

· 方法学与临床评价 ·

基于隐结构的中医脾系问诊信息客观化分析

刘国萍^{1*} 邓 峰² 夏春明² 王忆勤¹ 许朝霞¹ 李福凤¹ 钱 鹏¹ 郝一鸣¹

(1. 上海中医药大学中医四诊信息综合研究实验室, 上海市浦东新区蔡伦路 1200 号, 201203;

2. 华东理工大学自动控制与电子工程系)

基金项目: 国家自然科学基金资助项目(30901897); 上海市卫生局青年科研项目(2008Y128);

上海高校选拔培养优秀青年教师科研专项基金(SYZ08003); 上海市教育委员会资助项目(S30302)

* 通讯作者: tianshanghuo3838@163.com, (021)51322286

[摘要] 目的 建立隐结构模型, 探讨中医脾系问诊信息客观化特征。方法 使用隐结构法的启发式双重爬山法进行自动搜索, 对已经进行数据转换和录入的 1310 例脾系样本, 选用 40 个参与分析的变量来进行数据分析, 观察分析过程中模型贝叶斯信息标准(BIC)评分的变化。结果 1310 例中医脾系疾病中以慢性浅表性胃炎及慢性萎缩性胃炎患者居多, 分别占 46.6% 和 14.7%。得到的初始模型记为 M_0 , 模型 M_0 中有 12 个隐变量, BIC 评分为 -20243.42806654384。结论 隐结构模型与中医理论及脾系疾病的证候特点比较吻合, 基本反映了脾系疾病的主要病因病机, 为中医问诊的客观化研究提供了有益的借鉴。

[关键词] 脾系; 问诊; 隐结构法; 四诊客观化

Analysis on the Objectification of Questioning Information on TCM Spleen System: Based on the Latent Structure

LIU Guoping¹, DENG Feng², XIA Chunming², WANG Yiqin¹, XU Zhaoxia¹, LI Fufeng¹, QIAN Peng¹, HAO Yiming¹

(1. Shanghai University of Chinese Medicine, Shanghai 201203; 2. Shanghai East China University of Science and Technology)

ABSTRACT **Objective** To explore the objectification of TCM spleen system questioning information based on a latent structure model. **Methods** Using a heuristic doubly hill-climbing method of latent structure to do automatic search, 40 variables participating in analysis were selected from 1310 spleen system samples transformed and recorded already for data analysis. The changes of Bayesian Information Criterion (BIC) score of models in the analysis process were observed. **Results** Of the TCM spleen system diseases in 1310 cases, most were chronic superficial gastritis and chronic atrophic gastritis patients, accounted for 46.6% and 14.7% respectively. The initial model obtained was marked as M_0 . In the model M_0 there were 12 latent variables and their BIC score was -20243.42806654384. **Conclusion** The latent structure model is in accordance with the syndrome characteristics of Chinese medicine theory and spleen system diseases, basically reflected the main etiology and pathogenesis of spleen system diseases, and provided the beneficial reference for the objectification studies on TCM questioning information.

Key Words spleen system; questioning; latent structure model; objectification of four diagnosis

中医辨证缺乏客观标准, 辨证技能很大程度取决于经验, 辨证结果主观因素影响很大, 不同医师对同一患者所作的诊断往往不同^[1]。因此, 建立中医辨证客观的、定量的标准十分必要。中医问诊信息占中医诊断信息的 70% 以上, 本研究旨在以中医脾系问诊为切入点, 应用隐结构法对中医脾系问诊信息进行客观分析。

1 资料

1.1 研究对象

收集 2009 年 3 月至 2010 年 5 月上海中医药大学附属龙华医院、上海中医药大学附属曙光医院、上

海交通大学医学院附属仁济医院、上海市中医医院门诊及住院患者, 共采集有效脾系病例样本 1310 例。其中男 521 例, 平均年龄 (50.57 ± 18.32) 岁, 女 789 例, 平均年龄 (48.88 ± 16.84) 岁。病程 1 个月 ~ 30 年, 平均病程 (59.30 ± 85.54) 个月。

