

# 关联规则在中医药方剂挖掘中的应用<sup>\*</sup>

张 庆

轩 扬

马恒毅

(济宁医学院医学信息工程学院  
日照 276826)

(济宁医学院管理学院  
日照 276826)

(济宁市第一人民医院  
济宁 272011)

〔摘要〕 以肠痈方剂作为数据来源,运用关联规则分析、Apriori 算法得出方剂中最常见的单味药、对药及配伍规律,为传统中医论治时的合理组药、阐述并解释其浅显机制提供较为严谨的理论依据。

〔关键词〕 关联规则; Apriori 算法; 配伍规律; 肠痈方剂; 数据挖掘

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2018.11.015

**Application of the Association Rules in the Mining of Traditional Chinese Medicine Prescription** ZHANG Qing, School of Medical Information Engineering, Jining Medical University, Rizhao 276826, China; XUAN Yang, School of Management, Jining Medical University, Rizhao 276826, China; MA Hengyi, Jining NO. 1 People's Hospital, Jining 272011, China

〔Abstract〕 Taking the acute appendicitis prescription as the data source and making use of the association rule analysis and the Apriori algorithm, the paper reaches the conclusion of the most common single herbs, matched pair of medicine in prescriptions and compatibility laws, so as to provide relatively precise theoretical basis for rational drug pairs in the determine treatment of Traditional Chinese Medicine (TCM) as well as the expatiation and explanation of its plain mechanism.

〔Keywords〕 association rule; Apriori algorithm; compatibility law; acute appendicitis prescription; data mining

## 1 引言

数据挖掘是应用一系列技术从大型数据库中提取人们事先未知而潜在有用的隐含的信息和知识的过程。医学领域数据量庞大、更新速度快、结构复杂多样且蕴含着大量有价值的信息,利用数据挖掘技术处理这些信息可以取得事半功倍的效果<sup>[1]</sup>。常用的数据挖掘方法主要有关联规则、人工神经网络、决策树及聚类分析等<sup>[2]</sup>。在医学领域关联规则

常被用来发现数据间隐含的关联关系,如疾病与症状间的关联规则<sup>[3]</sup>、疾病并发症的关联研究<sup>[4]</sup>、用药和方剂配伍规则研究<sup>[5]</sup>、病症-辨证-处方关联研究<sup>[6]</sup>、药物间相互作用分析<sup>[7]</sup>、药品分类和剂型关联分析<sup>[8]</sup>、症状与药物间的关联关系<sup>[9]</sup>以及利用关联规则辅助检测、分析心理健康<sup>[10-11]</sup>等。本文旨在利用关联规则分析方法对肠痈方剂的配伍规律进行关联分析,揭示药材之间复杂特殊的关系,将隐藏的特点进行直观化呈现,以期揭示治疗肠痈的用药规律。

## 2 数据来源与方法

### 2.1 数据来源

本研究以《中医方剂大辞典》为依据,选取主

〔收稿日期〕 2018-07-13

〔作者简介〕 张庆,硕士,讲师,发表论文 7 篇,参编著作 3 部。

〔基金项目〕 济宁医学院校级科研计划项目(项目编号:0835/083501)。

治病症索引中治疗肠痛的方剂 101 首。由于中药材的俗称繁多杂乱,本文以《中药大辞典》为主要依据,辅以公认医学典籍作为参考,以其中记载的正名为标准对中药名称进行规范整理,如将“西黄”、“犀黄”、“犀牛黄”统一为“牛黄”,“苡米”、“薏仁”统一为“薏苡仁”等,避免同种药材因异名导致其出现频率降低从而影响研究结果的可靠性。对方剂中的中药进行排序处理,按照药材名称降序排列,整理后得到方剂数据集 P。

