隐结构模型对 2 型糖尿病病机规律的探讨*

□龚燕冰** （中国中医科学院广安门医院内分泌科　北京　100053）
张连文 （香港科技大学计算机科学及工程学系　香港　）
高思华 （北京中医药大学　北京　100029）
罗增刚 （中国中医科学院　北京　100700）
倪　青** （中国中医科学院广安门医院内分泌科　北京　100053）
谢雁鸣 （中国中医科学院中医临床基础医学研究所　北京　100700）

摘　要:目的:探索 2 型糖尿病的病机规律。方法:以 2501 例 2 型糖尿病的临床数据为基础,运用隐结构模型的方法加以研究。结果:得出隐结构模型关于 2 型糖尿病“症-证—病机”相互关系及病机规律。结论:隐结构模型是一种非线性的方法,这对于研究非线性、复杂性的中医证候无疑是一种比较合适的手段。

关键词:2 型糖尿病　病机规律　隐结构模型　决定性水平曲线　累积互信息曲线　两两互信息曲线　隐类

一、研究对象与方法

1. 研究对象

病例选取自 2003 年 6 月~2005 年 2 月中国中医科学院广安门医院内分泌科、北京中医药大学东直门医院、东方医院、山东中医药大学附属医院、济南市中医医院、潍坊市中医院、保定市中医院、河南中医学院附属医院、新疆中医院等医院的 2 型糖尿病患者,共 2501 例。

对上述原始病例资料进行双人双机 epidata 录入,建立 excel 表格。数据库包括病人的入院信息、体格检查、理化检查、症状、证型、中西药运用、出院诊断七个模块。数据库中不含有文字,每个变量唯一编码,并通过填充空缺值、纠正非法值、数据变换等进行严格的质量控制。

2. 方　法

隐结构法是一种研究中医证候分布规律的新方法。它由显变量、隐变量及隐结构图等组成。诸如“腰膝酸痛”、“舌红少苔”、“脉细数” 等症状可以通过望闻问切而直接观察到,是显变量,“肾阴虚”的证候迄

收稿日期: 2008-06-25
修回日期: 2009-02-16
* 国家自然基金面上资助课题(90209012):糖尿病中医证候的量化标准的研究,负责人:高思华;香港研究资助局项目(622408):有监督学习与无监督学习中隐结构挖掘的一个共同框架,负责人:张连文;香港研究资助局项目(622307):加快多层隐类模型的学习,负责人:张连文;科技部基础性工作专项基金资助项目(2001DEA20010):中药材标准及中医临床疗效评价标准,负责人:王永炎。
** 联系人:龚燕冰,博士后,主要研究方向:中医内科内分泌代谢专业,E-mail:norah121212@yahoo.com.cn;倪青,主任医师,硕士研究生导师,主要研究方向:中医内科学内分泌代谢病的研究,Tel:010-88001023,E-mail:niqing669@163.com。

今无法直接观察,是隐变量。除了证候以外,诸如“肾阴虚失养”、“肾阴虚失滋润”、“阴虚内热”的病机特点也是隐变量。隐变量与隐变量之间以及隐变量与显变量之间的关系构成了一个隐结构[1]。

基本步骤如下:(1)系统收集临床数据;(2)利用隐结构模型对数据进行分析;(3)对所得到的模型进行诠释;(4)将结果用于指导辨证论治。主要讨论第3步。

二、结 果

将数据库中涉及到的症状、舌、脉等86个变量纳入隐结构模型的分析中,计算得到最优隐结构模型如图1所示。

隐结构模型是通过计算机运算之后得出的直观结构图形,其真正的内涵可以用三线图和隐类谱来表示。三线图和隐类谱中所描述的信息,有的不会完全在结构图中显示出来,所以要正确、客观地诠释隐结构模型,首先要了解三线图的原理和隐类谱中包含的内容。

1. 通过三线图对隐变量的诠释

隐变量之所以有意义,是因为它们与显变量有关系,每一个隐变量都会与很多显变量有关系,其密切程度在三线图中能观察到。三线图中会涉及到两两互信息曲线、累积互信息曲线以及决定性水平曲线。