1.2 诊断标准

中医脾胃病诊断标准参考《中医病证诊断疗效标准》^[2]、《中医临床诊疗术语·证候部分》^[3]、中医诊断学、内科学教材^[4-8]。西医常见消化系统疾病, 如慢性胃炎、胃溃疡、急、慢性结肠炎等疾病的诊断标准参考《内科学》^[9]。

1.3 辨证标准

参考中医诊断学教材、中医内科学教材^[4-8]及《中医临床诊疗术语·证候部分》^[3]制定中医脾系病证辨证标准,共 21 个证候,即寒邪客胃、食滞胃脘、胃火炽盛(上逆)、瘀血停胃、胃阴虚、脾气虚、脾阳虚、脾虚气陷(中气下陷)、脾不统血、寒湿困脾、湿热蕴脾、痰饮内阻、肝气犯胃、肝胃郁热、肝气乘脾、脾胃虚弱、脾胃虚寒、脾肾阳虚、脾肺气虚、心脾两虚、肝脾不调(肝胃不和)。

1.4 纳入及排除标准

纳入符合中医脾系病证诊断及西医常见消化系统疾病诊断者,知情同意者。排除精神病患者及伴有其他系统重度疾病者;语言表达能力较差,病情叙述有困难者。

2 方法

2.1 量表的制作

由上海市资深中西医结合临床消化系统专家、临床医生及研究者组成研究小组。检索时间为 2000—2010 年,以“关键词”或“题名”为主要检索方式,检索词为“胃痛”、“慢性胃炎”、“腹痛”等中医及西医病名,分别在中国知网(CNKI)、维普网(VIP)等数据库进行检索涉及中医脾胃病的相关文献,同时参考中医诊断学、中医内科学等^[2-8]相关教材,初步制定出临床流行病学调查表。并经两轮的专家咨询及相关的统计学检验,完善修改量表。中医脾系问诊量表共包括寒热、汗、头身胸腹、二便、饮食口味、睡眠、情绪、妇女 8 个维度,还有既往史、望诊、切诊等内容,共 100 余个变量。其中问诊一级症状变量 63 个。

2.2 调查方法

采样人员经统一培训。同时对量表中症状给予明确的定义,指出问诊时的具体操作方法和顺序。为了保证在调查过程中的统一,小组成员定期集中,对典型病例的资料进行讨论,以尽可能保证所采集资料的一致性。

2.3 数据输入及处理

采用 Epidata 3.0 软件建立数据库,将问诊症状进行 0,1 转换。独立双遍录入,对两份录入数据进行对比核查。逻辑检查,检查调查表填写错误。

2.4 模型建立方法

2.4.1 分析原理 多层隐类模型(hierarchical latent class model, HLC 模型),是一种特殊的隐变量模型^[10],用 HLC 模型来分析数据即是假设显变量

的背后是一个 HLC 模型,并且基于数据对这个模型进行推测。数据中只有显变量的取值,而无隐变量的取值,即没有对隐变量的直接观测。用 HLC 模型分析数据是对样本同时进行多维聚类。因此隐结构模型的分析是一种基于模型的聚类分析。

我们利用的是双重爬山(doubly hill climbing, DHC)算法,即把模型优化问题分解成势学习问题和预模型学习问题。DHC 算法不是直接在所有正则 HLC 模型的空间中搜索,而是利用势学习和预模型学习这两个问题把该空间分成了两个层次,在每个层次中分别进行搜索。

2.4.2 分析工具 HLCM 是香港科技大学张连文教授编写的一套 JAVA 语言的隐结构分析学习工具箱,这个工具箱包括 3 个部分: Hlcm.zip//学习包里的封装好的类; LearnHLCM.java//运行的主程序; sampleData.Txt//提供的参考数据格式。该软件的运行需要下载装 JAVA 运行环境 JDK 工具,下载安装完以后还要设置好环境变量,运行平台可以利用系统的 DOS 命令窗口。使用 HLCM 工具箱前,需要把我们的格式按照工具所需要的格式改写,在计算时,得到的模型数据被保存在单独的文件中。

