

· 数据挖掘 ·

关联度最强药物配伍的中医止呕类方数据挖掘

黄颖琦, 贾恒, 何前松, 冯泳*
(贵阳中医学院, 贵阳 550025)

[摘要] **目的:** 在古今中医文献中寻找止呕方剂配伍规律与用药特点, 为中药止呕新药的开发提供理论支持。**方法:** 收录古今止呕类方剂 985 首建立止呕类方剂数据库, 运用相关置信度规则, 对中医止呕方剂药物配伍的数据进行挖掘, 利用剪枝方法筛选关联度最强的数据。**结果:** 最常用的单味药物为生姜使用频率高达 61.23%。关联性最强的核心药对是茯苓配伍姜半夏, 其相关置信度为 0.114 4。关联度最强的药组为生姜、姜半夏、茯苓。**结论:** 生姜、姜半夏、茯苓, 其相关置信度为 0.295 4 是中医止呕方剂中最常合用的药物配伍, 其 3 种药物间存在极强的关联性, 张仲景创制的小半夏加茯苓汤被证实是中医止呕的核心药组。

[关键词] 止呕类方; 配伍规律; 数据挖掘; 小半夏茯苓汤; 中医知识发现; 药对

[中图分类号] R287 **[文献标识码]** A **[文章编号]** 1005-9903(2012)14-0001-04

[网络出版地址] <http://www.cnki.net/kcms/detail/11.3495.R.20120515.1527.007.html>

[网络出版时间] 2012-05-15 15:27

Rule of Remedy for Vomiting in Science of Traditional Chinese Medicine Formu as by Data Mining Based on Both Association and Correlation Rule

HUANG Ying-qi, JIA Heng, HE Qian-song, FENG Yong*
(Guiyang College of Traditional Chinese Medicine, Guiyang 550025, China)

[Abstract] **Objective:** To look for the rules of compatibility of medicines in historical prescriptions which treat vomiting, mining out new knowledge in science of traditional Chinese medicine (TCM) formulas, providing support for making new TCM remedy for vomiting. **Method:** Created a TCM formulas database to treat vomiting by collecting 985 TCM formulas. A threshold level datamining method which based on both association and correlation rule is used, to mining rule of compatibility of medicines in database of historical prescriptions for vomiting. **Result:** The most used drug is *Zingiber officinale*, the using frequency of *Z. officinale* is 61.23%. The most association and correlation couple of drugs are *Poria cocos* (Schw.) Wolf and *Pinellia ternate* (Thunb.) Breit. the correlation - confidence of this couple is 0.114 4. The most association and correlation group of drugs are *Z. officinale* and *P. cocos* and *P. ternate*, the correlation-confidence of this group is 0.295 4, **Conclusion:** The group of drugs which contain *Z. officinale*. and *P. ternate* and *P. cocos* is most used to treat vomiting. The ternate added in the *P. cocos* soup, which is ancient famous TCM formulas created by ancient great doctor ZHANG Zhong-jing, is proved that it is the key compatibility of drugs to treat vomiting.

[Key words] prescriptions for vomiting; key compatibility of drugs; data mining; the pinellia add tuckahoe formulas; KDTCM; couple drug

[收稿日期] 20120320(003)

[基金项目] 国家自然科学基金项目(30960472/C190105);贵州省科学技术基金项目(黔科合 J 字[2011]2293);贵阳中医学院科研项目(贵中医院内[2011]25号)

