

应用 Rosetta 对中医喘证用药经验规律的挖掘研究

马会霞¹, 路振宇², 刘保相¹, 赵斗贵², 王国权³, 于荣霞¹, 王巍¹, 包巨太^{1*}

(1. 河北联合大学, 河北 唐山 063000; 2. 河北省邯郸市传染病医院, 河北 邯郸 056002;
3. 唐山职业技术学院, 河北 唐山 063000)

[摘要] 目的: 在遵循中医以药测症、方证对应规律的基础上, 应用粗糙概念格聚类技术和方法挖掘喘证用药经验中的基础数据, 通过整理、分析和归纳药症之间的对应关系, 总结喘证的用药规律。方法: 利用 Rosetta 计算机软件对喘证用药经验进行数据挖掘和聚类解析。结果: 得到喘证各个症状与用药之间的关联规则和用药规律。结论: 利用粗糙概念格聚类技术和方法可以有效地挖掘出中医喘证临症经验中药味的用药规律, 为中医临床用药经验规律的数据挖掘研究提供了一个新模式。

[关键词] 喘证; 用药经验; 粗糙概念格; 数据挖掘; 聚类分析

[中图分类号] R287 [文献标识码] A [文章编号] 1005-9903(2012)14-0005-05

[网络出版地址] <http://www.cnki.net/kcms/detail/11.3495.R.20120515.1528.008.html>

[网络出版时间] 2012-05-15 15:28

Rosetta Application on Traditional Chinese Medicine Dyspnea Syndrome with Experience of Medication

MA Hui-xia¹, LU Zhen-yu^{2*}, LIU Bao-xiang¹, ZHAO Dou-gui²,

WANG Guo-quan¹, YU Rong-xia¹, WANG Wei¹, BAO Ju-tai^{1*}

(1. Hebei United University, Tangshan 063000, China;

2. Infectious Disease Hospital in Handan City, Handan 056002, China;

3. Tangshan Vocational & Technical College Tangshan 063000, China)

[Abstract] **Objective:** In the traditional Chinese medicine (TCM) with the theory of drug intervention and formulas corresponding to syndromes, and on the basis of application rough concept lattice clustering technique and method for mining out the basic data of medication experience, the medication rules on dyspnea syndrome are summarized through the sorting, summarizing and analyzing the corresponding relationship between medicine syndrome. **Method:** Rosetta computer software is used for data mining and cluster analysis on the experience of medication. **Result:** The association and medication rules between all symptoms and various drugs are obtained. **Conclusion:** Rough concept lattice clustering technique and method can effectively dig out the rules of medication and clinical experience in treating with dyspnea syndrome by TCM, providing a new model for data mining study in the rules of medication and clinical experience by TCM.

[Key words] dyspnea syndrome; medicine experience; rough concept lattice; data mining; clustering analysis

喘证是以呼吸困难, 甚至张口抬肩, 鼻翼煽动, 不能平卧为临床表现的一种病证^[1]。可见于多种

[收稿日期] 20120211(013)

[基金项目] 2009 年河北省自然科学基金项目 (F2009001276)

[第一作者] 马会霞, 讲师, 从事中医量化、信息化研究, Tel: 15031523933, E-mail: mahuixia107@163.com

[通讯作者] * 包巨太, 教授, 主任医师, 硕士生导师, 从事名老中医用药经验量化研究, Tel: 15133969965, E-mail: 549292966@qq.com

急慢性疾病过程中,如肺炎、喘息性支气管炎、肺气肿、肺结核、矽肺等以呼吸困难为主要表现时,均可参照辨证施治。历代医家对本病经过了长期实践,积累了丰富的临床经验,对于喘证的治疗在传统经方的基础上进行相应的药物配伍,多有良好的疗效,对后世医家的诊疗具有相当大的影响。

