,引入累计互信息和信息覆盖度的概念。对在 1 和 n 之间的任何一个整数 i 计算 Y 与前 i 个显变量的互信息 $I(Y; X_1, X_2, \dots, X_i)$,并把它们也描记在坐标里,用线连起来,就得到 Y 的累计互信息曲线。划分 W_1, W_2 和 W_3 的累计互信息曲线见图 4。另外, $I(Y; X_1, X_2, \dots, X_i)$ 与 $I(Y; X_1, X_2, \dots, X_n)$ 的比,即 $\frac{I(Y; X_1, X_2, \dots, X_i)}{I(Y; X_1, X_2, \dots, X_n)}$,称为前 i 个显变量对 Y 的信息覆盖度。为了理解这个概念,考虑 Y 划分中各个隐类之间的区别, $I(Y; X_1, X_2, \dots, X_i)$ 度量的是前 i 个显变量反映出来的区别,而 $I(Y; X_1, X_2, \dots, X_n)$ 度量的是所有显变量反映出来的总区别,两者之比就是前 i 个显变量所反映出来的区别占总量的百分比,因此称为信息覆盖度。如果信息覆盖度是 1 或接近 1,那么前 i 个显变量就反映出了各个隐类之

间全部或者几乎全部的区别,从而可以基于它们进行诠释。

作为使用信息曲线和信息覆盖度的例子,察看图 4 中的 3 幅图。左边的图清楚显示,与 W_1 关系最密切的变量是“图形形状”,而且它对 W_1 的信息覆盖度已达到 100%,所以可以将 W_1 视为基于“图形形状”的划分。同样,中间的图明确显示, W_2 可以视为基于“图形大小”的划分。右边的图显示,与 W_3 关系最密切的变量是“图形形状”。但是,由于这个变量对 W_3 信息覆盖度只有 53%,不能够把 W_3 视为仅仅基于“图形形状”这一个因素的划分,还需要考虑其他因素。根据信息曲线,与 W_3 关系第二密切的变量是“图形大小”。它和“图形形状”加在一起,对 W_3 的信息覆盖度达到 100%,所以可以将 W_3 视为基于“图形形状”和“图形大小”两个因素的划分。

3 隐结构法应用于中医证候研究的步骤

下面通过抑郁症的研究实例,介绍将隐结构法

应用于中医证候研究的基本思想是：通过对基于临床调查的抑郁症的症状数据进行多维聚类分析，建立隐结构模型，结合中医专业知识对模型的内容进行分析，希望发现其中的一些隐类具有中医辨证意义，从而说明可根据辨证对抑郁症患者进行分类；为提取抑郁症的常见中医证候（和/或证候要素）及建立抑郁症的中医辨证规范（标准）提供较为客观的、量化的依据。其具体步骤包括以下 5 个方面：① 进行较大样本的临床流行病学横断面调查，收集抑郁症患者的中医症状数据；② 数据库的建立及对收集的数据进行预处理；③ 对数据进行隐结构分析；④ 对获得的隐结构模型进行诠释；⑤ 探索研究结果的意义和用途。

3.1 临床流行病学调查

从病证结合研究入手，在系统回顾近 10 年有关抑郁症证候研究的相关文献的基础上，结合临床经验，借鉴现代心理学量表的研究方法，编制了抑郁症中医四诊信息采集表，通过全国多中心的临床流行病学横断面调查，对符合抑郁症纳入标准的 604 例患者的症状信息进行了采集。

3.2 数据库的建立及数据预处理

应用 Epidata 3.0 软件建立数据库，由 2 人分别独立进行数据录入。由于隐结构分析复杂度高，受计算机运算能力的限制，分析数据时只选择了出现频率较高的 100 个显变量；又因为样本量相对较小，将调查症状由多值变量转为二值变量（0 或 1）。

3.3 数据的隐结构分析

预处理之后的数据先用 EAST 算法分析，得到一个含 34 个隐变量的隐树模型，其 BIC 分是 -268.07。接着用 LTM-EA 算法进行加边，BIC 分增加至 -267.47。最后形成的隐结构模型结构图如图 6 所示。以下称这里的数据为抑郁症数据，称所得到的模型为抑郁症的隐结构统计模型，简称抑郁症模型。

