

中药信息量子化研究^{*}

★ 吴园园 章新友^{**} (江西中医学院 南昌 330006)

摘要:从中药信息的特点出发,探讨了如何对中药信息进行“量子化”处理,并就中药信息量子化及其应用进行了阐述。

关键词:中药信息;量子化;数据库

中图分类号:R 28 **文献标识码:**A

Research on Quantization of Traditional Chinese Medicine Information

WU Yuan-yuan, ZHANG Xin-you

Jiangxi University of Traditional Chinese Medicine, Nanchang 330006

Abstract: This article begins with the characteristics of Traditional Chinese Medicine (TCM) information, discusses how to quantize the information of TCM, and elaborates its applications.

Key words: TCM information; Quantization; Database

1 引言

随着现代计算机技术以及互联网信息的不断普及,中药的各种信息正以惊人的速度迅速膨胀。面对呈指数级增长的中药信息,利用计算机进行处理已成为提高信息获取效率的有效途径,而如何处理和解析这些海量数据,从中挖掘出有用信息和未曾被人发现的隐含信息已成为人们迫切需要解决的难题。本文论述的中药信息“量子化”技术,即是将海量的中药信息“量子化”为一个不可再分的“量子”,这些“量子”经处理转换成一组可供计算机识别、计算的代码即“数字”后,计算机即可对这些“量子”进行“数字化”处理。中药信息的量子化是中药信息数字化、标准化和现代化的重要前提,不但有利于提高数据库的信息处理速度,而且也有利于提高数据库智能化水平,为进一步的数据挖掘奠定坚实基础。

2 中药信息的特点

2.1 信息量多,且呈指数级不断增长

2.1.1 浩如烟海的中药古籍文献信息资源 中药古籍文献是中国人民防病治病经验总结,也是中医药重要的信息资源。据统计,目前国内收藏的辛亥革命以前的中医学古籍

文献高达 13 000 多种,其中在社会上流通较广的古籍近 1 000 种^[1]。

2.1.2 迅猛增长的中药现代信息资源 近些年,随着各种实验技术与网络技术的迅速发展,使得原本就浩如烟海的中药信息正以前所未有的速度增长。比如,在中药化学研究中,高效液相色谱、气相色谱、光谱、质谱、核磁共振及其联用技术已经被广泛采用,人们可以应用各种分析仪器得到大批量化学测量数据,形成巨量化学数据集^[2]。而且各种书刊、会议资料、专利文献、学位论文等种类繁多,使中医药信息量迅速增加。

2.2 信息复杂,缺乏统一性和规范性 中药信息的复杂性源于中药本身的复杂多样。例如,一种天然植物药一般含有 100 种以上化合物,且同一植物在不同的生长环境下其成分有所不同;常用方剂一般由十多种的单味中药组成,且所含中药的数量及组成比例的差异会产生不同的功效;同一中药在不同地区可能有不同的名称,不同中药在不同地区也可能有同一名称,等等。中药知识的复杂多样以及中药术语的不统一、不规范和未标准化,使中药信息缺乏统一性和规范性。

2.3 信息模糊,缺乏精确性和定量性 中药知识是我国人

^{*} 江西省教育厅科学技术研究项目(项目编号:赣教技字[2006]221号)

^{**} 通讯作者:章新友, Tel: (0791) 7118716, Fax: (0791) 7118716, Email: jzxinyou@163.com

民在长期实践过程中积累的经验总结,所以对中药的很多信息是靠经验来描述,这就造成了中药信息的模糊、不精确性。例如,中药的“毒性”只是被单纯描述为有毒、无毒、大毒、小毒、剧毒等,而对各种毒性缺少定量描述,如半数致死量(LD_{50})是多少?它们之间的划分标准又是什么?在中药信息中并没有给出精确、定量的描述;又如,中药的“味”可分为甘、微甘、苦、微苦、咸、微咸等,但每个人的味觉是有差别的,同一种中药有些人觉得甘,可能另一些人觉得是微甘,这之间就缺少了一个定量的划分标准。这些都使中药信息变得模糊而不够精确。

2.4 信息可分,缺乏量子性和统计性 信息可分是中药信息的又一特点,尤其是大文本信息,如“药理作用”、“不良反应”及“临床应用”等信息都是可分而未分的,即目前的中药信息还缺乏“量子”特性。正是因为信息的划分缺少了定量的划分标准,使目前的中药信息很难具备“量子”性。如果中药的各种信息能解析成一个个不可再分的、不连续的、具有独立内涵的信息单位,即当中药信息具有“量子”特性时,将更有利于中药信息的统计分析和数据库的知识发现。