数据挖掘(data mining)是从大量的、不完全的、有噪声的、模糊的、随机的数据中提取隐含在其中的、人们事先不知道的、但又是潜在有用的信息和知识的过程^[2];粗糙集理论是一种较新的软计算方法,作为一个处理不确定性的数学工具相比之前的概率论、模糊集、证据理论等数据挖掘方法,其知识数据库的分类更加合理。概念格是一种有效的数据分析 and 知识发现的数学工具,基于概念格发现的知识模式有:关联规则、分类规则等,现在已获得的成果有基于概念格的关联规则挖掘算法;多概念格分类规则融合方法等^[3]。

中医医案中包含有大量模糊的、不确定的药症及方证对应信息,这是中医学经验知识发现过程中的特点,而粗糙集正是研究和解决不确定、不完备信息系统分类问题的理论。所以我们以名老中医喘证用药经验为研究对象,在传统中药药性与功效的研究基础上,将中医学经验知识视为具有多元模糊关系的不确定的信息系统,以中药属性和药症之间的对应关系为出发点,利用数学方法对中医喘证药、症的模糊对应关系进行关联规律挖掘,为下一步聚类得出喘证方证对应关系,进而建立中医喘证的粗糙集数学模型奠定了基础;其目的是构建一种应用粗糙概念格聚类技术研究中医临床经验的挖掘模式,为临床筛选有效方药提供指导,进一步拓宽传统喘证用药的临床应用范围。

1 研究方法

本研究选取以《中华医典》所收录的 75 部医案著作中所涉及的喘证医案为研究对象。通过选取和整理医案记录,应用 Excel 数据软件建立喘证症候与用药的数据库;然后应用粗糙概念格理论与方法对喘证临床经验的模糊关系进行规律挖掘,获得药、症之间的关联规则;利用软件的统计功能将各个关联规则按用药进行分类,对用药出现的次数和规则条目的多少进行统计,依次排列,为下一步建立中医喘证方证对应关联规则的数据库奠定基础。

2 构建以中药及其属性为研究对象的不确定信息系统

2.1 喘证信息数据来源 以《中华医典》所收录的

75 部医案著作中所涉及的喘证医案作为数据来源。《中华医典》是迄今为止规模最为宏大的中医类电子丛书,收录了民国以前中国历代医学古籍 600 余部,涵盖了至清末为止的中国医学文化建设的主要成就,具有权威性。以“喘”、“喘息”、“喘逆”、“咳嗽”、“上气”等主题词进行检索,共检出喘证相关医案 1 200 余例。

2.2 信息记录的整理 包括确定医案选择的标准,剔除记录过于简单的医案、删除重复的医案等;经过课题组全体成员的商议,最终确定能够用于研究的医案记录为 900 余条。然后按照喘证的证候诊断标准和八纲辨证原则确定每个医案的证候类型,治法、症状、处方和药物。名词术语规范化,包括对每一条记录进行整体的证候、治法分类;对各个单一类别的症状群进行标化;对药物进行标化,顺序按记录原貌处理;对方剂详细记录原方中的方名及药物加减,无方名的以君药加数学代码暂定名称以示区别。将整理好的各项记录输入 Excel 数据软件中备份,以备挖掘使用。

2.3 数据挖掘工具——Rosetta 及相关概念

2.3.1 Rosetta 是一款基于粗糙集理论框架的表格逻辑数据工具包,它提供了决策表补齐、决策表离散化等多种数据预处理功能,利用粗糙集中常见的约简和规则的获取算法,支持从数据预处理到预测和分析规则的全过程,是一款很直观的数据挖掘和知识获取图形操作软件^[4-5]。我们按照中医药基础理论的指导,根据 Rosetta 软件处理数据的特性,通过查找相关资料和专家咨询,将收集整理到的与喘证相关的数据进行了预处理,对医案各条记录中的症状设定为:喘主症,喘轻重,痰象,哮鸣,伴随症状寒热、汗出、口渴、大小便、饮食,其他伴随症状,舌象,脉象,病程,起病缓急,诊断病位,诊断病因等,并且按照第七版《中医诊断学》^[6]的标准对各项症状进行标化;对医案记录中的全部用药依照原貌顺序进行设定,并依照第七版《中药学》^[7]的药名进行标化整理。