3.4 对获得的隐结构模型进行诠释

抑郁症模型中有 34 个隐变量，每个隐变量代表一个对数据样本的划分，每个划分由二个或以上的隐类组成，模型诠释就是从中医角度，把握这些划分以及各个隐类的含义。

3.5 探索研究结果的意义和用途

在结合中医理论对抑郁症隐结构中 34 个隐变量及相关隐类进行诠释的基础上，进一步探讨这些内容和结果在中医证候研究中的意义和用途。

4 抑郁症模型诠释示例

以抑郁症模型中的隐变量 Y_{33} 为例，示范隐变

量和隐类诠释的方法和步骤。

4.1 考察隐变量的信息曲线

在抑郁症模型的结构图中（图 5）， Y_{33} 处于右下角位置。它有两个取值，所以代表一个将抑郁症患者分为两类的划分。把 Y_{33} 的两个状态分别记为 s_0 和 s_1 ，而把 Y_{33} 划分中的两个隐类分别记为 $Y_{33} = s_0$ 与 $Y_{33} = s_1$ 。 Y_{33} 的信息曲线如图 6 所示，在图中可见，胸胁胀满情况（ X_1 ）、腹痛情况（ X_2 ）、胃脘胀满情况（ X_3 ）、胸胁胀痛情况（ X_4 ）和腹胀情况（ X_5 ）是与 Y_{33} 关系密切的显变量，它们对 Y_{33} 的信息覆盖度达到 98%。这表示，胸胁胀满情况等 5 个显变量反映出了隐类 $Y_{33} = s_0$ 与隐类 $Y_{33} = s_1$ 之间的几乎全部区别。因此，可以说 Y_{33} 代表的是一个基于胸胁胀满情况等 5 个显变量对抑郁症患者所做的划分。

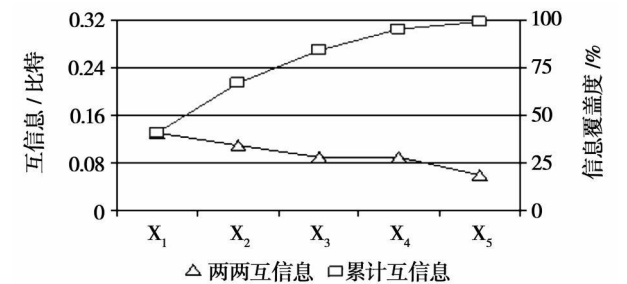
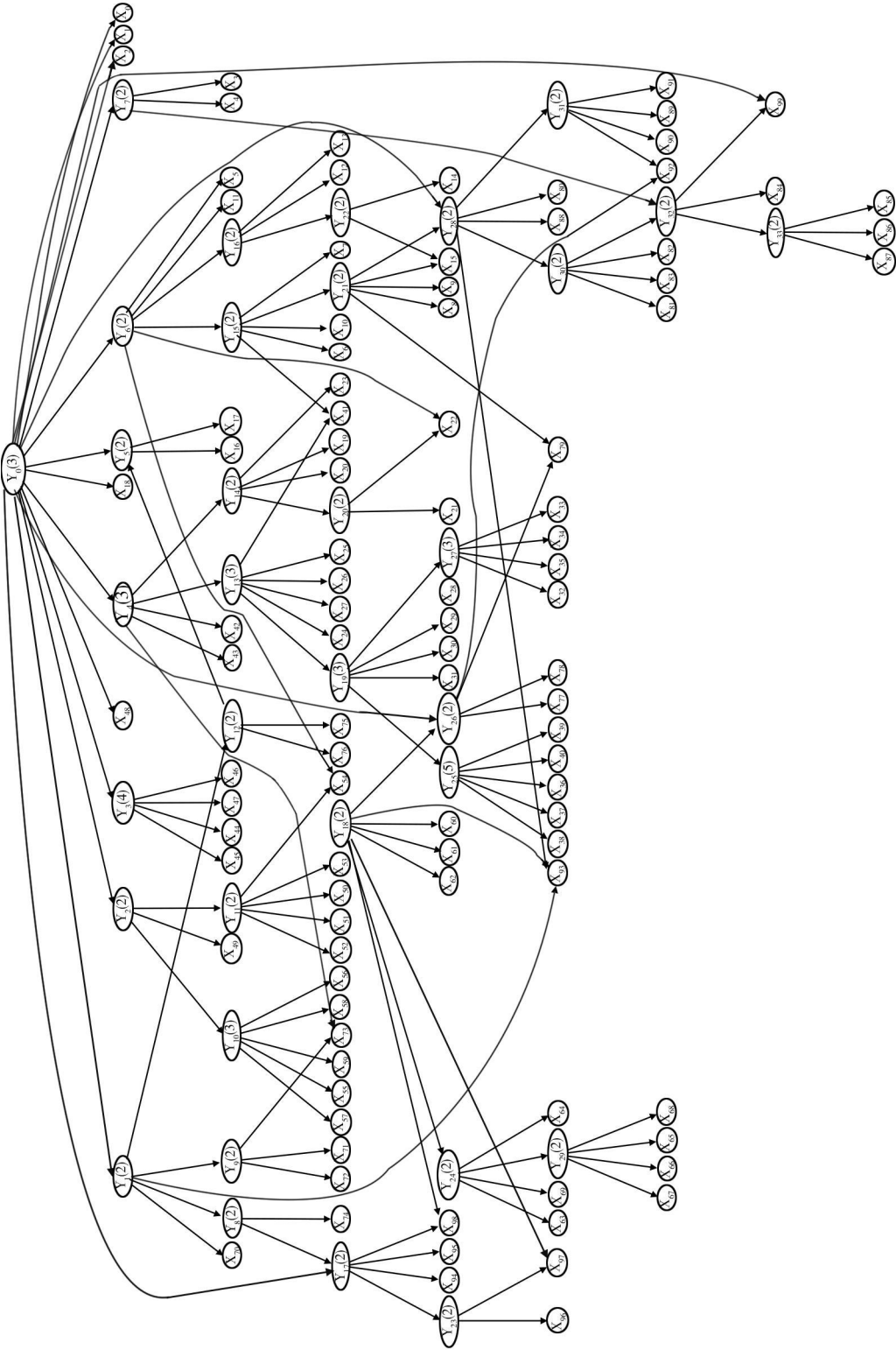