(1)两两互信息曲线。

互信息,顾名思义是两个离散变量X和Y相互关系紧密程度的度量,当X和Y相互独立时,$I(X;Y)$取值是0,为最小;当X和Y相互完全决定对方时,即其关系最紧密,$I(X;Y)$取值最大。假如X是隐结构模型中的隐变量,那么与X的关系最紧密的显变量Y_1的互信息就最大,其次为Y_2、Y_3……,Y_n把X与Y_1、Y_2、Y_3……,Y_n的互信息描述在一个坐标里,用一条线将它们连接起来,就得到了关于隐变量X的两两互信息曲线(pairwise mutual information curve),两两互信息曲线是单调下降的。

(2)累积互信息曲线。

隐变量X的含义首先应该考虑Y_1,其次是Y_2,依次类指,那么,要到哪里停止呢?这就要考虑隐变量X与一组显变量之间的互信息,X与Y_1、Y_2两个显变量的互信息,不会超过X分别与Y_1、Y_2的互信息之和。例如两个圆盘投影到墙上,影子的总面积不会超过两个影子的各自面积之和,如果有重叠,就会小于后者。把X与前I个显变量Y_1、Y_2、Y_3……,Y_i的互信息,记为$I(X;Y_1-Y_i)$,并用一条曲线连接起来,就得到隐变量X的累积互信息曲线(cumulative mutual information curve),为一条单调上升的曲线。

$I(X;Y_1-Y_i)$是前i个显变量所揭示的信息,而$I(X;Y_1-Y_n)$是X所含关于所有隐变量的信息。因此,$I(X;Y_1-Y_i)/I(X;Y_1-Y_n)$是前i个显变量所揭示的部分占总信息的比例,称为前i个显变量的信息覆盖度(degree coverage)。随着显变量的增加,信息覆盖度逐渐增高。信息覆盖阈值α是一个接近1的数字,如$\alpha=0.95$,如果在考察前i(α)个显变量之后,信息覆盖度已达到阈值α,说明已经掌握了关于X的主要信息。

(3)决定性水平曲线。

在考察前i(α)个显变量之后,子集覆盖度已经接近1,此时考虑更多的显变量,信息不会显著增加,但并不意味着通过前i(α)个显变量已经可以很好地把握隐变量X的意义。利用决定性水平曲线可以判断通过前i个显变量是否足以把握隐变量X的涵义。分别计算前I个显变量对隐变量X的决定性水平,将其用一条曲线连接起来,就得到关于隐变量X的决定性水平曲线,为单调上升的曲线。

本文仅以两个隐变量的三线图为例加以说明,见图2、图3,其他省略。

分析所得的2型糖尿病的隐结构模型,有些隐变量有明确的含义,另外一些含义不是很清楚,不可诠释含义明确的隐变量有:X3、X5、X6、X9、X13、X17、X18、X19、X22、X24、X30、X31、X32、X33、X34、X36;不可诠释的隐变量有:X0、X1、X2、X4、X7、X8、X10、X11、X12、X14、X15、X16、X20、X21、X23、X25、X26、X27、X28、X29、X32、X35。分析其信息曲线,发现这些隐变量在不同意义的板块之间有桥梁作用。

一个隐变量的含义往往是由它周边的显变量决定的,但也有很多例外,例如:消谷善饥、多食易饥与X21相连,但在信息曲线中与X21的关系不大;相反,面色晦暗与X36不相邻,却对X36有显著影响。另外,与桥梁变量密切相关的显变量更是来自模型的各个部分。