## 2.2 研究方法

2.2.1 关联规则及关联分析 (1) 关联规则。指在同一个事件中出现的不同项的相关性,是形如  $X \rightarrow Y$  的蕴含表达式,其中 X 和 Y 是不相交的项集,即  $X \cap Y = \emptyset$ 。关联规则的强度可以用支持度和置信度度量。支持度确定规则可以用于给定数据集的频繁程度,而置信度确定 Y 在包含 X 的事务中出现的频繁程度<sup>[12]</sup>。通过设定支持度和置信度的阈值可以获取联系密切的药材间的关联关系。(2) 关联分析。用于发现隐藏在大型数据集中的联系,所发现的联系可以用关联规则或者频繁项集的形式表示。关联规则挖掘就是从大量的数据中挖掘出描述数据项之间相互联系的有价值的有关知识。令  $P = \{p_1, p_2, \dots, p_n\}$  是方剂数据集中所有项的集合,而  $T = \{t_1, t_2, \dots, t_n\}$  是所有方剂的集合,即  $n = 84$ 。每个方剂  $t_i$  包含的项集都是 P 的子集,如 3 - 项集 {黄芪, 当归, 甘草}。

2.2.2 Apriori 算法 挖掘频繁项集的基本算法,最早由 Agrawal 等人提出<sup>[13]</sup>。该算法利用层次顺序搜索的循环方法来完成频繁项集的挖掘工作。即利用  $k$  - 项集来产生  $(k + 1)$  项集。具体做法是首先找出频繁 1 - 项集,即为  $L_1$ ; 然后利用  $L_1$  来挖掘  $L_2$ , 即频繁 2 - 项集,不断如此循环下去直到无法发现更多的频繁  $k$  - 项集为止<sup>[14]</sup>。

2.2.3 研究工具 IBM SPSS Modeler 是一款功能丰富的数据挖掘软件,利用上述的关联规则与关联分析以及 Apriori 算法对方剂数据集 P 进行关联规则挖掘分析。具体操作过程如下:运用 Excel 对

方剂数据集 P 中的项进行统计分析,得出药物频数表。将该表导入 IBM SPSS Statistics 中,利用数据重组技术工具将原始数据的格式从变量到个案转化成为数据集,对于当前数据中的每个变量组,已重组的文件中都会有 1 个目标变量,最终形成交易数据库。将交易数据导入 Modeler 工作界面,从外源部分读取数据的源节点、连接类型,过滤无需进行分析的字段,如方剂出处、朝代、编号及症型种类等,保留所需字段,按照 ID 汇总,设置真值为 T,否则为 F,数据流流动至“设为标志”字段即可成功转换为布尔型二项 T - F 数据集。将源文件用 Apriori 算法进行分析,设定支持度与置信度阈值,得出频繁项集。即根据使用频率较高的对药、药组得出治疗肠痛方剂的核心药物。

## 3 结果

### 3.1 单味药数据

运用 Excel 对单味药的频次进行统计分析可得到治疗肠痛的方剂中各味药材的使用频率,由高到低排列出现频次  $\geq 10$  的中药,见表 1。当归入甘、辛之味,药用性质温和平补,在润肠通便方面极有助益,中医方剂中多数用于治疗血虚气羸、肠燥干燥、便秘不畅,同时可以增强肠胃的吸收消化功能,能对肠痛发病时人体系统运化失职的症状进行改善,促进身体新陈代谢循环的流程,将体内废物排出体外。甘草则多用来补脾益气,对于脾胃孱弱者有补中固本的疗效,因其化学激素作用见效迅速,与芍药配伍应用对于痢症毒急发者可以显著增强清热解毒、缓急止痛的疗效,缓解药物毒性、烈性从而对诸药进行调和,对于已呈现出发病趋势的患者可以减轻强烈刺激,也可单独取用,内服或外敷。肠痛发病多因血腐肉败,积淤之物停滞在肠内不能化散,从而引发细菌性感染,大黄中具有的有效成分可杀灭多类病原微生物,如革兰氏阴性菌,从而可对抗多种感染症状,消解炎症,恢复人体免疫防线、活血行瘀、苦寒解热、涤荡痞积,对于肠部肿痛也有较为良好的缓解作用与治疗效果。

表 1 治疗肠痛的 101 首处方中最常见的单味药（频次≥10）

| 中药名称 | 频次 | 百分比（%） | 中药名称 | 频次 | 百分比（%） |
|------|----|--------|------|----|--------|
| 当归   | 32 | 38.10  | 没药   | 12 | 14.29  |
| 甘草   | 30 | 35.71  | 川芎   | 10 | 11.90  |
| 大黄   | 25 | 29.76  | 肉桂   | 10 | 11.90  |
| 地黄   | 12 | 14.29  | -    | -  | -      |

