

中药特性信息数据挖掘系统设计

胡建军

(广东商学院信息学院, 广州 510320)

摘要: 为推动传统中医药的研究与临床应用, 开发从中药药方中挖掘方剂药理信息的软件系统, 阐述该软件系统的系统结构、数据库设计、预处理设计。采用归一化数据处理技术与统计技术相结合的方法, 挖掘出中药方剂的性、味、归经。采用空间矢量量化方案, 结合传统中药理论, 挖掘出中药方剂的功效。

关键词: 数据挖掘; 中药方剂; 空间矢量

Design of Data Mining System for Chinese Medicine Characteristics Information

HU Jian-jun

(Information Science School, Guangdong University of Business Studies, Guangzhou 510320)

【Abstract】 In order to improve the research and application of traditional Chinese medicine, a software system is developed to mine the pharmacology from traditional Chinese medicine prescription. The system architecture, the designation of database and data preprocessing are described. Normalized data processing and statistical techniques are used in the system to mine the properties, flavors, channel from traditional Chinese medicine prescriptions. Dimensional vector technique combining the theory of traditional Chinese medicine is used to quantize and mine the information about effect and syndromes of a prescription.

【Key words】 data mining; Chinese medicine prescription; dimensional vector

中医药在中国已有几千年的应用实践, 然而其理论还有很多不完善的地方, 传统研究方法在中医药领域遇到了很多困难, 严重影响了对传统中医药的继承和发展^[1]。数据挖掘是最新的数据处理技术, 把数据挖掘技术应用于传统的中医药方剂学理论的研究^[2-3], 从古今大量验方中挖掘出对中医药研究有价值的信息, 不仅可以为数据挖掘技术开辟新的应用领域, 还可以为中医药的研究提供新途径、新思路。

1 中药特性信息数据挖掘系统结构

中药特性信息数据挖掘系统基于 Windows 平台, 用 Delphi 语言实现了对中药方剂的性、味、归经、功效、用药规律的挖掘。主要分为数据预处理、个方分析和类方分析 3 个模块。整体结构如图 1 所示。

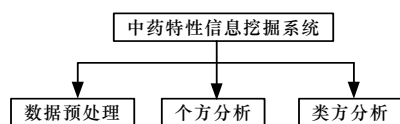


图1 中药特性信息数据挖掘系统结构

系统首先对方剂数据进行预处理, 在此基础上, 分析挖掘了单个方剂和一类方剂的药理特征和组方特点。个方分析与类方分析模块包含各自的性味挖掘、归经挖掘、功效挖掘。

2 系统的数据库设计

中药特性信息挖掘系统中涉及的数据是多方面的, 既有药的信息, 又有方的信息, 还有证的信息, 它们之间的关系更是千变万化、错综复杂。本着既要减少系统冗余, 又要便于系统快速处理, 还要具有灵活性和可扩展性的原则, 本系统设计的数据库中包含以下数据库表: 方剂表, 药物表, 症

状表, 方剂性能表, 基本药物表, 基本证候表, 症状规范表, 功效规范表, 药名规范表, 剂量转换标准表, 特殊剂量表, 类方筛选表, 类方筛选历史表。数据库中的表可分为 4 类: 基本表, 预处理表, 用户输入表, 输出表。

3 中药方剂数据挖掘过程

系统根据规范库对录入的原始方剂数据库进行预处理, 包括: 对药名、症状的规范化, 剂量的标准化转换, 过滤噪声数据, 形成数据规范、完整的方剂库。从预处理后的方剂库中选择待分析的个方或类方, 采用不同的数据挖掘方法从中挖掘出方剂的性、味、归经、功效、症候等药理特性和组方特点。系统分析流程如图 2 所示。

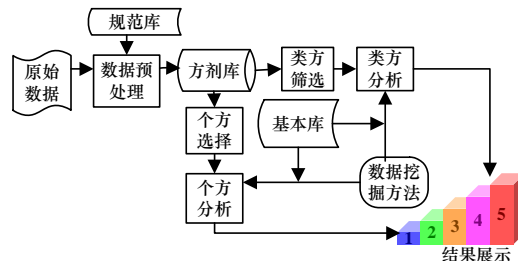


图2 系统分析流程

基金项目: 国家自然科学基金资助项目(60763012); 广东省自然科学基金资助项目(07006474); 广东省科技攻关计划基金资助项目(2007B010200044); 广东商学院博士启动基金资助项目(07BS52002)

作者简介: 胡建军(1970-), 男, 副教授、博士后, 主研方向: 商务智能, 智能信息处理, 数据库与知识工程

收稿日期: 2009-02-26 **E-mail:** jianjun_hu@163.com

4 数据预处理

对于不同朝代、不同医学专家传承下来的方剂，所用的剂量单位、药名、功效说明、症状描述等差别很大，并且有的方剂中还有错误信息和缺损数据。为了提高系统挖掘结果的可信度，在实施数据挖掘前，必须对用户录入的数据进行全面的预处理，使方剂库中的数据符合数据挖掘算法的要求。

该模块就是针对中药方剂数据库中的方剂表、药物表和症状表中诸如空白字段、错误数据、重复记录、药名和症状名等不规范名称等问题，进行过滤和修正^[4]。在整个工程中，这部分难度不大，但占用时间却比较多。