图 6 抑郁症模型中隐变量 Y_{33} 的互信息曲线

4.2 考察隐类的类条件概率分布

图 3 给出的 3 个划分都是硬划分，在硬划分中，一个事物要么属于一个类，要么不属于该类，中间无第三种可能性。与此不同，隐结构模型中隐变量所代表的划分是概率软划分，一个事物可能有一定概率属于某个类，同时也可以有一定概率属于另外的类。要把握概率软划分中各个类的特点，就需要考察它们的类条件概率分布。

Y_{33} 的两个隐类 $Y_{33} = s_0$ 和 $Y_{33} = s_1$ 的类条件概率分布见表 2。“ $P(X=1 | Y_{33} = s_0)$ ”是在隐类 $Y_{33} = s_0$ 中，某显变量 X 取值为“1”（即症状出现）的概率；而“ $P(X=1 | Y_{33} = s_1)$ ”是在隐类 $Y_{33} = s_1$ 中，显变量 X 取值为“1”的概率，表中显变量的选择基于 Y_{33} 的互信息曲线图。通过分析，发现两个隐类的主要区别在于：所有 5 个症状在 $Y_{33} = s_1$ 中出现的概率均较高，而在 $Y_{33} = s_0$ 中，胸胁胀满（ X_1 ）、腹痛（ X_2 ）、胸胁胀痛（ X_4 ）这 3 个症状出现的概率几乎为 0；胃脘胀满（ X_3 ）、腹胀（ X_5 ）这 2 个症状有一定出现概率。因此，隐变量 Y_{33} 是基于胸胁胀满情况等 5 个显变量把抑郁症患者划分为两大类。



X 显变量 Y 隐变量
图 5 抑郁症的隐结构模型结构图

表 2 隐类 $Y_{33}=s_0$ 和 $Y_{33}=s_1$ 的隐类条件类概率分布		
X	$P(X=1 Y_{33}=s_0)$	$P(X=1 Y_{33}=s_1)$
X_1	0.06	0.57
X_2	0.02	0.41
X_3	0.29	0.80
X_4	0.01	0.32
X_5	0.23	0.66

