

数据挖掘在中医诊断学中的应用

薛飞飞¹ 陈家旭²

(1. 暨南大学医学院中医学系, 广东省广州市黄埔大道西601号, 510632; 2. 北京中医药大学诊断教研室)

[摘要] 认为数据挖掘在中医诊断学中有非常广泛的应用, 主要体现在中医四诊客观化和中医证候规范化两个方面。应用数据挖掘技术应该注意以下三个方面: ①数据挖掘专家与医学专家相结合; ②原始数据的全面性与有效性相结合; ③中医文献的原始性和标准化相结合。

[关键词] 数据挖掘; 中医诊断学; 四诊客观化研究

中医药是中华民族的瑰宝。有关中医药的古医籍和现代期刊为中医的发展积累了大量的原始数据, 但这些数据大多都以零散的形式, 存储在各种各样的文献或资料中。面对如此海量而且无序的数据, 仅凭个人思维综述分析结果, 会受主观因素、认识水平、思维方式及其研究范围和研究方法的限制, 缺乏全面系统的总结。在知识经济时代的今天, 利用先进的信息技术, 加速中医药信息的全球性传播和利用, 加快有效信息转化为知识, 促进中医药知识快速更新, 已成为中医药学向前发展的关键所在^[1]。

数据挖掘(Data Mining, DM)是为了解决“数据丰富, 知识贫乏”状况应运而生的, 是从海量数据中获取知识的可靠技术。随着数据库技术的飞速发展及数据库管理系统的广泛应用, 各个领域的数据库或数据仓库里面都收集了海量数据, 现在人们已经不再满足于对数据库进行简单的查询, 而是希望借助现代信息处理技术, 能够得到隐藏在数据中反映事物本质和预测事物发展趋势的有用知识, 并以这些知识为基础辅助科学决策^[2]。

正确的诊断对于确立治则、指导用药及调理康复无疑是非常重要的。在临床中有些疾病错综复杂, 数据挖掘的有关分类分析可以应用于疾病的诊断, 如采用贝叶斯学习分类法对男女患者的CT图像进行自动诊断; 用推导机器学习方法对重症监护者呼吸的压力、体积曲线进行分析等。数据挖掘作为中医药现代化研究的重要组成部分, 在中医诊断学中同样有着非常广泛的应用, 主要体现在中医四诊客观化和中医证候规范化两个方面。

1 数据挖掘在四诊客观化研究中的应用

四诊客观化是中医诊断客观化的基础条件, 对古

今大量四诊文献进行数据挖掘也是四诊客观化的一种手段。如将中医医案与中医的舌诊、脉诊理论相结合, 通过建立中医病案OLAP系统, 对从一些著名医案著作中收集到的上千例案例进行实例研究, 从而了解当代名医大家对于舌脉诊的认识, 舌脉象与病证的相关性^[3]。对脉诊的客观化研究也是中医诊断领域中的一个重要内容。如利用脉图数据对平、弦、滑等脉的诊断, 采用逐步判别分析法、最大似然法、Fisher逐层判别分析法等3种多因素分析方法建立判别式, 结果符合程度较高^[4]。也有研究将明清及近现代3000例医案的相关内容分为病名、证型、舌象、脉象、症状等5种模式, 尝试采用Fayyad模型, 成功建立了病名与脉象之间、脉象与病名之间、证型与脉象之间、脉象与证型之间的联系^[5]。

这种将传统的研究方法与数据库技术、人工智能技术、统计学知识的有机结合, 将是解决大量数据难以处理这一研究瓶颈的行之有效的方法, 将其运用于中医四诊的研究中, 可促进中医诊断学的规范化研究, 也为中医的现代化注入了新的活力。

2 数据挖掘在中医证候研究中的应用

从20世纪50年代起, 医学界就开始了对于中医证候规范化的研究, 其方法和手段层出不穷。数据挖掘这个新兴的工具问世以来, 从开始单纯的技术层面, 逐渐进入了证候的研究领域。有学者^[6]提出了面向中医辨证规范的数据挖掘基本内容, 其中病证所属症状体征的规范包括病证所属症状术语规范、症状间逻辑关系规范、症状体征分级规范(三级量表的建立)、症状体征测量方法规范、症状体征诊断规范、体征(舌、脉象)诊断客观化等, 很有指导意义。

