

# 关联挖掘技术在中药药性及其他属性间关系的应用研究

于红艳<sup>1</sup>, 许成刚<sup>2\*</sup>

(1. 河南省中医院, 郑州 450000; 2. 河南中医学院信息技术学院, 郑州 450000)

**[摘要]** **目的:**采用关联规则挖掘技术研究中药药性与其他属性间的关系规律。**方法:**以《中国药典》、《中药学》教材等书籍为信息来源,结合临床用药经验,采集、规范中药相关属性术语,构建中药信息数据库,利用改进的 Apriori 算法,对药性与其他属性之间的关系进行统计挖掘。**结果:**建立了药性-药理作用、药性-主治、药性-西医病名、药性-功能、药性-药味、药性-归经、药性-化学成分等频繁项目集,统计出中药药性与其他属性间的支持度和支持率,如药性“温”与药味“辛”支持度为 117,支持率为 10.578。**结论:**应用关联规则挖掘方法分析中药药性及其他属性间关系数据的关系,数据挖掘结果可分析性较强,与中药药性传统认识相符。

**[关键词]** 关联规则; 中药药性; 属性关系; 频繁项目集; 数据挖掘

**[中图分类号]** R285.5 **[文献标识码]** A **[文章编号]** 1005-9903(2013)14-0343-04

**[doi]** 10.11653/syjf2013140343

## Relationship between Nature and other Properties of Traditional Chinese Medicine Based on Association Rule

YU Hong-yan<sup>1</sup>, XU Cheng-gang<sup>2\*</sup>

(1. Chinese Medicine Hospital in Henan Province, Zhengzhou 450000, China;

**[收稿日期]** 20121009(017)

**[第一作者]** 于红艳, 硕士, 主管药师, 从事制剂-药品质量研究, Tel:13526600485, E-mail:yuhongyan009@126.com

**[通讯作者]** \* 许成刚, 硕士, 副教授, 从事互联网应用技术、数据挖掘研究, Tel:0371-65680187, E-mail:xuchenggang001@126.com

多药耐药相关的 MDR1, Bcl-XL, Bcl-2 等基因有下调作用<sup>[5-7]</sup>。第三类以黄芪、当归、龟甲、阿胶、麦冬、五味子为代表的补益类药物, 当归黄芪为当归补血汤核心药物, 龟甲阿胶养血补血, 麦冬五味子养阴益气。第四类健脾和胃类药, 主要包括砂仁、郁金、焦三仙、鸡内金、茯苓等。第五类安神药物, 常用的有远志、柏子仁、生牡蛎等。化疗间期白血病突出特点邪弱正虚, 邪毒内伏, 兼有脾胃损伤, 情志不遂, 所以治则扶正祛邪, 解毒化痰, 搜风剔络, 顾护后天之本, 调理情志。杨文华教授针对病人病理变化, 拟以解毒药为君药, 化痰祛风药为臣药, 佐以健脾和胃药, 安神药, 使以黄芩、甘草。组方严谨, 不变中求变, 辨证论治不同病人。

“中医传承辅助系统”得到的药物核心组合以及新处方, 有利于临床疗效研究等进一步评判, 追本溯源, 提炼新处方, 新对药, 新辨证论治体系, 为继承挖掘名老中医经验, 提高临床疗效等提供了客观、科学的新方法, 值得进一步推广应用。

### [参考文献]

- [1] 王海燕. 内科学[M]. 北京: 人民卫生出版社, 2004:9.
- [2] 李健, 卢朋, 唐仕欢, 等. 基于中医传承辅助系统的治疗肺病方剂组方规律研究[J]. 中国实验方剂学杂志, 2011, 18(2):254.
- [3] 卢笑晖, 单崎玮. 基于中医传承辅助系统分析卢尚岭教授治疗头痛病用药经验[J]. 中国实验方剂学杂志, 2012, 18(9):5.
- [4] 唐仕欢, 陈建新, 杨洪军, 等. 基于复杂系统熵聚类方法的中药新药处方发现研究思路[J]. 世界科学技术——中医药现代化, 2009, 11(2):225.
- [5] 杨文华, 郝征, 杨向东, 等. 蝎毒多肽提取物对白血病 NOD/SCID 小鼠 MMP2、MMP9 表达的影响[J]. 天津医药, 2009, 37(10):856.
- [6] 杨文华, 王慧娟, 杨向东, 等. 蝎毒多肽提取物对白血病细胞粘附力影响的实验研究[J]. 中国临床药理学杂志, 2009, 25(3):235.
- [7] 杨文华, 王慧娟, 杨向东, 等. 蝎毒多肽提取物对白血病小鼠 E-钙黏蛋白、CD49d 和 CXCR4 表达的影响[J]. 时珍国医国药, 2010, 21(2):259.