含义明确的隐变量之间也会有信息重叠,例如:X31所覆盖的显变量是X30所覆盖的显变量的子集,前者便可舍去。

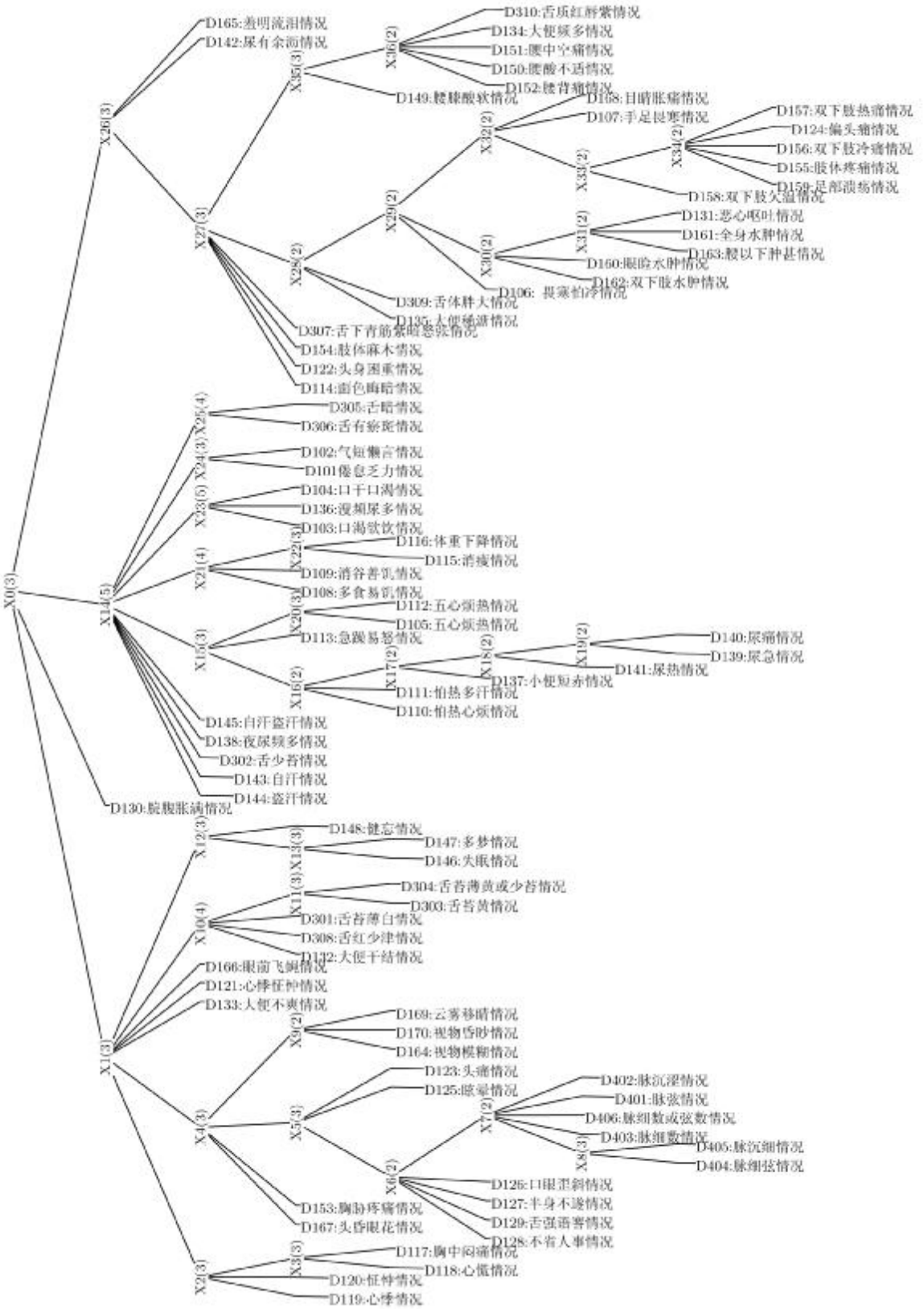

图 1 2 型糖尿病隐结构模型

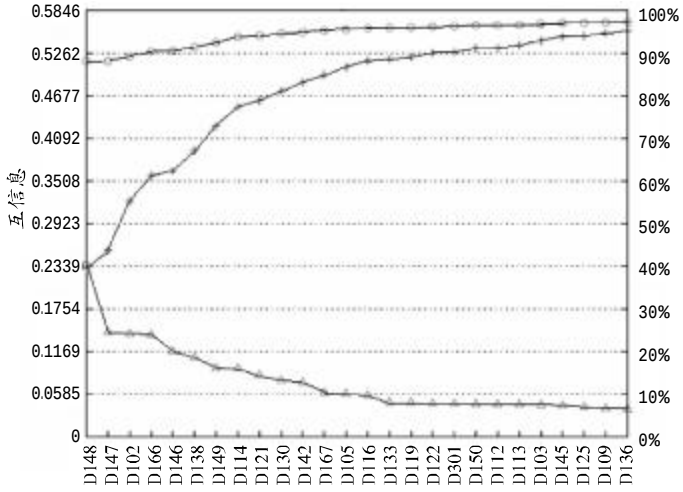

图 2　隐变量 X0 的两两互信息曲线(Δ),累积互信息曲线(+),以及决定度曲线(0)