2.1 医学多元统计学

多元分析是定量分析事物复杂关系的一种数理统计方法。目前人们在中医药研究中运用的多元分析方法主要有聚类分析法、判别分析法、回归分析法、因子

分析法、相关分析法等。多元分析在中医药研究中,尤其在中医证候学研究方面的应用日趋广泛并显现出其重要性。

2.1.1 聚类分析法 聚类分析是研究“物以类聚”的数理统计方法。如将与心气虚证所含症状有重叠的常见中医证候进行聚类挖掘,通过分析它们与心气虚证的相似程度,证明肺气虚证与心气虚证的相似程度最高^[7]。或运用流行病学和变量聚类分析的数理统计方法,将中医肝证的几个主要证型最后分类为实证、风证和虚证等3类^[8]。研究结果不仅科学地解决了中医肝证的分类层次问题,而且从统计学角度明确了中医肝证分类的正确性。

2.1.2 判别分析法 判别分析法是根据某些指标的观测值对所研究的对象判断其归属类别的一种统计分析方法,在中医理论研究中,多类判别较为常用,这主要是由于中医学认识事物的方法和特点所决定,中医药研究中的应用主要集中在诊断、证候以及病的计量研究方面。如应用贝叶斯网络方法通过分析上百例血瘀证临床诊断数据进行血瘀证定量诊断,发现了血瘀证的7个关键症状,并定量计算其诊断贡献度,基于这些关键症状建立的简单贝叶斯分类器模型对血瘀证诊断的准确率达到96.6%^[9]。

2.1.3 回归分析 回归分析是研究随机变量之间的统计相关关系的一种数理统计方法。近年来回归分析在中医学中的应用主要集中在对于证候的诊断和鉴别上^[10]。如对多种疾病的脾气虚证的诊断因素做逐步回归分析,对回归方程中的17个脾气虚证的诊断因素进行了分析,并将这17种因素确立为脾气虚证诊断标准的主要参考因素。也可用 Logistic 逐步回归分析方法模拟专家对疾病诊断过程,结果与专家诊断的总符合率很高^[11]。

2.2 神经网络方法

神经网络方法是继机器学习方法、统计学方法之后的另外一种数据挖掘的手段。由于传统的医学统计方法具有局限性,研究者们开始尝试运用模糊智能技术探索中医证候分类及疗效评估的方法,并用实验的方法论证了其方法上的可行性。如有研究^[12]构造了一个神经网络分级系统,并用于类风湿性关节炎寒热错杂证的病情分级诊断,获得了很好的临床吻合率。说明应用数学建模、神经网络、基于案例的推理、图像处理、遗传算法、智能数据与时序分析、模式识别等多种技术深入地研究某种疾病,以建立更为实用、高效的疾病诊断智能系统^[13]。同时还可对疗效评价、方案选优之类的问题进行辅助决策,充分发挥了数值分析的优势,即

定量和定性的分析。

3 思考和展望

对中医药信息进行文本数据挖掘是促进中医药信息结构化的途径之一。中医的特殊性决定了其从中医诊断、辨证到组方的各个环节,文字描述的定性内容占很大比例,使中医信息在一定程度上具有不确定和不完全性,同时也使对这些定性的内容进行计算机量化显得尤为困难。因此,选择合适的工具对中医药大量数据进行多层面智能分析,是进一步更新和完善中医药知识的有效途径^[12]。对以古语言和纯文本为主的中医药理论和实践进行结构化解析是中医药信息化研究的重要内容,其中的某些内容可以通过对文本的数据挖掘来实现。