[第一作者] 黄颖琦, 硕士, 讲师, 从事数据挖掘及复杂系统信息技术研究, E-mail: hyqly-xly@sohu.com

[通讯作者] * 冯泳, 教授, 从事中药复方配伍研究, E-mail: fy668@sina.com

呕吐是临床常见症、多发病,古今中医文献资料记载了许多临床行之有效的止呕方剂。利用当今数据挖掘技术,在浩如烟海的中医止呕方剂文献和当今临床报道资料中寻找止呕方剂配伍规律与用药特点,可以为中药止呕新药的开发提供有力的理论支持。现有中医方剂数据挖掘大多基于传统的数据挖掘方法如分类、聚类、关联规则等^[1]。近年来方剂配伍规律研究主要是利用关联规则的支持度和置信度,对药对、药组进行数据挖掘研究,难于从大量数据挖掘结果中选取关联度最强的数据,并且方剂配伍的数据挖掘需要在大量的方剂数据基础上进行^[2]。如何在中医理论指导下,结合临床用药实践经验,从大量的挖掘结果中选择出最有价值的数,是基于支持度和置信度的关联规则由于自身缺陷所导致的难点所在。利用支持度-置信度框架的传统关联规则挖掘,其中有些挖掘结果不但没有实际意义,反而会使判断发生误差^[3]。为此,本课题组联合多学科研究人员反复实践,在关联且相关规则^[4]理论上,提出基于相关置信度的筛选方法,用于中医止呕药对、药组的挖掘研究,探讨中医止呕方剂的用药特征和配伍规律。

1 方法

1.1 方剂来源 方剂数据库共收录古今止呕类方剂 985 首,其中 768 首方剂录自《中医方剂大辞典》,方源包括《金匱要略》、《诸病源候论》、《备急千金要方》、《外台秘要》、《普济本事方》、《三因极病证方论》、《丹台玉案》、《寿世保元》、《普济方》、《片玉心书》、《证治汇补》、《惠直堂方》、《医学心悟》、《医方类聚》、《景岳全书》、《幼幼新书》等;今方选自成药处方集、公开发表的临床疗效报道或者医案,如《全国中药成药处方集》、《中国中医急症》、《国医论坛》、《中医药临床杂志》、《中国医药学报》、《实用中医内科杂志》、《中国中医药信息杂志》等。另收录古今呕吐类方剂 217 首,方源为《普济方》、《仁斋直指》、《圣济总录》、《太平圣惠方》、《丹溪心法》、《饲鹤亭集方》、《儒门事亲》、《景岳全书》、《医学入门》、《济阴纲目》、《叶天士女科证治秘方》、《奇效良方》、《古今医鉴》、《御药院方》等。

1.2 数据收录及标准化处理 分别按痰饮内阻、肝气犯胃、饮食停滞、脾胃气虚、脾胃阳虚、胃阴不足、外邪犯胃、小儿呕吐、妊娠呕吐、产后呕吐、呕吐类 11 个证型进行呕吐方剂的收录。其中在方剂古文文献记载中,对呕吐证型未做具体描述的,仅指明“治疗呕吐”而未进行具体辨证的古方剂,或者古方记

载为一方治疗多种病症,其中包含“治疗呕吐”的,数据库中笼统归为“呕吐”类,单独列为“呕吐”类证型。药名均进行标准化处理,如“炙甘草”和“甘草”统一称为甘草;“茯苓”和“茯神”统一称为“茯苓”;“陈皮”和“橘皮”统一称为“陈皮”。

1.3 数据挖掘方法

1.3.1 频数分析 数据的频数分析是一种描述性统计分析。本研究通过频数分析数据库中单味药物使用情况,其中中药频数是中药在方剂数据库中总共出现的次数,反映了该味药物在数据库中的使用频率。^[5]

1.3.2 关联分析 通过关联分析挖掘处方中药对之间的配伍关系,即发现处方中一种中药与另一种中药的关联关系强弱。数据挖掘领域中置信度 confidence(A→B),表示 A 条件下 B 出现的条件概率 P(A|B),支持度 support(AB)反映 AB 同时出现的概率大小。统计止呕类方常用药物,用药物同时出现的概率表示,即用支持度说明^[6]。止呕类方使用药物 A 条件下使用药物 B 的概率,用置信度表示。