2.3.2 不确定信息系统 不确定信息系统:模糊的、不能精确定义的一组信息类型,它们的成员没有精确定义的判别准则。

2.3.3 关联规则 信息处理过程中借助挖掘技术对海量数据进行挖掘和分析,发现数据中每个集合的支持度和置信度均满足其阈值,即每个事物的概率和条件概率均满足最小支持度阈值和最小置信度阈值,则认定事物之间有内在对应规律。

2.3.4 决策属性和条件属性的关系及应用 在信息融合与归纳学习研究中,利用数学模型和计算机工具进行分类时,通过某种等价关系或偏序关系,按照条件属性和决策属性进行归类,如果条件属性的某类包含于决策属性的某一类,则形成确定性的关联规则^[8]。

2.3.5 中药属性 中药的认识和使用是以中医理论为基础,具有独特的理论体系和应用形式,而其中性能(包括四气、五味、升降浮沉、归经、毒性)、功效和临床应用^[6]是每一味中药的侧重点,针对临床应用功效及喘证的临床用药特点将中药中纷繁复杂的性能、功效和临床应用汇聚到一起,构成数据挖掘中的属性板块。

2.3.6 构建以中药及其属性为研究对象的不确定信息系统 应用基于粗糙集理论框架的表格逻辑数据工具 Rosetta 进行数据挖掘。根据粗糙集理论处理模糊^[9-10]和不确定性知识的特性,在大量纷繁复杂、浩若烟海的喘证诊疗用药经验记录中,在保持数据原始分类不变的前提下,通过知识约简,找出药症之间的关联规则——药症对应关系。

2.3.7 构建药性决策表和方剂决策表^[9-12] 以中药为研究对象,以中药的属性为条件属性,以组方中药物的功效为决策属性,构建中药的不确定信息系统;以方剂为研究对象,以药物为条件属性,以证候为决策属性,构建方剂的不确定信息系统。

2.3.8 构建关于药和方的概念格^[13] 以药性的不确定关系,方、药的不确定关系,方、证的不确定关系,药、证的不确定关系构建相应的概率近似粗糙集。

2.3.9 挖掘关联规则 建立中医喘证医案数据库并完成分类知识的挖掘,确定决策属性与条件属性之间的关联规则。探索常用中医喘证临床用药经验规律,在此基础上建立粗糙概念格的数学模型。

3 处理过程

以喘证医案为研究对象,以医案各个症候为条件属性,以组方中药物的功效为决策属性,构建中药的不确定信息系统;以方剂为研究对象,以证候为条件属性,以药物为决策属性,构建方剂的不确定信息系统。利用 Microsoft Excel, Rosetta 数据处理平台,建立资料数据表,通过数据挖掘功能来处理喘证药症之间关联规则的相关信息。

因为我们的研究目的是借助数据挖掘技术对海量中医医案中的用药经验记录数据进行挖掘和分析,从而发现数据内已知的和未知的规律,即药症和方证之间的关联规则,故我们按照关联规则挖掘技

术的要求,对喘证医案记录数据的挖掘过程分为两个阶段:第一阶段先从大量资料集合中找出所有的高频项目组(frequent itemsets),即以医案中诊断病性(八纲辨证)为中介,假定记录中的所有症状描述或用药为条件属性,诊断病性为决定属性,得到以病性为枢纽的喘证症状和用药的高频集合;第二阶段再由这些高频项目组中产生关联规则(association rules)。在最小信赖度(minimum confidence)的条件门槛下,若一规则所求得信赖度满足最小信赖度,称此规则为关联规则。即在喘证症状和用药的高频项目集合中挑选出所有的药-症对应关系:以喘证各个单一症候群为条件属性,逐一对应以喘证用药群为决策属性进行初次规律挖掘,之后以喘证症候群组为条件属性,用药群组作为决策属性进行二次规律挖掘,然后将所得数据通过数据补齐、数据离散化、数据约简、上下近似集获取处理后,生成喘证的药-证关联规则,然后运用规则挑选的方法做成喘证药-证关联规则数据库。