2.4.3 数据筛选 由于启发式 DHC 法进行自动搜索,该算法无法处理所涉及的全部变量(问诊信息)。在做数据分析之前首先对数据进行筛选,即变量的降维处理。筛选变量的原则是:既要反映中医脾系疾病的临床发病特征,也要体现中医脾系病证的辨证(病性、病位辨证)理论,充分体现中医理论和中医专家的指导性作用。数据筛选主要通过频次分析和专家评分相结合进行,最终筛选出 40 个症状变量来进行数据分析。具体方法如下:①频次统计分析:对 1310 例脾系患者问诊信息,采用 SPSS18.0 进行处理,按照问诊症状出现的频次进行排序。②中医专家评分:制订专家咨询表,邀请 5 位上海市中医脾胃病资深专家,请专家对问诊症状对脾系病证辨证的重要性按“重要(3 分)、次要(2 分)、不重要(1 分)”3 个水平进行评分,将症状按照专家评分的总分进行排序。

2.5 分析过程

使用启发式 DHC 法进行自动搜索,观察分析过程中模型贝叶斯信息标准评分^[11](Bayesian Information Criterion, BIC 评分)的变化。1 个 HLC 模型的质量可以用 BIC 评分来度量。BIC 评分一方面要求模型与数据拟合、揭示数据中隐含的规律,另

一方面要求模型尽量简单。用 HLC 模型分析数据就是要寻找 BIC 分数最高的最优 HLC 模型^[12]。在搜索过程中,模型与数据的拟合程度不断改进,从而 BIC 评分逐步增加。当模型变得太复杂时,BIC 不升高反而降低,这时停止搜索。

3 结果

3.1 研究对象的疾病分布情况

表 1 示,1310 例中医脾系疾病中以慢性浅表性胃炎及慢性萎缩性胃炎患者居多,分别占 46.6% 和 14.7%,其余所占比例由多到少依次为溃疡性结肠炎,胃炎兼溃疡,胃溃疡,急、慢性结肠炎,十二指肠球部溃疡,其他消化系统疾病,胃出血,消化系统肿瘤。

表 1 疾病分布情况

疾病类型	频率(例)	百分比(%)
慢性浅表性胃炎	611	46.6
慢性萎缩性胃炎	193	14.7
溃疡性结肠炎	154	11.8
胃炎兼溃疡	98	7.5
胃溃疡	73	5.6
急、慢性结肠炎	72	5.5
十二指肠球部溃疡	33	2.5
其他消化系统疾病	30	2.2
胃出血	27	2.1
消化系统肿瘤	19	1.5
合计	1310	100

3.2 模型的构建结果

使用启发式双重爬山法进行自动搜索,对以上采集的 1310 例数据,选取 40 个变量参与分析。得到的初始模型记为 M_0 , BIC 评分为 -20243.42806654384。模型具体见图 1。

3.3 模型诠释

模型 M_0 中有 12 个隐变量,每个隐变量代表 1

个对数据样本的划分,每个划分由 2 个或以上的隐类组成,模型在定性方面的评价就是从中医角度,解释这些划分以及各个隐类的含义。每个隐变量都有两个密不可分的方面:一方面代表一个划分,另一个方面代表划分产生的机制。

Y_0 因子是总的隐类,说明胁肋疼痛、口干口渴、口苦是脾系疾病的共同发病基础之一; Y_1 因子影响泛酸、口黏腻、口臭、大便不爽、喜热饮变量的出现,是脾胃湿热的表现; Y_2 因子影响畏寒、肢冷两个症状的出现,是机体阳气虚衰的表现; Y_6 因子影响胀痛、痛无定处变量的出现,是气滞的证候要素特点; Y_8 因子影响刺痛、痛有定处变量的出现,是瘀血的表现; Y_{11} 因子影响身体困重、乏力懒言、小便频数变量的出现,可归纳为气虚湿阻的病机特点; Y_{12} 因子影响呃逆/暖气、胃脘胀满、嘈杂变量的出现,可归纳为气滞的证候要素特点; Y_4 因子影响潮热、盗汗、自汗症状的出现,是气阴两亏的表现; Y_{10} 因子影响恶心呕吐、食欲不振、口淡乏味、失眠、多梦、手足心热的出现,是心脾两虚的表现。以上的因子与中医理论基本吻合,同时反映了脾系病证的病因病机特点,说明脾系疾病多和肝、心等脏腑关系密切,且肝胃不和、肝脾不调出现肝郁化热是脾系疾病的重要病因病机。