[责任编辑 邹晓翠]

2. Institute of Information Technology, Henan University of Traditional Chinese  
Medicine, Zhengzhou 450000, China)

**[Abstract]** **Objective:** Based on the association rule, to find the relationship between the traditional Chinese medicine (TCM) nature and other properties. **Method:** Many text book were used as the information source, such as ‘Chinese pharmacopoeia’, ‘Chinese materia medica’, etc. And combined with the empirical clinical medication, we builded the Chinese medicine database. Based on the improved Apriori algorithm, we analyzed the relationship between the TCM nature and other properties. **Result:** We have found the closely frequent itemsets in Nature-Pharmacology, Nature-Indication, Nature-Western medicine name, Nature-Function, Nature-Flavour, Nature-Channel tropism, Nature-Chemical composition, etc. And we calculated support and support rate in the TCM nature and other properties. **Conclusion:** Through analyzing the relationship between the TCM nature and other properties based on the association rule, the analyzability of results are reasonable which is consistent with the traditional understanding of the TCM nature.

**[Key words]** association rule; nature of traditional Chinese medicine; relationship of properties; frequent itemsets; data mining

中药药性是中医药理论的基础和核心,是中医、中药衔接的桥梁,是临床应用的纲领,是指导中医临床用药的重要理论依据<sup>[1-2]</sup>。随着中医药现代化的发展,药性药味理论的现代研究越来越成为国内学者关注的热点,而目前很多中药属性同药性理论的联系尚缺乏可以遵循的规律<sup>[3]</sup>。借助数据挖掘等技术方法揭示其科学内涵,挖掘相关中药数据的潜在信息,是中药药性理论现代研究的首要任务。

数据挖掘是从大量的数据中发现有效的、新颖的及潜在有用的信息并且能够转化为可理解的知识的过程<sup>[4]</sup>。将数据挖掘技术应用于中药属性间关系研究,不仅有利于中药信息结构化,促进中药质量标准化研究,还有利于揭示中药多层属性联系的原理。周密等<sup>[5]</sup>把数据挖掘用于中药缺失药性研究,采用决策树算法,得出“功能”-“药性”关联规则,对 20 味有功能与性味而无归经记录的中药,进行了归经预测。尚尔鑫等<sup>[6]</sup>采用 Apriori 算法和改进的数据库间关联规则算法,试图在中药的性味、归经及功效等属性间发现有意义的规则,为分析药对内部的组成关系提供数据支持。大量实验表明,利用数据挖掘技术可以揭示中药众多知识的隐含关系<sup>[7]</sup>。

针对药性在中药属性研究中的重要地位,本文以《中国药典》、《中药学》、《中药学》(教参)、《中华本草精选本》、《中药药理学》和《新编中药志》等书籍作为相关信息来源,以获取到的中药药性及其他属性的数据为基础,建立数据模型,进行挖掘分析,发现药性与主治、西医病名、药理作用、功能、药味、归经、化学成分等属性间的有意义的规则,为相关人

员提供更全面更可靠的信息平台,同时也为中药的现代研究和应用提供数据支持。

### 1 主要工作

关联规则挖掘是数据挖掘研究的一个重要分支,关联规则挖掘主要反映了事物之间的关联性,可发现大量数据中项集之间的关联或相关联系,这些关系是预先未知的或者被隐藏的<sup>[6]</sup>。

在进行数据挖掘之前,首先确定数据挖掘的目标,然后进行数据收集、数据预处理、数据分析等相关操作,挖掘出中药药性与主治、西医病名、药理作用、功能、药味、归经、化学成分等属性的关联关系。

**1.1 数据预处理** 数据预处理主要目的是消除或减少数据噪声和处理缺失数据的信息,尽量减少数据“不一致性”、“不完整性”等干扰因素影响,提高数据质量<sup>[8]</sup>。要有效利用数据挖掘技术进行中药属性间关系的知识发现,必须对相关中药数据进行必要的预处理,使其符合算法的要求。

针对中药的同药异名、同名异药,以及部分中药的功能、主治、药理作用术语也不统一的现象,首先进行数据准备、数据清洗工作,以《中国药典》、新版《中药药理学》教材、《中华本草精选本》为参照基础,结合临床用药经验,采集、归纳、规范、输入中药相关属性术语,将主治、药理作用和功效术语按照作用机制相同或相似等原则作规范,将相应术语词汇进行数据收集、数据清洗、数据集成及变换等预处理操作<sup>[9]</sup>,为构建中药数据信息库中的中药属性关系的探讨提供基础。