4 挑选结果及统计分析

4.1 挑选结果 数据病症的规则抽取结果示意图,见图1。数据处方的规则抽取结果示意图,见图2。

图1 数据疾病的规则抽取结果示意图

图2 数据处方的规则抽取结果示意图

4.2 统计分析 喘证症候群组用药关联规则 17 084条;喘证痰象-用药关联规则 78 条;喘证轻中重-用药关联规则 62 条;喘证主症-用药关联规则 31 条;喘证哮喘有无-用药关联规则 42 条;喘证病程-用药关联规则 58 条;喘证伴随症状-用药关联规则 162 条;喘证其他伴随症状-用药关联规则 229 条;喘证起病缓急-用药关联规则 5 条;喘证舌苔象-用药

关联规则 149 条;喘证脉象-用药关联规则 176 条;
喘证病位-用药关联规则 104 条;喘证病因-用药关
联规则 168 条;喘证病性-用药关联规则 92 条;喘证
治法-用药关联规则 275 条。

5 喘证用药分析

历代医家大量的临床用药经验是中医学的精
髓,疗效确切的药物在诊疗过程中会被优先选择,高
效药物会有更多的使用机会。因此对主要药物的分
析,核心药物的提取,有助于帮助我们筛选疗效确切
的临床药物,可优先在高关联药物中进行,这也是本
文研究的目的之一。为了筛选疗效确切的临床药
物,利用 Rosetta 软件中的 statistics 功能将各个关联
规则按用药进行分类,依照用药出现次数和规则条
目的多少依次排列。喘证用药较广泛,此次挖掘出
的关联规则中共涉及 22 类 165 味常用中药,反映了
喘证证型的复杂性和临证处方的多变性。下面就其
用药规律按类别进行阐述。

5.1 化痰止咳平喘药 化痰止咳平喘药的药-证关
联规则多,用药比例大。其中前胡、瓜蒌、杏仁、旋覆
花、法半夏、竹茹、桔梗、白前、川贝母、葶苈子、枇杷
叶、苏子、桑白皮的关联规则少、症状集中单一而用
药比例最多,分类集中明显。提示治疗喘证的首选
基本药物是化痰止咳平喘药。痰饮既是病理产物,
又是继发病因,外感六淫之风、寒、热、湿、燥、火,内
伤之七情、饮食、劳逸均可损伤机体造成痰饮内停。
也就是说,大多数喘证患者均有不同程度的痰饮存
在,所以喘证的症状描述单一变化少,用药明显突
出,规则提取简单且分类集中。

5.2 补虚药 具有补虚作用的药-证关联规则多,
用药比例大,在收集整理的古代有关喘证的医案记
录中,无论实证还是虚证都用到大量具有补益作
用的中药,如北沙参、茯苓、生地黄、人参、白芍、鳖甲、
海蛤粉等,补益药在喘证核心用药中无论关联规则
还是用药比例都占有明显大的优势,表明很多医家
在治疗喘证时常配伍补虚药,且变证较多。总结原
因有三。

5.2.1 喘证多由机体正气不足,痰饮内停所致,正
如《证治准绳》所云:“肺虚则少气而喘。”说明肺
气虚是导致喘证发生的重要原因。又如《诸病源候
论·伤寒上气候》中指出:“肺主气,肺虚为邪热所
客,客则胀,胀则上气也。”喘证多因感受外邪而
发病,但病变的本质仍是正气不足。《时方妙用》更
进一步阐述:“肺虚力弱,不能下引……则喘作矣。”
由此可见,正虚是贯穿喘证发病全程的根本机制。

5.2.2 肺为娇脏,病邪传变较快,由于失治误治,极
易导致病情由表入里,由实转虚,或虚实夹杂,虚喘
病位主要在肺、肾(尤其是肾阳亏虚)两脏,故有关
附子的对应关联规则最多。