此外, Y_3 因子、 Y_5 因子、 Y_7 因子很难用单一的证候要素解释,还有待进一步研究。

4 讨论

中医数据研究中,往往特别关注症状与症状间以及症状与证候间的关联及规律。即特别关心隐变量与隐变量之间以及隐变量与显变量之间的关系。中医问诊中的胃脘痛等症状是显变量,而对应的脾气虚等证候是隐变量。隐变量与隐变量之间以及隐

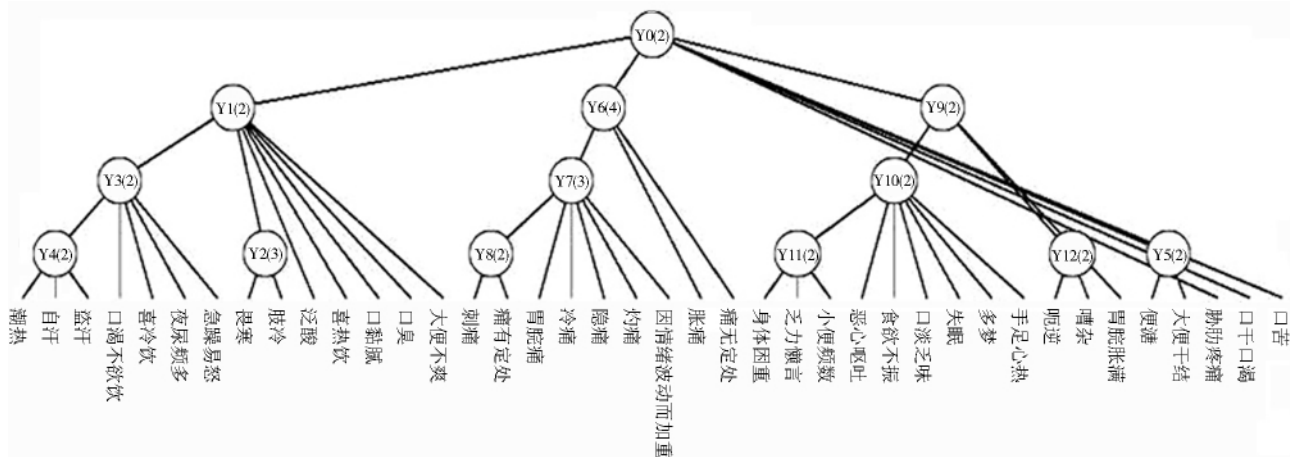


图 1 M_0 模型

变量与显变量之间的关系构成了一个隐结构。隐结构法^[10]是用电脑来取代人脑进行数据分析,通过临床流行病学调查,系统收集病例;然后用计算机对所得数据进行分析,找出规律,并且构造隐结构模型来解释这些规律;最后用所得到的隐结构模型来指导辨证。是一种基于模型的多维聚类分析。由于中医症状的出现与否,不是相互独立而是相互联系的。通过对症状在众多患者身上出现情况的观察,可以发现它们之间的关联关系,这些关系为寻找症状背后的隐结构提供基础。用电脑构造的隐结构来指导辨证的好处是一定程度上排除人为因素对辨证结果的影响,客观的反映症状间的关联关系。