数据挖掘在中医学上的应用远远不止中医诊断这个领域,它几乎涉猎了中医学的所有内容,无论是对古医籍的整理,还是对临床疾病诊断、治疗、方剂和中药的探索,各领域的学者们都做了大量的努力,使中医现代化研究越来越活跃。尽管如此,数据挖掘技术仍然面临着许多问题和挑战。总结起来,应该注意以下几个方面。

3.1 数据挖掘专家与医学专家相结合

数据挖掘的结果最终需要转换成为能够被用户理解的知识,并加以解释和评价,确定所发现的东西是否有价值。脱离了中医理论指导的挖掘结果是没有意义的。我们所挖掘到的知识必须是精确的,而且使用者是可以理解的,合乎医学常识的,可以为使用者进行决策提供坚实的基础,这就决定了整个挖掘过程需要医学专家的审查和指导,使医学数据库知识发现和医学专家知识之间真正达成一致。

3.2 原始数据的全面性与有效性相结合

数据挖掘要求所需要挖掘的原始数据应该涵盖可以影响挖掘目标的所有的知识元,以保证数据的完整性,这就要求收集资料的人员应该非常熟悉相关领域的背景知识。而有效性是指在数据筛选和预处理过程中,尽量保持数据的一致性,去除可能会影响到结果的噪音数据,如有需要可单独分析。

3.3 中医文献的原始性和标准化相结合

中医的文献描述多数是主观的、模糊的、定性的,比如:纳差、纳呆、厌食、少食等。而进行数据挖掘的资料必须是经过标准化之后的定量数据或者是二分类数据,因此,如何使数据经过清洗后仍然保留最多的信息,也是当今数据挖掘的一个需要考虑的问题。

总之,我们可以看出数据挖掘作为一个在海量数据中获取知识的有力工具,目前已经不再仅仅限于技

术层面的研究,而是越来越多地扩展到不同领域的应用中,进行中医药的数据挖掘是一个有着非常美好前景但又充满挑战的研究方向,无论在疾病的诊断和治疗方面,还是医学科研与教学方面,甚至医院的管理方面,数据挖掘都将会发挥出巨大的作用。

参考文献

[1]田琳,阎英杰,朱建贵,等.数据挖掘及其在中医药领域中的应用[J].中国中医基础医学杂志,2005,11(9):710-712.
[2]姚美村,袁月梅,艾路,等.数据挖掘及其在中医药现代化研究中的应用[J].北京中医药大学学报,2002,25(5):20-23.
[3]陈涛.基于当代名医医案数据库的舌脉诊研究.北京中医药大学博士研究生学位论文,2006:5.
[4]王洁贞,束怀符,肖琪.几种常见中医脉象的计量诊断方法[J].山东医学院学报,1983,21(4):22-35.
[5]刘晋平,黄宇虹,陆小左.数据挖掘在中医脉诊中的应用[J].天津中医学院学报,2003,22(3):9-10.
[6]王波,张斌,魏伟杰,等.面向中医辨证规范的交互式数据挖掘框架[J].世界科学技术——中医药现代化,2006,8(1):24

-30.
[7]刘强,刘璨,朱文锋.数据挖掘:心气虚证的鉴别诊断研究[J].福建中医药,2005,36(3):5-6.
[8]张世筠,沈明秀,王先春,等.中医肝证的变量聚类分析[J].中国中西医结合杂志,2004,24(1):75-76.
[9]瞿海斌,王阶.一种基于数据挖掘的中医定量诊断方法[J].北京中医药大学学报,2005,28(1):4-7.
[10]刘士敬,杨维益,朱倩.中医各科脾气虚证诊断因素的多元逐步回归分析[J].辽宁中医杂志,1996,23(9):395-397.
[11]刘凤斌,郝元涛,方积乾. Logistic 逐步回归分析方法在模拟专家辨证诊断中的应用研究[J].中国中医基础医学杂志,2001,7(2):58-59.
[12]周志坚,毛宗源,邓兆智.神经网络在类风湿性关节炎病情分级中的应用初探[J].生物医学工程学杂志,1999,16(4):479-482.
[13]张琴,刘平,张文彤.数据挖掘技术在中医证候学研究中的应用[J].上海中医药杂志,2006,40(3):3-5.