5.2.3 脾为化痰之源,因痰由脾肺气虚而生,脾运
强健则痰自消,无论实证、虚证多配以燥湿利水、健
脾化痰之药,如核心药物的茯苓。

5.3 解表药 解表药药-证关联规则以紫苏、荆芥、
桂枝、细辛、桑叶、浮萍等为代表,其中紫苏、桑叶规
则数目多用药比例大。而且发散风寒药无论规则数
目还是用药比例均明显高于发散风热药,提示喘证
在表实阶段以寒证为主。

紫苏性辛,温;归肺、脾经;长于发汗解表,行气
宽中。该药既能宣肺止咳平喘,又能醒脾宽中、行气
化痰,应用广泛,无论在表还是在里,风寒还是痰饮
均可加减配伍,所以有关紫苏的关联规则较多。而
桂枝、细辛、浮萍的对应规则只有一条,但涉及它的
用药数量很大,说明医家对桂枝、细辛、浮萍的用药
目的明确,应用单一。

5.4 理气药 理气药中与喘证治疗相关的用药以
陈皮、薤白、橘红、大腹皮为代表。配伍理气药主要
用于平降上逆气机。理气药的选择,常选用兼能化
痰之品,主要药物中的陈皮、青皮、化橘红、大腹皮
的使用,除取其理气之功外,也取其化痰之效。《素
问·至真要大论》指出“诸气愤郁,皆属于肺”。《诸
病源候论》进一步解释“肺主于气……气有余则喘
满逆上。”指出喘证发生的病机在于肺气上逆,配伍
理气药主要用于平降上逆气机。故喘证方剂常配用
理气药理气降逆以平喘。

5.5 清热药 清热药中与喘证密切相关的用药以
芦根、生地黄、玄参、竹叶、牛蒡子、连翘、射干、石膏、
青蒿、金银花、黄连为代表,其中以芦根、生地黄的用
药关联规则最多,芦根清热生津,归肺、胃经,能清透
肺胃气分实热,生津养胃,还可祛痰排脓。而生地
黄、玄参等清热凉血药均有滋阴的功效,配伍平喘药
可起到滋阴润肺、清热化痰的双重功效。

5.6 利水渗湿药 与喘证治疗相关的利水渗湿药
不多,生成的关联规则也有限,但茯苓、猪苓、冬瓜仁
的对应规则数目在所有用药中比较靠前,可见单味
茯苓、猪苓、冬瓜仁的使用意义重大。进一步揭示了
痰饮在喘证的发病中占有重要地位,喘证发生可以
形成痰饮,同时痰饮又可以成为喘证的发病病因,故
治疗宜配伍利水渗湿药。另外痰阻息道也是导致喘
证的主要病机。

5.7 温里药 以附子为代表,关联规则多,用药所占比例大,而干姜、肉桂、丁香、炮姜等用药规则比例较小。温里药性温而多燥,即可助阳又可化湿利水,温里药的使用符合痰饮产生的病理机制。正如《金匱要略》所言“短气有微饮者,当从小便以去之”。痰饮发生正虚是其本质,阳气虚则阴寒内生,寒邪闭肺,肺失宣降,精微不布,水道不通,水津内聚,终致痰饮郁肺。肺阳虚寒,寒痰阻肺是温里药使用的病理基础。然温里药易耗气伤阴,故使用温里药常配伍收涩之品,如配伍五味子,既敛肺止咳,又敛阴生津。

综上所述,喘证用药多集中于化痰止咳平喘药、补虚药、清热药、解表药、理气药,可以认为化痰止咳平喘法、补虚法、理气法、解表法是治疗喘证的常用方法,上述药物类别构成喘证用药基本框架。在喘证的药、症关联规则中某些中药在治疗喘证及其兼证上具有独特的优势,愈后疗效良好,进一步研究与开发这些药物对喘证的中医药治疗具有深远的意义。

