

中医临床个体化诊疗信息特征 与数据挖掘技术应用分析*

□胡镜清** 刘保延 王永炎 (中国中医研究院 北京 100700)

摘要:从信息学的角度分析中医临床个体化诊疗信息多维与多阶、多态与多粒度和缺失、折叠与混杂的特征,归纳了应用数据挖掘技术开展中医药临床个体诊疗规律研究的主题应包括疾病变化规律及其影响因素、证候特征与诊断规范化、中医临床诊疗经验总结、药物/疗法有效性和安全性评价与治疗方案的优化和卫生经济学等方面,旨在进一步开启多学科、多视角进行中医临床研究的思路,促进数据挖掘和中医临床领域专家对相关问题的深入研究。

关键词:中医临床个体化诊疗 数据挖掘 信息技术

提出临床诊断和治疗的思维决策过程是个体化的。医生详细地收集患病或功能失调个体的临床表现(信息),选择合适的诊断方法认识疾病或失调状态,再根据所患疾病或失调状态的类型、阶段、原因和其社会、心理、经济情况等,给予恰当地治疗或调理,追求最为有利的结局。从信息学的角度观察,其实质就是信息的收集、加工和科学决策过程。中医学“辨证论治”是

典型的个体化诊疗模式,是符合临床思维的本质特征和临床医学的发展方向。建立中医药临床个体化诊疗信息采集平台,全面整合分析其中的规律,是建立符合中医药临床模式的方法学体系和推动中医药现代化发展有益尝试^[1]。本文试图从信息学和数据挖掘技术应用的层面,分析中医临床个体化诊疗信息特征,探讨数据挖掘技术在中医临床个体化诊疗信息分析与挖掘过程中的可能应用主题。

一、中医临床个体化诊疗信息特征

中医学在临床诊疗的过程中,不仅强调注意辨析病人的特征信息,还强调全面收集病人整体信息以及相关外部环境的信息。不仅如此,假如我们要建立全面整合分析病人诊疗信息的数据仓库与分析平台的话,还涉及到病人信息在现代医疗卫生保健体系中“流动情形”的分析。我们认为,中医临床个体化诊疗具有以下信息特征:

收稿日期:2004-01-02

* 北京市重大科技计划项目(H020920010130):中医药防治重大疾病临床个体化诊疗信息平台及其标准规范的研究,负责人:刘保延。

** 联系人:胡镜清,中国中医研究院临床评价中心,副研究员,从事中医药临床疗效评价及其方法学研究,Tel: 010-64014411-2407, Email: hujingqing@mail.cintcm.ca.cn。

1. **多维与多阶**。要准确辨析病人的病理变化本质,病人信息收集是否全面是至关重要的。这其中包括病史、症状、体征、特殊实验室检查、治疗情况,以及发病季节、发病影响因素,还包括体质、既往患病情况、心理因素和社会、政治、经济、环境等信息。可以说,临床诊疗的信息维度非常多。同时,在诊疗过程中,信息的时间性即多阶的特征也是临床个体诊疗信息的重要特征。对疾病本质的认识和医疗决策的价值来说,截面所获取的信息相对于流动的信息往往只能提供较有限的认识。多维多阶是我们面对临床信息是所遇到第一个难题,降维降阶是对多维多阶数据进行预处理时经常采取的方法。但有时为增加对数据和结果的分辨能力,提高判别效率,合理的升维升阶也是必要的。

2. **多态与多粒度**。临床信息的多态性主要表现在两个方面:一是数据类型的复杂性。临床诊疗信息应该包含了所有不同的数据类型。其次是信息采集的个性化。由于人体信息的采集永远不可能采用工业样品检测那样的方式完全由机器来完成,信息因为采集者不同而表现出多态性的情形非常常见,所谓“观察者再现于观察结果中”了。这一点在中医临床诊疗过程中表现似乎更为突出。数据粒度是数据挖掘单元的大小。它好像是我们进行数据挖掘时才应该考虑的问题。但实际上为了今后的数据挖掘,我们必须在信息采集时先期考虑数据的

粒度。不同属性的粒度常有差异,即使同一属性在不同的数据挖掘目标下也需要定义不同的数据粒度。经常很难在数据挖掘前就明确挖掘目标,解决的权宜之计往往只能确定较小的数据粒度以保证信息采集的深度和方便数据粒度的组合。这常需要我们面对不同粒度的数据。

3. **缺失、折叠与混杂**。临床诊疗信息全面采集的另外一个困难是信息的不完整性即缺失。信息缺失的概念应该不仅仅是在数据的“量”的完整性上,而应该注重在符合研究目标要求的系统的“质”上。我们常常在需要进行统计分析和数据挖掘时发现因为信息收集时覆盖的广度不够而导致信息严重缺失。但应该注意的是,有些信息的缺失恰恰是因为我们对信息的广度提了不切实际的要求所导致的。另外一种情形是,很多时候病人会连续在不同的医院/部门/医生处就诊(比如一个病人可以因为胸痛在心血管科挂号就诊后再到呼吸科重新挂号就诊)。这样就会在不同的医院/部门/医生处出现该病人相似的病情信息和有部分差异的(医生)诊疗信息。我们暂且将这种情况称为“信息折叠”,以便与完全的数据重复区分开。这种信息流动过程中所产生的特殊情形给我们的数据处理技术和分析技术都提出了新的要求。临床诊疗信息的混杂也是常见的。人具有生物属性,同时还具有社会属性。疾病和健康的影响因素也是错综复杂互为