4.3 结合中医理论进行诠释

根据中医的辨证学理论一般认为：①胃脘胀满提示病变部位在胃，属胃脘气滞的表现。导致胃脘气滞的原因有多种，如痰、饮、水、湿、宿食、寒、热等邪气停滞于胃，或胃的阳气不足，均可导致胃中气机的郁滞不畅而出现胀满。②腹胀提示病变部位多在脾，也可能在大肠、肝等，是各种原因，如阳气不足，或各种邪气内阻影响腹部气机运行不畅所致。③胸胁胀满和胸胁胀痛提示病变部位多在肝、胆，常见于气滞、湿热、热邪等邪气阻碍肝、胆气机疏泄所致。④腹痛提示病变部位多在脾，也可能在大肠、肝等，常因各种虚，如阳虚、气虚等和实如气滞、寒凝、瘀血等因素引起局部气血阻滞所致。

每个隐变量都有两个密不可分的方面，一方面它代表一个划分，另一方面它代表划分产生的机制。作为产生划分的一个机制， Y_{33} 与上述 5 个症状变量密切相关。根据中医知识，初步认为导致这 5 个症状出现的机制最有可能是脾胃气滞、肝郁气结（或肝、脾胃不和）。由于脾胃气滞、肝郁气结的症状特征，除上述 5 个症状以外，还有许多其他方面。所以可以认为， Y_{33} 这个划分机制是脾胃气滞、肝气郁结的一个侧面。

综合前面讨论的内容，可以认为， Y_{33} 是基于胸胁胀满情况等 5 个显变量把病例样本划分为两个类，这个划分是脾胃气滞、肝郁气结一个侧面在症状上的反映。划分中的 $Y_{33}=s_0$ 这个类表现出脾胃气滞、肝郁气结表现的概率较高，而另一个类 $Y_{33}=s_1$ 则很少表现出肝气郁结的症状，但有一定的概率表现出脾胃气滞的症状。从 Y_{33} 所反映的病机层面看，病位在脾、胃与肝（可以理解为构成证候的病位类证候要素），直接相关的病性为气滞（可以理解为构成证候的病性类证候要素）。

5 讨论

一个隐变量代表一个划分，划分由两个或以上隐类组成。对一个隐变量进行诠释，首先基于信息曲线，考察哪些显变量与该隐变量关系密切以及它们信息覆盖度的高低，从而明确划分中各个类之间

的区别主要在哪些方面；其次考察各个隐类的条件概率分布，把握各个隐类的特点；最后结合中医专业知识，确定隐变量和隐类的中医学含义。在这个过程中，每一步都需要主观判断，存在一定主观性。但是，隐结构模型分析的结果是在客观临床调查数据基础上，由计算机在统计学原则指导下自动搜索而产生的，不是某一个人或几个人的经验，反映的是数据共性特征，因此具有较强的客观性。虽然对这些结果的诠释难以避免的会受主观因素影响，但它们对中医的研究还是很有意义的，主要有以下几方面。

第一，它能显示中医证候 病机的客观性。隐结构法的基本思想是，在统计学原则的指导下，寻找多个角度，对未经医生辨证的病例数据进行多维聚类分析，或者说多维划分。所得到的每一个划分都是数据在某一方面特征的反映，具有客观性。如果划分所反映的特征是某个证候 病机的特征，那么就说明该证候 病机具有客观性。例如，在抑郁症模型中隐变量 Y_{33} 代表的是一个主要依据胸胁胀满等 5 个症状对样本数据的划分。根据中医辨证学理论，这种划分所的特征可以认为是脾胃气滞、肝郁气结证（或肝、脾胃不和）的一个侧面。由于该划分的依据，来源于对一定样本的抑郁症的客观症状数据的分析，因此，体现了脾胃气滞和肝郁气结的客观存在性。需要指出的是，并不是所有隐变量所反映出来的数据特征都有中医意义。

第二，通过分析模型中具有中医辨证意义的隐变量或隐类可能涉及的证候及证候要素，为提取该疾病的常见证候及证候要素提供依据。

第三，通过运用模型中的数学信息分析一些隐变量与其所含显变量（症状）的关系，可以了解症状对于相应证候或证候要素的贡献度，从而为建立证候或证候要素的诊断标准提供依据。