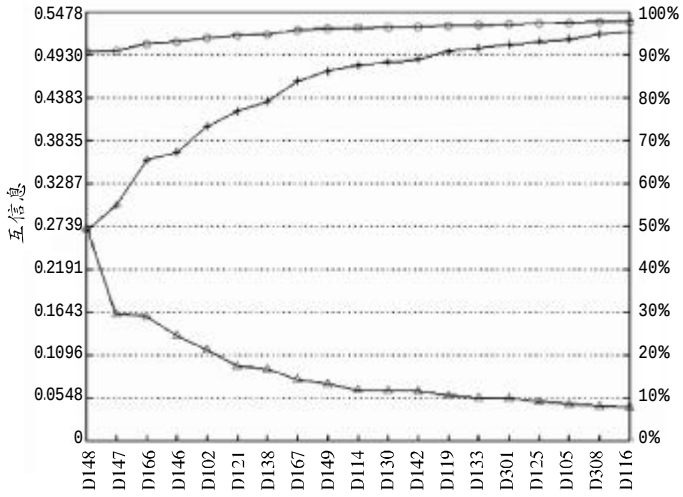

图 3　变量 X1 的两两互信息曲线(Δ),累积互信息曲线(+),以及决定度曲线(0)

含义明确隐变量的诠释:

X3 病位在心,可以理解为心血瘀阻。心脉闭阻,致心失所养,心脉不畅,心神失养引起引起心慌、胸中闷痛、心悸;

X5 病位在肝肾,可理解为肝肾阴虚,脑髓失充,头目失养,故眩晕、健忘、头晕眼花;肾阴虚,水不涵木,肝阳偏亢,气血上冲,则见头痛;

X6 病位在脑,可以理解为风痰阻络,风痰流窜经络,血脉闭阻,气血不通故见半身不遂,舌强语謇,口眼歪斜;

X9 病位在肝,可以理解为肝开窍于目,肝阴血亏虚,不能上荣目窍,可见视物模糊、视物昏缈;头目失养故健忘、头晕眼花、眩晕;

X13 病位在心,可理解为肾阴亏虚,水不济火,心火偏亢,心神不宁,故可见夜梦多,失眠;肾阴虚脑髓失充,可见健忘;

X18 病位在膀胱,可理解为心火亢盛,火下结于下焦,影响膀胱的气化,气郁化火,煎熬尿液,故小便短赤、尿热;心火郁于内故怕热心烦;

X19 病位在膀胱,可理解为湿热蕴结下焦,膀胱气化失司,尿液受其煎熬,故可见尿热,热盛伤络故见尿痛,尿急;

X22 病位在脾,可理解为脾阴亏虚,则消谷而不能充肌肉,津液无所生而燥热内炽,上灼肺液,下耗肾阴,日久则体重下降,消瘦;

X24 病位在肾,可理解为肾阳亏虚,元气不足,机体功能减退,可见气短懒言,倦怠乏力;肾阳虚失于摄纳,可见夜尿频多;阳不化气,阳损及阴,阴精不足,脑髓不充,故见健忘;

X30 病位在脾肾,可理解为脾肾阳虚,阳不化气,水湿下聚,故见水肿以下肢为甚;腰膝以下,肾气主之,故腰以下为甚;眼睑为肉轮,属脾,脾失健运,水湿停于眼睑,则水肿;脾肾阳虚,不能温煦四末,故手足畏寒,畏寒怕冷;

X31 属于 X30 的子集,即为脾肾阳虚水泛,出现腰以下肿甚,双下肢水肿,全身水肿,眼睑水肿;

X32 病位在脾肾,可理解为脾肾阳虚,出现手足畏寒,畏寒怕冷,双下肢欠温,双下肢水肿;脾阳亏虚,运化无力,可见大便稀溏;X32 也可以理解为 X31 和 X33 之间的桥隐变量;

X33 病位在脾肾,可理解为脾肾阳虚,失于温煦,可见双下肢欠温,手足畏寒,畏寒怕冷;阳虚血运无力,血流瘀阻于下肢,则见双下肢冷痛;

X34 病位在肾,可理解为肾阳虚衰,不能温养四末,可见双下肢欠温,手足畏寒;阳虚血瘀,不通则痛,故见双下肢冷痛,肢体疼痛;足部血瘀,经络虚涩,不能濡养肌肉,易发足部溃疡;