3 “量子化”与中药信息

3.1 “量子化”的由来 量子化概念源自物理学,普朗克是“量子物理学”的开创者和奠基人,创立了“量子理论”。1900年,普朗克抛弃了能量是连续的传统经典物理学观念,导出了与实验完全符合的黑体辐射经验公式。在理论上导出这个公式,必须假设物质辐射的能量是不连续的,只能是某一个最小能量的整数倍,普朗克把这一最小能量单位称为“能量子”,简称“量子”。由此产生了“量子”的概念。而在1905年,爱因斯坦根据光电效应推断,光能也不是连续的。对光的“量子化”就是认为光是以一个一个微小单位的形式存在和传播的,被称为光量子(简称光子)。这就是早期量子论中的“量子化”概念。

3.2 中药信息的“量子化” 面对复杂繁多的中药信息,本文引进了物理学中的“量子化”概念,但又与物理学中的“量子化”有所不同。中药信息的“量子化”是指将纷繁复杂的中药信息合理解析成具有独立内涵的、不可再分的最小信息单位。这种最小信息单位即称为“量子”。将中药原始数据“量子化”的方法不仅使中药复杂数据简洁化、精确化、规范化和有效化,而且为中药信息的数据库知识发现提供了必要准备。

4 中药信息量子化应用

4.1 为中药信息“数字化”做好准备,提高计算机的信息处理效率 中药信息经量子化后,通过对每个量子进行编码,使各种文字信息转换成“数字”后,计算机即可对这些“量

子”进行“数字化”处理。所谓“数字化”是指将许多复杂多变的的信息转变为可以度量的数字、数据,再以这些数字、数据建立起适当的数字化模型,把它们转变为一系列二进制代码,引入计算机内部进行统一处理,这就是数字化的基本过程。例如,中药种类可处理为“植物类中药”、“动物类中药”、“矿物类中药”和“中药复方制剂”四个量子,经过编码,“植物类中药”可编码为“00”,“动物类中药”编码为“01”,“矿物类中药”为“10”,“中药复方制剂”为“11”等。中药信息的量子化是信息数字化、标准化和现代化的重要前提,它不但有利于提高计算机的信息处理速度和处理质量,而且为后续的知识发现做好了必要准备。

4.2 为数据库知识发现奠定基础,提高中药数据库的智能化水平 中药信息量子化是将纷繁复杂的中药信息处理成简洁、有效的信息单位,这在一定程度上解决了信息量大、数据复杂不易处理的困境,并为进一步的数据库知识发现奠定了基础。数据库知识发现是指从大量数据中提取出可信的、新颖的、有效的并能被人理解的模式的处理过程^[3]。从概念中可以看出,数据库知识发现技术的发展,使我们有可能从浩如烟海的中药信息中获取能够反映中医药理论的隐含的、未被人知的规律和模式,这将大大提高中药数据库的智能化水平,方便人们对中药各种知识的发现和挖掘,从而进一步加快中药现代化的进程。中药信息的量子化又是数字化和标准化的重要前提,而中药信息的数字化和标准化为数据库知识发现创造了有利条件,由此可见,中药信息的量子化为中药数据库知识发现奠定了坚实基础,并最终将有力推动中药的现代化进程。

5 结束语

伴随着中药信息数量的空前增长,将量子化技术应用于中药信息的加工整理,虽然可以使纷繁复杂的中药数据简洁化、精确化和规范化,但同时也带来了中药信息损失率增加的问题。为了克服这一缺陷,将中药量子化技术与中药全文检索技术相结合,互补短长是其有效的解决途径。我们相信,随着科学技术的不断进步和研究方法的日益创新,中药的现代化和国际化进程将逐步加快。

参考文献

- [1] 顾东蕾. 浅议网络环境下的中医药学古籍文献资源共享[J]. 中国中医药信息杂志, 2005, 12(1): 109-110.
- [2] 程翼宇, 范晓辉, 瞿海斌. 论建立与发展中药信息学[J]. 中国中医药信息杂志, 2003, 10(2): 84-86.
- [3] 张春华, 王阳. 数据挖掘技术、应用及发展趋势[J]. 信息化与网络建设, 2003, (4): 47-50.

(收稿日期: 2008-01-18)

欢 迎 投 稿 ! 欢 迎 订 阅 !