

数据挖掘在中医脉诊中的应用

刘晋平,黄宇虹,陆小左

关键词:数据挖掘;病名;脉象;证型;脉象软件

中图分类号:R241.1 文献标识码:A 文章编号:1005-7145(2003)03-0009-02

随着科学技术的迅猛发展及中医脉诊客观化的深入研究,人们积累了大量的原始数据,而伴随着数据的不断增加,人们发现往往难以通过简单的总结来理解数据中包含的大量信息,从而形成了“资料丰富而知识缺乏”的局面。在传统中医理论的指导下,人们往往是事先知道结果,再去寻找其理论根源和实践佐证,结果大多是凭经验而来的,这是与数据挖掘最主要的区别,当数据丰富,超越了人们固有的思维习惯时,人们对其认识是不全面的,散在的,通过数据挖掘才能提取出其新颖、未知的知识,即数据挖掘的事先未知性、隐含性。笔者认为在中医的脉诊研究中,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办法,将数据库技术等计算机基础研究方法运用于中医的脉诊等的研究,也为中医的现代化注入了新的活力。

本文以明清、近现代 3 000 余例病案为研究对象,运用计算机数据挖掘技术,将病案分为病名、证型、脉象、舌象及症状几项,然后进行统一化及规范化处理,以中医理论为指导,通过对各项因素的单一统计、相关性分析来探求疾病与其各个构成因素以及各个因素之间的相关影响,进行大量统计、运算分析,从一个全新的角度来认识其本质及规律。

1 资料与方法

数据挖掘在中医脉象研究中的运用是本文需要具体讨论的内容,本文对 3 006 例明清及近现代医案、医话的分析、统计是基于数据挖掘模型而提出的。

1.1 数据(病例采集) 针对确切的明清及近现代

3006 例医案的一个集合,它可以基本描述疾病的特点、情况,这些原始数据基本上是有据可依的,来源于古代医案、医籍。

1.2 方法

1) 模式:对于原始病案来讲,可分为五大类别,分述其各自特点。在这里五大类分别为:病名,证型,舌象,脉象,症状,此称为五种模式。

2) 数据挖掘的处理过程

本文所提到的数据挖掘与病案原始数据库中的知识是客观的,具有同样广泛的意义,在从整个原始数据到最终知识的处理过程中包括原始医案数据的采集、数据的预处理运行挖掘算法,知识表达(脉象软件)等。目前有几类处理手段模型,如多处理过程的 Fayyad 模型、John 模型等。本文采用 Fayyad 模型。是对 Fayyad 模型的简说明,本文是把病案分析看作一个多阶梯处理过程,每个过程之间是相互关联的。

2 结论

本研究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。

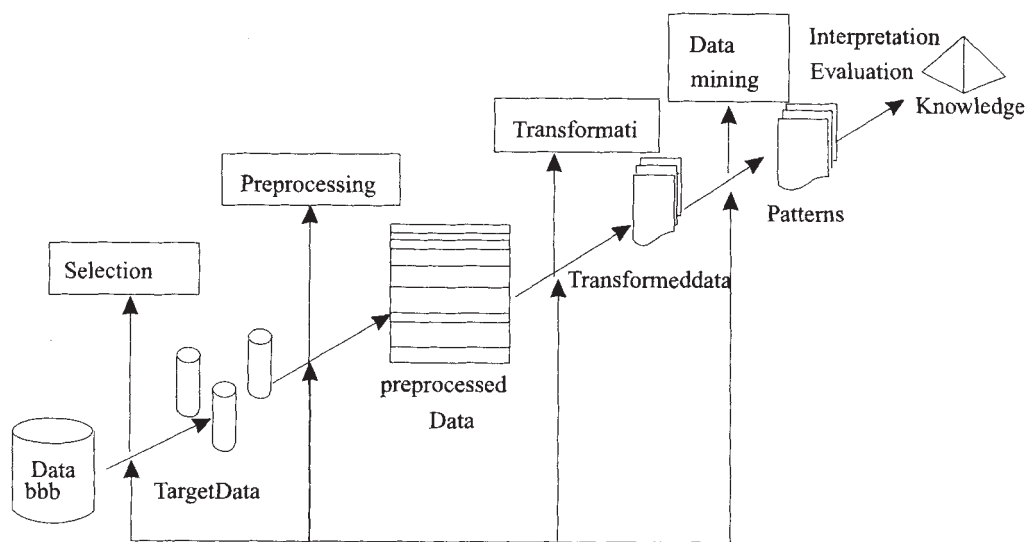
1) 从病名推出相关脉象,并显示各脉象出现的频度,可分析得出病名与各脉象之间的联系,与哪一种脉象关系最密切,与哪一种脉象关系最疏远。