X36 病位在肾,可理解为腰为肾之府,肾气亏虚,腰府失养,可见腰酸不适,腰膝酸软,腰中空痛;肾气虚,不能上荣于面,肾主色黑,故见面色晦暗。

桥梁变量是连接不同隐变量之间的桥梁, 如果通过同一个桥梁变量连接的隐变量之间的关系可能比不同桥梁变量连接的隐变量之间的关系紧密这种假设成立的话, 便可得出更多的关于应证组合的规律,今后的工作中可进一步探讨。

2. 通过隐类对隐变量的诠释

可以解释的隐变量由隐类构成，隐类把隐变量分为不同的等级，可以加深对隐变量的理解。一个隐变量有两个或以上的可能取值，每个取值代表一个隐类，本文只选择有明确含义的隐类进行诠释。假设隐变量 X 有两个可能取值 S_0 和 S_1，属于该隐变量的前 i 位显变量，在 S_0 和 S_1 中的分布会有不同，于是 S_0 和 S_1 就有不同的含义。

S_0 和 S_1 的诠释结果有两种可能：第一种，S_0 和 S_1 所包含的症状发生同步变化，在一类里症状出现的频率很低甚至几乎为零，可诠释为“无”，在另一类里则可诠释为“有”；第二种，症状在每类中出现的频率都不低，只是在程度上有差异，可分别诠释为轻度、中度或重度。

用隐类谱来诠释隐类的含义，相当于典型成员的方法，即通过将每一个隐变量单独拿出来，对其下显变量的分布情况逐个分析。举例说明：

“有”、“无”的例子：以 X33 为例，X33 包含双下肢欠温，手足畏寒，畏寒怕冷，双下肢冷痛。如前文所述，X33 病位在脾肾，可理解为脾肾阳虚，失于温煦，见双下肢欠温，手足畏寒，畏寒怕冷。阳虚血运无力，血流瘀阻于下肢，则见双下肢冷痛；

X33 有两类，分别为 $X33=S_0$、$X33=S_1$，其显变量分布分别见表 1、表 2：

在 $X33=S_0$ 中，四个症状中至少有三个均为零；在 $X33=S_1$ 中，均有双下肢欠温症状且为中度，其他症状可出现 1 到 3 个。前面把 X33 诠释为肾阳虚程度，所以 $X33=S_0$ 可理解为没有肾阳虚，$X33=S_1$ 可理解为有肾阳虚。

轻、中、重程度的例子：X3 中包含心慌、胸中闷痛、心悸，其 X3 病位在心，可以理解为心血瘀阻，心脉闭阻，致心失所养，心脉不畅、心神失养，引起心慌、胸中闷痛、心悸。

X_3 分为 3 类，分别为 $X3=S_0$、$X3=S_1$、$X3=S_2$，其显变量分布分别见表 3、表 4 和表 5：

可以看出，$X3=S_0$ 3 个症状中，最多有一个症状出现；$X3=S_1$ 3 个症状中，至少有两个症状出现，且程度基本在中度以下；$X3=S_2$ 中，至少有一个症状为重度，或有两个中度以上的症状。前面已经把 X_3 诠释为心血瘀阻的程度，所以可以把 $X3=S_0$ 的含义理解为没有或轻度心血瘀阻症状，把 $X3=S_1$ 理解为中度心血瘀阻症状，把 $X3=S_2$ 理解为重度心血瘀阻症状。

表 1 $X33=S_0$ 的情况

D158	D107	D106	D156	例数	最大可能	最小可能
0	0	0	0	1982	0.973997	0.575876
0	0	1	0	24	0.973935	0.537640
0	1	0	0	13	0.973935	0.601645
0	0	2	0	18	0.929392	0.534997
0	0	0	1	39	0.878780	0.540072
0	0	0	2	22	0.811203	0.517354

表 2 $X33=S_1$ 的情况

D158	D107	D106	D156	例数	最大可能	最小可能
2	0	0	0	72	1	1
2	0	0	1	1	1	1
2	0	0	2	18	1	1
2	0	1	0	1	1	1
2	0	2	0	9	1	1
2	0	2	2	1	1	1
2	1	0	0	4	1	1
2	1	0	2	1	1	1
2	1	1	0	3	1	1
2	1	1	1	1	1	1
2	1	1	2	2	1	1
2	1	2	2	1	1	1
2	2	0	0	18	1	1
2	2	0	2	3	1	1
2	2	0	3	1	1	1
2	2	1	0	1	1	1
2	2	2	0	8	1	1
2	2	2	1	1	1	1
2	2	2	2	3	1	1
2	3	0	0	1	1	1