目前,在本中药信息库中,共收录了常见中药

717种,药性主要有“温”、“寒”、“平”、“微寒”、“凉”、“微温”等13个属性值;另外,对于中药主治属性,主要包括“肺燥咳嗽”、“肌痿无力”、“劳嗽”、“咯血”等内容,和中药对应项有2928条记录;对于西医病名属性中,主要包括“积滞”、“痢疾”、“疟疾”、“慢性气管炎”等内容,和中药对应项有3289条记录;对于药理作用,内容如“抗过敏”、“抗炎”、“影响免疫功能”、“杀虫”等,和中药对应项有3611条记录;对于功能,内容主要有“止血”、“滋阴”、“消积”、“化痰”等,和中药对应项有3023条记录;中药药味中,主要包括“辛”、“苦”、“甘”等内容,和中药对应项有1110条记录;归经中,主要包括“肺”、“肝”、“肾”、“脾”、“胃”、“心”等内容,和中药对应项有1610条记录;化学成分中,主要包括“氨基酸”、“明胶蛋白”、“葡萄糖醛酸”、“有机酸”等,和中药对应项有1840条记录。

**1.2 研究方法** 在数据集中,若大量记录在具有特征属性 $A$ 的同时,也频繁出现特征属性 $B$ ,则称特征属性 $A$ 和 $B$ 构成频繁模式,表示 $A$ 和 $B$ 之间的关联性,而这些模式可以用关联规则来观察和分析。关联规则的质量一般由规则的支持度(support)和置信度(confidence)两个参数来度量,它们分别反映所发现的规则的有效性和确定性。

规则 $X \Rightarrow Y$ 在数据库 $D$ 中的支持度是交易集中同时包含 $X, Y$ 的事务数与所有事务数之比,记为 $\text{support}(X \Rightarrow Y) = \text{support}(X \cup Y)$ ,可简化表示为 $P(A \cup B)$ 。支持度描述了 $X, Y$ 这两个项集在所有事务中同时出现的概率,即在所有事务集 $D$ 中 $A$ 和 $B$ 这两个项集同时出现的概率。

规则 $X \Rightarrow Y$ 在事务集中的置信度是指同时包含 $X, Y$ 的事务数与包含 $X$ 的事务数之比,它用来衡量关联规则的可信程度。记为 $\text{confidence}(X \Rightarrow Y) = \frac{\text{support}(X \cup Y)}{\text{support}(X)}$ ,可简化表示为 $P(B|A)$ ,即在事务集 $D$ 中出现项集 $A$ 的同时,也出现项集 $B$ 的概率。

Apriori算法是R. Agrawal等<sup>[10]</sup>于1994年提出,其核心是基于两阶段频繁集思想的递推算法。该关联规则在分类上属于单维、单层、布尔关联规则。该算法使用一种逐层搜索的迭代方法,并通过“连接”和“剪枝”在大量数据分析结果中寻找令人感兴趣的关联规则。但是,Apriori算法可能产生大量的候选集,以及可能需要重复扫描数据库。

本课题在Apriori算法的基础上进行了改进,以上述医学书籍作为相关信息来源构建中药信息数据

库,首先将收集整理好的数据库划分成 $n$ 部分,先是针对每个部分单独产生一组频繁项目集,最后再将这些项目集合并为一个全局的候选频繁项目集,然后以候选频繁项目集为基础,执行连接及剪枝处理等循环处理,直至产生强关联规则。

## 2 结果

本文以Microsoft SQL Server为后台服务软件,关联挖掘程序自行编写,运行于Matlab R2010a平台,对建立的数据库进行分析,挖掘中药药性与其他属性间的关联关系。

通过大量实验,建立了药性-药理作用、药性-主治、药性-西医病名、药性-功能、药性-药味、药性-归经、药性-化学成分等频繁项目集,并统计相关中药信息。

在收录的717种中药中,对于中药药性,“温”在中药中出现频次最高,如阿魏、艾叶、八角茴香等189种中药药性皆为“温”;依次为“寒”、“平”、“微寒”等,而“冷”出现频次最低,仅有田螺和蛤蜊肉2种中药。

各规则中,中药的属性药味、归经和化学成分,和药性共现频次较高,其中,药性药味的共现频次最高的前10位数据见表1。从表1实验结果看,对于中药药性与药味,药性“温”和药味“辛”联系最为密切,较为密切的频繁项集依次为寒-苦-平-甘-温-甘-温-苦-寒-甘-平-苦等内容。

表1 药性药味共现频率前10位关联规则

No.	药性	药味	支持度/次	支持率/%
1	温	辛	117	10.57
2	寒	苦	107	9.67
3	平	甘	87	7.86
4	温	甘	61	5.51
5	温	苦	60	5.42
6	寒	甘	60	5.42
7	平	苦	50	4.52
8	微寒	苦	42	3.79
9	平	辛	32	2.89
10	凉	甘	31	2.80