1.3.3 关联相关频繁模式的剪枝方法

1.3.3.1 关联且正相关规则 选用该规则反映药对之间结合配伍的紧密相关程度。

设模式 X 和 Y 都是频繁模式,如果模式 XUY 的支持度 (support) 大于等于预先给定的最小支持度,对预先给定最小相关置信度 r,条件概率 P(Y/X) 和 P(X/Y) 都大于预先给定的最小置信度,且以下式子成立:

$$P(XY) = [P(XY) - P(X)P(Y)] / [P(XY) + P(X)P(Y)] \geq r$$

则称规则 X↔Y 为关联且正相关规则。

1.3.3.2 正相关模式相关性度量 c-conf(X):若模式 X 中任意子模式的相关置信度大于 0,则其为正相关模式。模式 X 中所有两两子集的相关置信度的平均值,即为此正相关模式相关置信度 c-conf(X)。

1.3.3.3 剪枝挖掘方法 本文提出剪枝关联相关频繁模式的挖掘方法,先挖掘频繁二次项和频繁项集组的频数和支持度,选出所有频数大于具有最高频数的三次项集的二次项,并按照公式计算二次项的相关置信度。在关联正相关的候选多次项表中,选取大于单倍数据种类数,小于两倍数据种类书的候选多次项集,在候选多次项集中删除含有负相关置信度的二次项的项组,最后得到的药组即为关联相关频繁药组,即核心药组配伍。

1.4 程序设计方案 本研究前台采用 ACCESS 数据库,后台用 VB,设计 Apriori 算法程序,挖掘 1,2,次频繁项目集合。利用 1.3.3.3 所述剪枝方法,计算候选药对相关置信度,在 3 项候选药组中进行剪枝筛选,从具有强关联性的药对药组中,最后分析出具有真正研究价值的核心药对及药组^[7]。

2 挖掘过程及挖掘结果

挖掘频数大于等于 4 的频繁项,共挖掘到数据项 10 553 条,其中两项组成的药对为 1 191 条,3 项组成的药组为 4 007 条。

2.1 单味药物频数挖掘分析 挖掘止呕类方剂数据库中共享药物 203 味,其中使用频率最高的前 10 味药物及其所占比例见表 1 所示。

表 1 单味药物频数统计

排序	频数	中药	所占比例/%
1	616	生姜	61.23
2	575	陈皮	57.16
3	543	甘草	53.98
4	483	姜半夏	48.01
5	444	人参	44.14
6	410	茯苓	40.76
7	309	白术	30.72
8	214	丁香	21.27
9	188	厚朴	18.69
10	179	干姜	17.79

2.2 药对相关置信度分析 具有最大频数的 3 次项模式为 X = {陈皮;生姜;甘草},对应频数为 223。从 1 191 条药对中筛选所有频数大于 223 的二次频繁项,并计算其相关置信度,药对的相关置信度见下表 2。

表 2 药对配伍分析

频数	药对 /A	药对 /B	置信度 A->B	置信度 B->A	支持度	相关置信度 c-conf
253	茯苓	姜半夏	0.617 1	0.523 8	0.251 5	0.114 4
232	人参	茯苓	0.533 6	0.577 9	0.235 5	0.113 2
274	茯苓	甘草	0.668 3	0.504 6	0.272 4	0.096 0
357	生姜	姜半夏	0.579 5	0.784 1	0.354 9	0.083 4
275	陈皮	茯苓	0.478 3	0.670 7	0.273 4	0.069 3
277	人参	甘草	0.623 9	0.623 9	0.275 3	0.061 8
347	陈皮	甘草	0.603 5	0.639	0.344 9	0.045 2
297	陈皮	姜半夏	0.516 5	0.614 9	0.295 2	0.026 0
263	生姜	茯苓	0.426 9	0.641 5	0.261 4	0.012 7
263	人参	陈皮	0.592 3	0.457 4	0.261 4	0.007 3
359	陈皮	生姜	0.624 3	0.582 8	0.356 9	-0.000 8
275	人参	生姜	0.446 4	0.619 4	0.273 4	-0.004 8
334	生姜	甘草	0.542 2	0.615 1	0.332	-0.008 3
260	姜半夏	甘草	0.478 8	0.538 3	0.258 4	-0.011 9