表 3 $X3=S_0$ 的情况

D118	D117	D119	例数	最大可能	最小可能
0	0	0	1713	0.980661	0.871105
0	1	0	168	0.97764	0.871188
0	0	1	57	0.962226	0.648171
0	1	1	16	0.95645	0.946416
0	2	0	60	0.74853	0.64442
0	0	2	45	0.637869	0.6311
0	0	3	4	0.637738	0.608051
0	1	2	3	0.602644	0.595969
0	2	1	7	0.599234	0.586161

表 4　X3＝S_1 的情况

D118	D117	D119	例数	最大可能	最小可能
1	1	1	28	0.997735	0.989637
1	1	0	63	0.991067	0.960924
1	0	1	29	0.98699	0.939128
1	2	1	11	0.965434	0.933798
1	2	0	14	0.923608	0.842329
2	1	2	1	0.892073	0.892073
2	1	1	6	0.892063	0.822961
1	3	1	1	0.800093	0.800093
2	1	0	7	0.729153	0.659351
2	0	1	6	0.594979	0.565104

表 5　X3＝S_2 的情况

D118	D117	D119	例数	最大可能	最小可能
0	3	0	8	1	1
0	3	2	1	1	1
3	0	0	2	1	1
3	0	3	1	1	1
3	3	0	1	1	1
3	3	3	1	1	1
2	0	2	14	0.999928	0.95277
2	2	2	5	0.999567	0.66306
2	1	2	5	0.998455	0.748778
2	0	0	40	0.997591	0.590255
2	3	2	1	0.997528	0.997528
1	2	2	2	0.98704	0.714321
2	3	0	1	0.982046	0.982046

这样，把每一个隐变量都分为不同的隐类，便可以对隐变量进行判定：包含哪些症状，属于什么程度，可以判断其类别。这为证候要素或者其判定标准的客观化提供了客观的手段。

三、讨　论

中医学中，症状和证候之间、证候与证候之间的关系以及疾病的病机规律都多为复杂的非线性关系，它们之间存在大量的多重共线性关系和协同关系。隐结构法是一种运用计算机手段实现中医证候研究的非线性方法，能使辨证客观性得到大大提高。本文所得隐结构模型，是从两个层次对结构图中的隐变量进行诠释的：其一是三线表从定性层次；其二是隐类表从半定量层次，二者分别从纵向和横向展现了 2 型糖尿病的病机规律。研究结果提示，隐结构法是一种比较合适研究中医证候及疾病规律的方法。

注：两两互信息曲线、累计互信息曲线、信息覆盖度、决定性水平、隐类等概念均引自：张连文，王天方、王焱、赵燕．隐结构模型诠释．技术报告 HUST-CS07-04，香港科技大学计算机科学及工程学系，2007 年 3 月．

参考文献

1　张连文，袁世宏．隐结构模型与中医辨证研究(I)–基本思想以及隐结构分析工具．北京中医药大学学报，2006，(29)6:365~369．

Study on pathogenesis of Type 2 Diabetes by the Latent Structural Model

Gong Yanbing[1], Zhang Lianwen[2], Gao Sihua[3], Luo Zenggang[2], Ni Qing[4], Xie Yanming[5]

(1. Guang'anmen Hospital, China Academy of Chinese Medical Sciences, Beijing 100700, China;
2. Department of Computer Science and Engineering, Hongkong University of Science and Technology, Hongkong, China;
3. Beijing University of Chinese Medicine, Beijing 100029, China;
4. China Academy of Chinese Medical Sciences, Beijing 100700, China;
5. Institute of Fundamental Clinical Medicine, China Academy of Chinese Medical Sciences Beijing 100700, China)

Abstract: Based on 2501 clinical cases of type 2 diabetes, the pathogenesis of type 2 diabetes was studied by the latent structural model, and, the symptoms - pathogenesis relationship was established. The results show that the latent structure model, which is nonlinear, may serve as a good tool for investigating the nonlinear and complex traditional Chinese medicine symptoms.

Keywords: type 2 diabetes; pathogenesis; latent structural model

（责任编辑：崔建华，责任译审：张立崴）