对于中药药性与归经,共现频次最高的前10位数据见表2。从表2可见,对于中药药性与归经,药性“平”和归经“肝”联系最为密切,较为密切的频繁项集依次为温-肝-温-脾-寒-肝-寒-肺-温-肾等内容。

对于中药药性和化学成分,关联规则药性“温”和化学成分“挥发油”最为密切,支持度为66,支持率为3.60%,较为密切的频繁项集依次为寒-有机酸-温-酯类-寒-生物碱-平-糖类-温-有机酸-平-有机酸等内容。药性化学成分的共现频次最高的前

表 2 药性归经共现频率前 10 位关联规则

No.	药性	归经	支持度/次	支持率/%
1	平	肝	86	5.35
2	温	肝	84	5.22
3	温	脾	83	5.16
4	寒	肝	80	4.98
5	寒	肺	74	4.60
6	温	肾	67	4.17
7	平	肺	61	3.79
8	温	胃	58	3.61
9	寒	胃	58	3.61
10	温	肺	57	3.54

10 位数据如表 3 所示。

表 3 药性化学成分共现频率前 10 位关联规则

No.	药性	化学成分	支持度	支持率/%
1	温	挥发油	66	3.60
2	寒	有机酸	48	2.62
3	温	酯类	47	2.57
4	寒	生物碱	41	2.24
5	平	糖类	40	2.18
6	温	有机酸	39	2.13
7	平	有机酸	39	2.13
8	温	糖类	37	2.02
9	平	黄酮类	36	1.97
10	温	无机物	35	1.91

同理,在实验结果中,对于中药药性及功能,关联规则药性“寒”和功能“清热”最为密切,支持度为 91,支持率为 3.04%,较为密切的依次为寒-解毒、温-止痛、寒-凉血、微寒-清热、寒-消肿、凉-清热等频繁项目集。对于中药药性及药理作用,关联规则药性“温”和药理作用“抗炎”最为密切,支持度为 57,支持率为 1.58%,较为密切的依次为寒-抗菌、温-抗菌、温-镇痛、寒-抗炎、温-影响心血管系统等内容。另外,经相关实验,中药药性与属性主治、西医病名与关联规则的支持率分别为 0.62%,0.49%。

### 3 小结

关联规则挖掘技术作为在重要数据中获取知识的有力工具,越来越多的应用于中医药领域的各个方面。然而中药自身的特点及多种不确定因素使

得数据挖掘技术在中药研究过程中有待进一步研究和探索。

本文以相关书籍、研究文献为基础,对获取的中药药性数据及中药其他属性数据进行预处理操作,搭建实验平台,应用关联规则挖掘方法分析中药药性及其他属性间关系数据的关系,数据挖掘结果可分析性较强,与中药药性传统认识相符,具有一定的现实意义。

### [参考文献]

[1] 高学敏. 中药学[M]. 北京:中国中医药出版社, 2007:69.

[2] 张冰,林志健,翟华强,等. 基于“三要素”假说研究中药药性的设想[J]. 中国中药杂志, 2008, 33(2):221.

[3] 梁茂新,范颖. 中药四性研究的现状、问题和对策[J]. 中华中医药杂志, 2008, 23(7):565.

[4] 胡志帅,曹慧,马金刚,等. 计算机科学在中药信息化建设中的应用进展[J]. 中国实验方剂学杂志, 2011, 17(13):263.

[5] 周密,王耘,乔延江. 利用数据挖掘方法预测中药缺失药性的初步研究[J]. 中国中医药信息杂志, 2008, 15(6):93.

[6] 尚尔鑫,叶亮,范欣生,等. 基于改进关联规则算法的中药药对药味间性味归经功效属性关系的发现研究[J]. 世界科学技术——中医药现代化, 2010, 12(3):377.

[7] 尚尔鑫,范欣生,段金殿,等. 基于三维图形化数据挖掘方法的四物汤类方配伍规律研究[J]. 中国实验方剂学杂志, 2011, 17(1):217.

[8] 秦莉花,王小云,陈晓阳. 基于关联规则绝经综合征睡眠问题的研究[J]. 中国实验方剂学杂志, 2011, 16(5):214.

[9] 金锐,林茜,张冰,等. 基于 Apriori 算法的中药气-味-效三维数据关联规则挖掘研究[J]. 中西医结合学报, 2011, 9(7):794.

[10] Jiawei Han, Mecheline Kamber. 数据挖掘理论与技术[M]. 2 版. 范明,孟小峰,译. 北京:机械工业出版社, 2010:146.

[责任编辑 邹晓翠]