因果的。在信息分析中特别是因果判断时必须时时注意相关与因果关系的甄别。

二、数据挖掘技术在中医药临床研究的应用

数据挖掘(Data Mining, DM)是适应大量信息储存之后进一步分析其内在规律的要求而发展起来的信息技术,是从数据中发现有用模式的过程。数据挖掘技术主要包括分类、估计、预测、组合或关联、聚类、描述与可视化^[2]等6大类。近年来决策树和神经网络等算法受到关注。在医药研究领域的应用中,较早应用数据挖掘方法于活性药物筛选^[3]、基因组及蛋白组分析^[4]中,但现正逐步扩展到临床医学研究领域^[5]。在中医药临床个体诊疗规律的研究中,如将数据的收集与临床诊疗过程结合起来,建立临床诊疗采集平台,进而开展数据挖掘是大有可为的。我们认为至少在以下几方面可以尝试应用数据挖掘技术开展中医药临床诊疗规律的研究(需要说明的是,由于应用数据挖掘技术开展中医药临床的研究刚刚开始,以下例子多为西医的研究。希望未来的3~5年内应用数据挖掘技术开展中医药的临床研究会有所增加)。

1. **疾病变化规律及其影响因素的研究**。包括疾病发生危险因素的研究;疾病变化规律(基本特征与变化阶段)的研究;疾病预后及其影响因素的研究等。如台湾Hsia^[6]应用神经网络算法建立肺

癌生存期影响因素与预测模型, 结果提示影响生存期的主要因素为临床确诊时的临床分期及其病理分型和对治疗的反应。模型预测的患者生存期 13.16 ± 1.77 月与随访的真实生存期 12.44 ± 7.95 月非常接近。

2. 证候特征与诊断规范化研究。证候研究是中医药理论与临床研究的关键之一, 其主要包括证候基本特征、诊断标准与规范化等研究。国内广泛应用聚类、主成分/因子分析等方法研究证候的基本特征。王炳和等^[7]在脉搏源-系统模型的基础上, 采用信号检测与系统分析原理提出一种估计人体脉搏系统声传递函数的简便有效的方法。结果表明正常人平脉脉搏系统通常具有 3 个共振峰, 滑脉脉搏系统有 2 个共振峰, 弦脉脉搏系统出现 4 个共振峰, 而细脉仅存在 1 个共振峰。共振峰频率反映了动脉系统的共振特性。对 63 例未知脉象识别和检验, 其中平脉的准确率为 76.5%, 滑脉准确率为 75.0%, 弦脉准确率为 88.5%, 细脉准确率为 62.5%。

3. 中医临床诊疗经验总结的研究。中医临床诊疗规律的研究与老中医临床经验的总结也是中医药临床研究的重要内容。包括辨证方法与辨证策略的研究; “四诊合参” 诊断规律的研究; 处方行为的研究; 用药经验总结等。20 世纪 80 年代前后开展的中医专家系统的工程研究也属于这方面的研究内容。但当时企图用计算机代替老中医的思维, 非常困难。但

它为应用信息技术开展中医临床思维的研究进行了有益的尝试。

4. 药物/疗法有效性和安全性评价与治疗方案的优化研究。包括药物/疗法临床疗效的评价; 临床疗效影响因素的研究; 个体化治疗方案的确定与疗效预测; 治疗方案的优化; 不良反应/事件监测等。如 Gardner SN 等^[8]提出应用机器计算模型可能是确立癌症个体化治疗方案的有效方法。Cerrito^[9]提出可以通过收集病人诊疗全程的情况, 应用数据挖掘的方法分析临床复合治疗用药的不良反应, 以补充临床试验中因严格限制观察病例及其用药难以观察到多种药物相互作用所可能产生的不良反应。

5. 卫生经济学研究。卫生经济学 (Health Economics) 是应用经济学的基本原理和方法, 研究以提高人民健康为宗旨的医药卫生服务资源如何最佳分配和有效利用的问题。近年来, 医疗费用持续上涨、人口老龄化对卫生资源的压力增加等因素更加重视卫生经济学研究。卫生经济研究的突出特点在于其不确定性: 疾病的发生具有很大的随机性; 而且对疾病的治疗也具有很大的不确定性, 治疗所需的时间和最终的疗效都是很难确定的。同时, 疾病及其治疗方法多种多样, 由此带来的投入产出的效率差异很大。而大量医院/疾病控制等部门海量数据的积累和数据挖掘技术的发展, 在一定程度上为解决卫生经济学中的不确定性提供了可能。