3.2 常见对药

两味中药材组成的集合规则称为对药，根据出现频次由高到低排列，见表 2。

表 2 治疗肠痛的 101 首处方中最常见的对药（频次≥9）

| 对药配伍    | 频次 | 对药配伍    | 频次 |
|---------|----|---------|----|
| 当归 - 甘草 | 16 | 丹皮 - 大黄 | 11 |
| 白芷 - 当归 | 13 | 人参 - 黄芪 | 10 |
| 桃仁 - 丹皮 | 12 | 川芎 - 当归 | 9  |
| 没药 - 乳香 | 11 | 茯苓 - 黄芪 | 9  |
| 人参 - 甘草 | 11 | 白芷 - 甘草 | 9  |
| 黄芪 - 甘草 | 11 | 桃仁 - 大黄 | 9  |
| 黄芪 - 当归 | 11 | -       | -  |

3.3 配伍规律

在 Modeler 中设定支持度大于 10%，最小规则置信度为 50%，运用 Apriori 算法进行分析，得出治疗肠痛的药物配伍规律，见表 3。“白芷 - 当归”这一对药的置信度为 1，也就是说当白芷被选入方剂使用时当归也必定同时被选用，白芷主要能缓解腰腹蚀脓之痛，治息肉及补血痢、肠风、痔瘕排脓等疮痍之症，二者皆具活血通窍之效，可治疗肠阻痢症时积血滞怠的症状。置信度排在第 2 位的“没药 - 乳香”，也显示出较为密切的配伍关系。

表 3 基于关联规则的治疗肠痛药物配伍规律的研究

| 序号 | 关联规则    | 置信度     |
|----|---------|---------|
| 1  | 白芷 - 当归 | 1       |
| 2  | 没药 - 乳香 | 0.916 7 |
| 3  | 川芎 - 当归 | 0.9     |
| 4  | 人参 - 甘草 | 0.733 3 |

续表 3

|    |               |         |
|----|---------------|---------|
| 5  | 黄芪、当归 - 甘草    | 0.909 1 |
| 6  | 茯苓、黄芪 - 人参    | 0.888 9 |
| 7  | 白芷、甘草 - 黄芪    | 0.777 8 |
| 8  | 川芎、当归 - 黄芪    | 0.777 8 |
| 9  | 川芎、当归 - 甘草    | 0.777 8 |
| 10 | 桃仁、大黄 - 丹皮    | 0.777 8 |
| 11 | 白芷、甘草、当归 - 黄芪 | 0.777 8 |

注：关联规则含义为“-”左侧药物出现时“-”右侧可能会出现 的药物；置信度含义为“-”左侧药物出现时“-”右侧药物出现的 概率

4 结语

随着数据挖掘技术的日益发展，越来越多的学者将其应用于中医药各大领域的研究中，运用相关分析方法挖掘其中隐含的信息，为临床诊断用药提供依据，促进中医药学的不断发展。方剂是在中医药性理论和病机理论指导下的数味中药的有机组合<sup>[15]</sup>，蕴含着中医药疗病祛疾的精髓。关联规则作为数据挖掘中的一种技术方法，主要用于探索事物之间的关联关系。肠痛是指发生于肠道的痈肿，是一种常见的 外科急腹症<sup>[16]</sup>。本文通过对肠痛方剂运用数据挖掘的方法进行关联规则分析，得出使用频率较高的单味药，从药性来看，多选用平性、微带寒性的药材。既有平性药材如当归、黄芪等发挥温补作用，又有药性与桃仁相近者可借寒驱热，另外具有补中益气功效的药材如人参也选入少数，说明古代医家认为该病与脾虚血滞、久热不散关系密切。通过关联分析得到肠痛方剂的核心组方，如白芷 - 当归、黄芪 - 当归 - 甘草、黄芪 - 甘草 - 当归 - 人参等，从配伍规律来看，治疗肠痛的方剂多从活血化瘀、补脾养气以促进血液循环的思路出发对

药材进行选择组合与配伍。本文通过对肠痈方剂运用数据挖掘的方法进行关联规则分析揭示肠痈方剂的配伍规律,为传统中医论治时的合理组药解释其浅显机制提供较为严谨的理论依据。