(收稿日期:2008-11-06;修回日期:2008-12-17)

[编辑:黄健]

镇咳宁胶囊对小鼠成瘾性的影响

张庆平¹ 石森林²

(1. 义乌市中医医院,浙江省义乌市城中中路15号,322000,2. 浙江中医药大学)

咳嗽是呼吸系统疾病的常见症状,有利于清除呼吸道分泌物和有害因子,但频繁剧烈的咳嗽对患者的工作、生活和社会活动造成严重的影响。我院多年来一直沿用中成药镇咳宁胶囊,临床反馈良好。为了观察该药是否有成瘾性,我们观察了镇咳宁胶囊对小鼠成瘾性的影响,现将结果报告如下。

1 材料和方法

昆明种小鼠共100只,体重20g±4g,雌雄各半,由浙江中医药大学提供。镇咳宁胶囊(成都康弘制药有限公司,批号070518);复方磷酸可待因溶液(香港澳美公司,批号0707313);5%糖盐水(安徽双鹤公司,批号070619)。

参照谢秀琼《中药新制剂开发与应用》,采用“自然戒断试验”的方法。将100只小鼠随机分为5组,每组20只,分别给予高、中、低剂量镇咳宁(剂量分别为:1.05g/d、2.10g/d、3.15g/d,相当于临床实际用量的10倍、20倍和40倍),复方磷酸可待因溶液(剂量为:3ml/d,相当于临床实际用量)和5%糖盐水(剂量为:6ml/d),每天3次,连续30天。在停药前24h到停药后48h之间,每隔4h观察记录动物的外观体征和行为活动、植物神经系统机能变化,并测量体重。镇咳宁胶囊(每粒镇咳宁胶囊装0.35g,含盐酸麻黄碱4mg)。每10ml复方磷酸可待因溶液,含磷酸可待因10mg,盐酸麻黄碱8mg,马来酸氯苯那敏2mg,氯化氨220mg。

2 结果

镇咳宁高、中、低剂量组,共60只小鼠,均未发生竖尾和自发性惊厥。糖盐水组,共20只小鼠,均未发生竖尾和自发性惊

厥。可待因组,共20只小鼠,其中的18只发生了竖尾反应和13只出现了自发性惊厥。经t检验,差异有统计学意义(P<0.01)。镇咳宁高、中、低剂量组,共60只小鼠,试验前后体重无明显变化。糖盐水组,共20只小鼠,试验前后体重无明显变化。可待因组,共20只小鼠,试验后平均体重出现了显著下降。经t检验,差异有统计学意义(P<0.01)。

3 讨论

镇咳宁胶囊系以宋代《太平惠民和剂局方》所载的“华盖散”为基础,结合历年临床应用经验,予以加减化裁而成,处方成分包括盐酸麻黄碱、桑白皮酊、桔梗酊和甘草流浸膏。麻黄性温散寒,是为君药,桑白皮和桔梗化痰利咽止咳,共为臣药,佐以甘草调和诸药,并寓有桔梗甘草汤之意,使本品外可疏风散寒解表,内可宣肺止咳护本,表里兼治,共奏散寒止咳之效。

现代药理研究发现,镇咳宁胶囊处方巧妙,各味成分间治疗作用表现为相加,而不良反应则相互抵消。这使得镇咳宁胶囊在临床上表现出了较其他中成药更理想的镇咳、祛痰作用。镇咳宁胶囊中所含盐酸麻黄碱是非选择性的肾上腺素能受体激动剂,而桑白皮酊的主要有效成分东莨菪内酯则为非选择性的胆碱能受体阻断剂,二者对眼压、血压等的作用相反,但都能舒张支气管平滑肌。镇咳宁胶囊中桔梗酊的主要有效成分桔梗皂苷,和甘草流浸膏中的主要有效成分甘草酸苷,均为肾上腺皮质激素类似物,具有明显的抗炎作用,但二者对代谢的影响却是相反的。

(收稿日期:2008-11-20)

[编辑:黄健]