隐结构模型^[10]是一个统计学模型,其质量的高低可以从几个方面来判断:模型是否贴切地反映了实际情况;模型的内容是否清晰;模型是否完备;模型是否简洁;模型是否有自相矛盾的地方。本研究主要从定性层面对模型与中医理论进行比较,判断模型的质量。通过对本研究中的隐结构模型分析可以看出,与中医理论及脾系疾病的证候特点比较吻合。在一定程度上体现了脾系疾病问诊信息中症状间的关联,以及症状与证候间的联系,反映了脾系疾病的内在病因病机。尤其是 Y0 因子说明肝郁可能是脾系疾病的共同发病基础,进一步突出脾系疾病与肝的密切关系,为中医问诊的客观化研究提供了有益的借鉴。

但是 Y3、Y5、Y7 因子比较不容易解释,可能是由于脾系病证包含的西医消化系统疾病比较广泛,

证候分布相对分散;同时本研究的样本含量可能尚不足以全面的反映脾系病证的证候特点,考虑扩大样本量,进行更深入的探讨。

参考文献

- [1] 张连文,袁世宏. 隐结构模型与中医辨证研究[J]. 北京中医药大学学报, 2006, 29(6): 365-369.
 - [2] 邓铁涛. 中医诊断学[M]. 上海:上海科学技术出版社, 1984: 1-190.
 - [3] 王忆勤. 中医诊断学[M]. 北京:中国中医药出版社, 2004: 8-189.
 - [4] 张伯臾. 中医内科学[M]. 上海:上海科学技术出版社, 2000: 103-116.
 - [5] 王永炎. 中医内科学[M]. 上海:上海科学技术出版社, 1997: 100-132.
 - [6] 周仲瑛. 中医内科学[M]. 北京:中国中医药出版社, 2003: 80-120.
 - [7] 国家中医药管理局. 中医病证诊断疗效标准[S]. 南京:南京大学出版社, 1994: 19-20.
 - [8] 国家技术监督局. 中医临床诊疗术语·证候部分[M]. 北京:中国标准出版社, 1997: 17-20.
 - [9] 叶任高. 内科学[M]. 5 版. 北京:人民卫生出版社, 2002: 373-500.
 - [10] 张连文,郭海鹏. 贝叶斯网络引论[M]. 北京:科学出版社, 2006: 194-229.
 - [11] Schwarz G. Estimating the dimension of a model [J]. Ann Stat, 1978, 6(2): 461-464.
 - [12] Zhang LN. Hierarchical latent class models for cluster analysis [J]. J Mach Learn Res, 2004, 5(6): 697-723.
- (收稿日期:2010-09-08;修回日期:2010-11-30)

[编辑:邓 媛]

第十一次全国中医妇科学术大会有奖征文通知

由中华中医药学会妇科分会主办“第十一次全国中医妇科学术大会”,拟定于 2011 年 10 月在河南省郑州市举行。现将有关征文事宜通知如下:

征文内容:1. 名老中医诊治妇科出血性疾病的学术思想及特色;2. 妇科出血性疾病中医诊治新进展;3. 新技术、新疗法、新工艺在诊治妇科出血性疾病及其在相关科学研究中的应用;4. 妇科出血性疾病的文献综述;5. 阿胶及其为主组方在妇科及其它相关学科领域的应用;6. 阿胶药效物质基础、提取制备工艺及质量跟踪分析等研究技术新进展;7. 阿胶及其为主组方文献综述等。

征文要求:1. 论文不超过 4000 字,并提供 200 字以内中文摘要,摘要按目的、方法、结果及结论 4 部分撰写。2. 仅接收电子投稿(fkfh189@163.com)。3. 截稿日期:2011 年 8 月 30 日。

论文一经选用,即收入学术大会论文集,颁发中华中医药学会论文证书,并通知作者出席会议,经考核合格授予国家级继续教育 I 类学分。届时还将评选优秀论文,颁发“东阿阿胶杯”一、二、三等奖。欢迎中医、中西医结合妇科及其他相关学科领域临床、教学与科研人员投稿。

联系人:刘丹卓,电话:13787132521,传真:0731-85381189,电子邮箱:fkfh189@163.com

网址: <http://www.zhzyyfkjw.com>