## 参考文献

- 1 肖辉,周征奇,肖革新,等. 公共卫生领域中的数据挖掘 [J]. 医学信息学杂志, 2013, 34 (12): 1-5.
- 2 沈培. 基于数据挖掘的甲肝医疗费用影响因素与控制策略研究 [D]. 武汉: 华中科技大学, 2012.
- 3 王华,胡学钢. 基于关联规则的数据挖掘在临床上的应用 [J]. 安徽大学学报 (自科版), 2006, 30 (2): 21-25.
- 4 余辉,张力新,刘文耀,等. 医学数据挖掘系统研究——糖尿病并发症流行病学知识发现 [J]. 计算机工程与应用, 2006, 42 (18): 229-232.
- 5 郭位先,吴嘉瑞,张冰,等. 基于关联规则和复杂系统熵聚类的颜正华教授治疗血瘀证用药规律研究 [J]. 中华中医药杂志, 2013, 20 (11): 2274-2277.
- 6 欧凤霞,王宗殿. 基于关联规则的数据挖掘技术在中医诊断中的应用 [J]. 河南工程学院学报 (自然科学版), 2011, 23 (2): 53-58.
- 7 胡双. 关联规则挖掘研究及其在药物相互作用中的应用 [D]. 昆明: 昆明理工大学, 2013.
- 8 郑银丽,相秉仁,赵国明. 关联规则技术在医药零售业

药品营销组合中的应用 [J]. 医学信息学杂志, 2011, 32 (4): 55-58.

- 9 李艳红,沈瑞琪,欧敬民. 基于文本挖掘技术的高血压用药规律研究 [J]. 微型机与应用, 2017, 36 (3): 103-106.
- 10 亓文娟,晏杰,黄书城,等. 关联规则挖掘在大学生心理健康测评系统中的应用研究 [J]. 湖南工业大学学报, 2013, 27 (6): 94-99.
- 11 Zhang LF, Zhang DN, Wang ZP. Application of Association Rule in Mental Health Test for Employees in a Petrochemical Enterprise [J]. Zhonghua Lao Dong Wei Sheng Zhi Ye Bing Za Zhi, 2017, 35 (10): 748-751.
- 12 崔雷. 医学数据挖掘 [M]. 北京: 高等教育出版社, 2006.
- 13 张红艳,都娟. 关联规则中 Apriori 算法的应用 [J]. 数字技术与应用, 2011 (8): 14-15.
- 14 王培吉,赵玉琳,吕剑峰. 基于 Apriori 算法的关联规则数据挖掘研究 [J]. 统计与决策, 2011, 11 (23): 19-21.
- 15 张真全. 论方剂与病机的相关性 [J]. 辽宁中医药大学学报, 2009 (6): 11-12.
- 16 张董晓,张燕生,燕美荣. 肠痈中医治疗进展 [J]. 辽宁中医药大学学报, 2012, 14 (2): 62-64.

## 2019 年《医学信息学杂志》征订启事

《医学信息学杂志》是国内医学信息领域创刊最早的医学信息学方面的国家级期刊。主管: 国家卫生和计划生育委员会; 主办: 中国医学科学院; 承办: 中国医学科学院医学信息研究所。中国科技核心期刊 (中国科技论文统计源期刊), RCCSE 中国核心学术期刊 (武汉大学中国科学评价研究中心, Research Center for Chinese Science Evaluation), 美国《化学文摘》、《乌利希期刊指南》及 WHO 西太区医学索引 (WPRIM) 收录, 并收录于国内 3 大数据库。主要栏目: 专论, 医学信息技术, 医学信息研究, 医学信息组织与利用, 医学信息教育, 动态等。读者对象: 医学信息领域专家学者、管理者、实践者, 高等院校相关专业的师生及广大医教研人员。

2019 年《医学信息学杂志》国内外公开发行, 每册定价: 15 元 (月刊), 全年 180 元。邮发代号: 2-664, 全国各地邮局均可订阅。也可到编辑部订购: 北京市朝阳区雅宝路 3 号 (100020) 医科院信息所《医学信息学杂志》编辑部; 电话: 010-52328673, 52328672, 52328686, 52328687, 52328670。

《医学信息学杂志》编辑部