总之将隐结构法用于中医证候的研究，是由张氏等^[1-3]于 2006 年首先提出的。本文系统阐述了关于隐结构法的目的、原理、步骤、结果诠释及意义方面的最新认识和进展。包括以下几个方面：①在以往主要强调隐结构概念的基础上，进一步明确了隐结构和多维划分在隐结构法中是两个不可分割的支柱。②更加明确了隐结构法的目的，即是基于数据特征，从多个角度对患者人群进行划分，从而为证候的规范化、标准化研究提供较为客观的依据。③在隐树分析算法方面有重要进展，而且研发了对树状模型进行加边的算法，从而摆脱了树状模型的限制。④从划分角度出发，深入探讨了隐结构模型

的诠释,使得现在可以准确地把握隐变量和隐类的含义。⑤本文通过一项具体研究实例,为读者使用隐结构法提供了一个范例。

参考文献:

- [1] 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及隐结构分析工具[J]. 北京中医药大学学报, 2006, 29(6): 365—369.
- [2] 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.
- [3] 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.
- [4] ZHANG N L, YUAN S H, Chen T, et al. Latent tree models and diagnosis in traditional Chinese medicine[J]. Artificial Intelligence in Medicine, 2008, 42: 229—245.
- [5] ZHANG N L, YUAN S H, Chen T, et al. Statistical Validation of TCM Theories[J]. Journal of Alternative and

Complementary Medicine, 2008, 14(5): 583—587.

- [6] 张连文,郭海鹏. 贝叶斯网引论[M]. 北京:科学出版社, 2007: 34—36.
- [7] ZHANG N L. Hierarchical latent class models for cluster analysis[J]. Journal of Machine Learning Research, 2004, 5(6): 697—723.
- [8] SULLIVAN P F, SMITH W, BUCHWALD D. Latent class analysis of symptoms associated with chronic fatigue syndrome and fibromyalgia[J]. Psychological Medicine, 2002, 32: 881—888.
- [9] WASMUS A, KINDEL P, MATTUSSEK S, et al. Activity and severity of rheumatoid arthritis in Hannover/FRG and in one regional referral center[J]. Scandinavian Journal of Rheumatology, 1989, 79(Suppl): 33—44.
- [10] SCHWARZ G. Estimating the dimension of a model[J]. Annals of Statistics, 1978, 6(2): 461—464.
- [11] 沈连丰,叶芝慧. 信息论与编码[M]. 北京:科学出版社, 2004: 22—24.

(收稿日期: 2009-02-20)

《北京中医药大学学报》第 33 卷(2010 年)征订启事

《北京中医药大学学报》系中央级的高级中医药学术刊物,1959 年创刊,由教育部主管、北京中医药大学主办。现分别属于中国自然科学核心期刊、中国中文核心期刊、全国中医药优秀期刊、全国优秀科技期刊、中国科学引文数据库首批入选的 315 种期刊之一、北京高校自然科学学报系统及全国高等医药院校学报系统优秀学报等。我刊 1998 年被俄罗斯《文摘杂志》(AJ)正式列为文摘刊源,1999 年被美国《化学文摘》(CA)正式列为文摘刊源,2007 年又被美国《国际药学文摘》(IPA)、美国《剑桥科学文摘》(CSA)、美国《乌利希期刊指南》(Ulrich PD)正式列为来源期刊。进入国际著名检索系统,标志我刊已在国际期刊界享有一定的声誉。2001 年我刊以“双效期刊”进入了中国期刊方阵。2006 年和 2008 年我刊两次被评为中国高校精品科技期刊。

本刊宗旨是为中医药教学、科研、医疗服务,主要反映中医药研究的新成果、新成就、新动态,贯彻“双百方针”,活跃学术空气,促进中医药学术发展。本刊最适合高级中医药工作者阅读参考。本刊以提高为主,兼顾普及。辟有“学科展望”“专家述评”“中医发展战略研讨”“博士之光”“科研思路与方法”“科技之窗”“理论研究”“临床研究”“文献研究”“中医药实验研究”“中药化学”“制剂与炮制”等专栏。

本刊为月刊,每月 30 日出版,国内外公开发行。A4 开本,72 页,每期定价 6.00 元,激光照排胶印。由北京报刊发行局向全国发行,国内读者请注意及时在当地邮局(所)订阅,国内代号 82-414;国外发行由中国国际图书贸易总公司(北京 399 信箱)办理,国外代号 M 734。

《北京中医药大学学报》始终是广大读者获取学术信息、开展学术交流的百花园,欢迎您订阅本刊。