数据预处理模块的结构如图3所示。

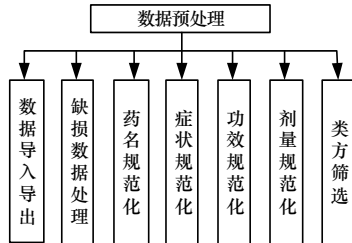


图3 数据预处理结构

5 个方分析

预处理操作完成后，方剂库中的数据已经成为标准统一、无噪声的数据。在此基础上可从中选择一首方剂进行相应的个方分析。

5.1 挖掘性味归经

5.1.1 缺损数据处理

在古方剂的记载中，有些方剂中各药的用量是没有的，有的是文字性的描述，如等量、适量、等分等。有些方剂的剂型是丸、散剂，其方中的用量和熟知的汤剂是不一样的，一般情况下是服用很多天的用量，然而又没有明确记载服用天数，这在计算机中是难以处理的。为了可以和其他方剂相互比较，本系统结合中医理论和数据挖掘算法，挖掘出方中每味药在同类方中的常用剂量，然后计算出该方中每味药的每日用量，从而把文字描述的剂量信息和丸、散剂的剂量都转换成标准的汤剂的每日用量。

5.1.2 作用度归一化处理

中药用量是药效作用发挥的基础，也是影响方剂功效改变的重要因素。不考虑用药人的意图，单就药物的质地、性味、采集部位的不同，方中各药用量也会千差万别。即使用量相同的药物，在方中发挥的作用也是不相同的。在方剂中各药的用量范围差别很大，有的药在方中每次仅用零点几克，最多只能用几克，而有的药在方中最少就要用几十克。用量范围主要是依据用药的有效性、安全性而确定的。

中药药典中记载了每味药使用的最高剂量和最低剂量，这只是常规的用药剂量范围。在临床实践中，使用的剂量有时不在这个范围内。当用量超过常规范围，则该药在方中作用显著增大，超过越多，其在方中所起作用越大；当小于常规用量时，则在方中作用显著降低^[5]。

为了使方剂中各药之间具有可比性，本文引入作用度这一概念。

定义 单味药对方剂性、味、归经的影响程度称为作用度，记为 G 。

作用度计算公式如下：

$$G = \begin{cases} \left(\frac{v}{N_{\min}}\right)^2 \times n & v < N_{\min} \\ \frac{v - N_{\min}}{N_{\max} - N_{\min}} \times (m - n) + n & N_{\min} \leq v \leq N_{\max} \\ \frac{v}{N_{\max}} \times m & v > N_{\max} \end{cases} \quad (1)$$

其中， v 表示该单味药在方中的剂量，以 g 为单位； N_{\min} 表示该药在药典中的最低用量； N_{\max} 表示在药典中的最高用量； m 为归一化上限，默认值为 60； n 为归一化下限，默认值为 10。

在上述定义中，把小于最低剂量的用药剂量映射到 $0 \sim n$ 区间内，并且当用量越小时，其作用度会显著降低。介于最高剂量和最低剂量之间的用药剂量映射到 $n \sim m$ 区间内。大于最高剂量的用药剂量映射到大于 m 的范围。

5.1.3 中药的四性、五味及归经

中药的四性、五味、归经代表了中药的药理特性，它与功效具有密切的联系，是临床用药的重要依据。

(1)性分为大热、热、温、微温、平、凉、微寒、寒、大寒，共有 9 个属性。

(2)味分为辛、苦、甘、咸、淡、涩、酸、微苦、微甘、微辛，共有 10 个属性。

(3)归经分为脾、肺、胃、肝、心、膀胱、肾、胆、大肠、心包、三焦、小肠，共 12 条经。

由于中药药理的不完善和配伍关系的复杂性，人们虽对单味药的性、味、归经有清晰的认识，但还不清楚由多味药组成复方的性、味、归经，对这一问题的认识目前还停留在个人的行医经验上。

四性分为寒热温凉平，其中寒又分为大寒、寒、微寒 3 个不同层次；热分为大热、热 2 个不同层次；温分为温、微温 2 个不同层次。按照中医理论，寒热温凉并不是通常所指的温度，它描述是人体内在体征的变化。因此，它不是一个数值量，而是一个范畴型数据，不存在一个数值上的偏序关系。例如，微温+微温=微温，而不等于温；即无论多少微温相加也不等于温，只是表示微温的量增加而已。因此，在挖掘方剂的性时，分别把性相同的单味药的作用度相加，然后比较各个性的总的的作用度，从而得出方中四性各个属性的作用度。

很多药具有 2 个以上的味，是因为它表现为 2 种以上的味倾向，它们之间没有必然的联系，也没有主次之分。这样的药在不同的方中所表现出来的味倾向是不同，并不是说只发挥了一种味的作用，而没有发挥其他味的作用。只是由于和方中其他药物相互作用后，一种味倾向和其他药的味倾向相须，以致这种味倾向得到了加强，而其他味倾向和其他药的味倾向相佐，以致被掩盖了而没有表现出来。因此，在分析具有多个味倾向的药的作用度时，应对不同味倾向分别计算。

味和归经的处理方式与性的处理方式相似。

5.1.4 个方性味归经的挖掘

在确定好每个属性上的作用度后，根据模糊评判规则，挖掘出方剂的性、味、组合性味和归经信息。

5.2 挖掘功效

5.2.1 中药功效

方剂功效是组方药物在人体内作用的综合反映，它是通
(下转第 273 页)