以上探讨, 旨在努力开启多学科、多视角进行中医临床研究的思路, 建立彰显中医特色与优势适应中医药临床诊疗模式的方法学体系。

参考文献

- 1 刘保延, 胡镜清, 谢雁鸣等. 中医药学现代个体诊疗体系建立的构想与研究. 世界科学技术·中医药现代化, 2003, 5(1): 1~5.
- 2 Michael J. A. Berry, Gordon S. Linoff (袁卫等译). 数据挖掘—客户关系管理的科学与艺术. 北京: 中国财政经济出版社. 2003年, 第5页.
- 3 Marx KA, O'Neil P, Hoffman P, et al. Data mining the NCI cancer cell Line compound GI(50) values: identifying quinone subtypes effective against melanoma and leukemia cell classes. J Chem Inf Comput Sci. 2003, 43(5): 1652~1667.
- 4 Peterson C, Ringner M. Analyzing tumor gene expression profiles. Artif Intell Med. 2003, 28(1): 59~74.
- 5 Lavrac N. Selected techniques for data mining in medicine Artif Intell Med. 1999, 16(1): 3~23.
- 6 Hsia TC, Chiang HC, Chiang D, et al Prediction of survival in surgical unresectable lung cancer by artificial neural networks including genetic polymorphisms and clinical parameters. J Clin Lab Anal. 2003, 17(6): 229~234.
- 7 王炳和, 相敬林, 杨颢等. 基于信号检测的人体脉搏系统传递函数的估计. 科学通报, 1999, 44(10): 1069, 1073.
- 8 Gardner SN, Fernandes M. New tools for cancer chemotherapy: computational assistance for tailoring treatments. Mol Cancer Ther. 2003, 2(10): 1079~1084.
- 9 Cerrito P. Application of data mining for examining polypharmacy and adverse effects in cardiology patients. Cardiovasc Toxicol. 2001, 1(3): 177~179.

(责任编辑: 柳莎 刘维杰)

Progress in Study on Technology for Sperfine Powder of Chinese Medicines and Consideration Caused Herefrom

Chen Changzhou (Institute of Traditional Chinese Medicine of Guangdong Province, Guangzhou 510095)

Guo Yongzhuang and Feng Yanni (Yifang Phamarceutical Co. Ltd. of Guangdong Province, Nanhai 528244)

The study and application of the technology for superfine powder of Chinese medicines have grown in a collaboration of interdisciplines and in various areas of industry recent years and encouraging successes have been made in them. Nevertheless, there are still a lot of problems to be considered and explored in its deep study and in the popularization of its application.

Key Words: Chinese medicine, technology for superfine powder

Application of Micro – wave Technology to Extraction of Effective Components of Chinese Medicinal Herbs

Wang Yinghao (Fujian College of Traditional Chinese Medicine, Fuzhou 350003)

Micro – wave technology has been commonly used in the processing of Chinese medicinal herbs, but people have just begun to explore how to use it as a new means to extract effective components of them. In this article a feasibility study of this method is made and some problems concerned and the measures of their resolution are discussed.

Key Words: micro – wave technology, Chinese medicinal herb, extraction

Analyses on Characteristics of Information from Individualized Diagnosis and on Application of Data – digging Technology in Clinics of Traditional Chinese Medicine

Hu Jingqing, Liu Baoyan and Wang Yongyan (China Academy of Traditional Chinese Medicine, Beijing 100700)

This article analyzes from the angle of information science the characteristics of multi – dimension and multi – stage, the polymorphism and multi – size as well as the lack and loss, the folding and the mixture of information from individualized diagnosis in the clinical treatment of traditional Chinese medicine, and indicates that in the study of the rules for individualized diagnosis in TCM by data – digging technology the subjects of the study should include the changing laws of diseases and their influential factors, the characteristics of symptoms and the normalization of diagnosis, the summing – up of experience in clinical treatment, the assessment of the effectiveness and safety of drugs/therapies, the optimization of therapeutic schedules, the hygienic economics , etc. with the aim of further opening ideas for the study of clinics in TCM at the level of multi – disciplines and from many viewpoints and promoting experts in the fields of data digging and clinics in TCM to engage in deep studies of relevant problems.

Key Words: individualized diagnosis in TCM, data digging

A Comparison between Traditional Five – elements Theory in China and Parameters of Statistic Dynamics in Distribution of Life Power Elements of Human Organs

—Annotation of Traditional Theories of TCM by Theories of Modern Chemicphysics and Quantum Statistic Dynamics (V)

Jin Riguang, Li Yuanzhu and Mou Xueyan

(Life Sciences Research Center of Beijing University of Chemical Technology,

Beijing Institute of Sub – cluster Life Power Technology , Beijing 100029,

China and School of Medical Science, Dae Zen University, Republic of Korea)

[*World Science and Technology / Modernization of Traditional Chinese Medicine and Materia Medica*] 77