2) 从脉象推出对应疾病,并显示各疾病所占的百分比,通过比较可得出此脉象在哪一种疾病中更常见。

3) 从证型推出相应的脉象,并显示各脉象的频度,可分析得出此证型与各种脉象之间的联系,与哪一种脉象的关系最密切,与哪一种脉象的关系最疏远。

4) 从脉象推出相应证型,并显示各证型所占百分比,通过比较可得出此脉象在哪一种证型中最常见。

其全面性、系统性、明确性、简便性明显优于祖国医学传统的依靠经验所总结的结果。



Data: 医案医籍;TargetData: 纳入病例;Preprocesseddata: 医案原始记载;Transformeddata: 统一规范化数据;Patterns: 五种模型;Knowledge: 结果。

3 讨论

本课题所纳入的病例均是随机抽取,结果显示脉象出现频率顺序依次为:细、数、弦、沉、浮、滑,其反映出 3 000 余例病案脉象出现规律。1) 此六种脉象基本上与临床常见脉象相吻合,是符合临床脉象出现规律的。2) 本文数据统计结果显示,细脉共出现 2 267 次,占 34.39%,而据有关资料显示,目前部分医院住院病人中弦脉的出现率达 60%左右,而本文统计结果显示弦脉共出现 1 033 次,占 15.67%,居于第 3 位,与其调查结果有所不同。

除去时代与人群差异的因素之外,脉象诊断技术的差异应为主要原因。所选病案诊脉者均为著名中医,其技术高超,判断比较准确,目前普通医院的

中医脉诊技术尚有提高的余地。“细脉者,脉细如线,但应指明显”,“弦脉者,端直而长,如按琴弦”,两者在脉形上有相似之处,因此给医生诊脉带来了困难,临床易于混淆。脉象的形成与脏腑气血密切相关,通过诊察脉象可显示出健康状况的细微变化,不同脉象所显示的疾病不同,明辨细脉、弦脉在临床研究与实践上均有重要意义。这也从另一侧面证实,本课题的研究工作是具有临床现实意义的。但本课题只是初步尝试了数据与脉象的相关性研究。至于随着输入病例数目的增加,脉象出现频率是否会发生变化有待进一步研究。

(收稿日期:2003-04-05)

2004 年《天津中医学院学报》征订启事

《天津中医学院学报》由天津市教育委员会主管,天津中医学院主办,为综合性中医药学术期刊,是天津市教委中心核心期刊。本刊理论与实践并举,继承与发展并重,设有理论探讨、中医教学、实验研究、研究生园地、雏鹰之窗、中医药在海外、中西医结合、临床报道、针灸与推拿、临证心得、综述等栏目。

本刊为季刊,每季末月 15 日出版,每期定价 4 元,全年 16 元(4 本)。国内外公开发行,国内读者在全国各地邮局均可订阅,邮发代号 6-153。漏订者可直接向本编辑部办理订阅,地址:天津市南开区玉泉路 88 号,邮编 300193,电话:(022)23051018,传真:(022)27470216,E-mail:xuebaobj@tjutc.edu.cn。