6 讨论

探析喘证方证之间的关联规则,对于喘证的临床诊疗颇有现实意义。中医医案中包含有大量丰富的临床诊疗经验,将名老中医喘证临床用药经验在传统复方的基础上进行相应的药物配伍,可以为拟定高效优质的喘证中药复方提供依据。针对喘证方证之间关联规则的挖掘和梳理为中医临证诊断疾病和筛选有效方药提供有效的针对性指导,进一步开拓了传统方药的实用价值。我们从喘证的症状-用药和用药-症状两个方面挖掘探讨名老中医的用药经验,既继承了传统中医诊疗的精髓,又对当今中医临床的发展开辟了一条新的道路。

应用粗糙概念格理论与方法对中医喘证临床用药经验的模糊关系进行规律挖掘是本文的特色之一,目的是探索一种应用数据信息技术构建中药作用效果及强度体系的新途径,为中医临床用药经验规律的量化研究提供一个新模式。中医学的发展有其自身的特点和理论内涵,必须在保持中医特色的情况下,在中医理论体系框架内开展对中医临床用药经验的研究,这是一个中医学继承和发展的根本前提。粗糙集理论及其挖掘技术在传统的中国经验医学规律挖掘中具有很大的优势,二者之间的理论内涵具有很高的相似性,中医药症及方证之间的模糊性为粗糙集理论的不确定软计算方法提供了广阔

的研究空间,也为研究中医其他疾病证候开拓了一个新的路径。

应用粗糙概念格理论与方法能够充分挖掘出中医临床用药经验中药与症之间及方与证之间的对应关系,本研究的结果初步显示中医喘证的各个分型与相对应的方药之间呈现独特的“一对一”的数学相关联,这样从数学理论角度在数据挖掘的层面上初步验证了中医“以药测症”“方证对应”规律的客观存在。为下一步在“阴阳球-八纲三级结构模型”中的三维空间坐标系进行名老中医喘证用药经验——喘证方证对应规律的图形化、量化研究打下坚实的基础,也对利用数学三维空间理论解析传统的中医用药规律提供了理论依据。

[参考文献]

- [1] 周仲瑛. 中医内科学[M]. 北京:中国中医药出版社, 2007:2.
- [2] 屈景辉,廖琪梅. 医学信息数据库的建立与数据挖掘[J]. 第四军医大学学报, 2001, 22(1):88.
- [3] 史开泉,刘保相. S-粗集与动态信息处理[M]. 北京:冶金工业出版社, 2005:2.
- [4] 徐袭,刘玉波,范学鑫. 基于模糊工具箱和 Rosetta 的粗糙集数据挖掘[J]. 微计算机信息, 2007, 23(6/3):174.
- [5] 陈素玲,付爽,高云,等. 基于粗糙集理论的原发性高血压肝阳上亢证辨证系统的建立[J]. 山东中医药大学学报, 2011, 34(3):201.
- [6] 高学敏. 中药学[M]. 北京:中国中医药出版社, 2007:2.
- [7] 朱文峰. 中医诊断学[M]. 北京:中国中医药出版社, 2008:2.
- [8] 李伟光. 知识表达系统中条件属性与决策属性关系的分析[J]. 计量与测试技术, 2010, 37(1):35.
- [9] 杨纶标,高英仪. 模糊数学原理及应用[M]. 广州:华南理工大学出版社, 2006:07.
- [10] 王佑华,陆金根,柳涛,等. 中医医案中的知识发现研究[J]. 中西医结合学报, 2007, 5(4):368.
- [11] 沈逸雄,周怡,张睿玲. 心脏超声诊断数据挖掘尝试[J]. 医学信息, 2005, 18(1):19.
- [12] 胡玉荣. 粗糙集理论在心脏病诊断中的应用[J]. 电脑知识与技术, 2007, 2(7):164.
- [13] 丁卫平,顾春华,石振国,等. 基于形式概念分析的不完备电子病历系统粗糙挖掘研究[J]. 计算机科学, 2009, 36(10):230.

[责任编辑 邹晓翠]