2.3 3 项药组筛选 从 4 007 条 3 味药配伍药组中,筛选出 22 项频数最高的记录(因为总共有 11 个类别的数据),如表 3 所示。

表 3 3 味药组配伍支持度统计

ID	频数	中药项 1	中药项 2	中药项 3	支持度
1.00	223.00	陈皮	生姜	甘草	0.23
2.00	219.00	陈皮	生姜	姜半夏	0.22
3.00	194.00	陈皮	茯苓	甘草	0.20
4.00	191.00	陈皮	茯苓	姜半夏	0.19
5.00	191.00	生姜	姜半夏	甘草	0.19
6.00	189.00	陈皮	姜半夏	甘草	0.19
7.00	185.00	生姜	茯苓	姜半夏	0.19
8.00	184.00	陈皮	生姜	茯苓	0.19
9.00	178.00	人参	陈皮	甘草	0.18
10.00	175.00	人参	陈皮	生姜	0.18
11.00	174.00	生姜	茯苓	甘草	0.18
12.00	171.00	人参	生姜	甘草	0.17
13.00	170.00	茯苓	姜半夏	甘草	0.17
14.00	164.00	人参	茯苓	甘草	0.17
15.00	152.00	人参	生姜	茯苓	0.15
16.00	152.00	人参	生姜	姜半夏	0.15
17.00	151.00	人参	白术	甘草	0.15
18.00	145.00	人参	陈皮	茯苓	0.15
19.00	145.00	白术	陈皮	甘草	0.15
20.00	143.00	人参	白术	陈皮	0.15
21.00	140.00	白术	茯苓	甘草	0.14
22.00	137.00	白术	陈皮	生姜	0.14

2.4 剪枝筛选 由于甘草在方剂中多作为辅助性用药^[8],所以核心药组中去掉含有甘草的项,最后剩核心药组两组:姜半夏、生姜、茯苓和姜半夏、陈皮、茯苓。见表 4,5。

表 4 剪枝后的 3 味药组

ID	频数	中药项 1	中药项 2	中药项 3	相关置信度
3	194	陈皮	茯苓	甘草	0.303 2
4	191	陈皮	茯苓	姜半夏	0.279 2
7	185	生姜	茯苓	姜半夏	0.295 4
8	178	人参	陈皮	甘草	0.300 2
14	164	人参	茯苓	甘草	0.265 0
17	151	人参	白术	甘草	0.236 5
18	145	人参	陈皮	茯苓	0.260 6
19	145	白术	陈皮	甘草	0.258 9
20	143	人参	白术	陈皮	0.230 5
21	140	白术	茯苓	甘草	0.230 1

表 5 关联度最强的高兴趣度模式

ID	频数	中药项 1	中药项 2	中药项 3	相关置信度
7	185	姜半夏	生姜	茯苓	0.295 4
4	191	姜半夏	陈皮	茯苓	0.279 2

3 结果与讨论

3.1 单味药物频数分析 从表 1 得出,单味药物频数统计中,呕吐类方剂生姜使用占总方剂 61.23%,是使用频率最高的药。生姜辛温,主入脾胃肺经,温中止呕,《药性论》中谓生姜:“疗时疾,止呕吐不下食”,《备急千金要方·呕吐秽逆》云:“凡呕者,多食生姜,此是呕家圣药”^[9]。数据挖掘结果显示生姜使用频率最高,与临床实践完全吻合,表明生姜是止呕类方剂首选的核心用药。其次是陈皮使用频率占总方剂 57.16%,名列第 2 位;姜半夏在呕吐方剂数据库中使用时名列第 4,高达 48.01%,故陈皮、姜半夏亦为呕吐类方剂常用药。甘草使用频率虽为 53.98%,名列第 3 位,但由于甘草一般均作为调和诸药的常用药配伍成方,方剂中多作为辅助性用药^[8],故其对止呕方剂的配伍意义不大。

3.2 核心药对配伍分析 从表 2 得出,具有关联性的药对中,相关置信度最大的药对为茯苓配伍姜半夏。《中药学》谓姜半夏:“姜半夏味苦降逆和胃,为止呕要药。各种原因的呕吐,皆可随证配伍用之”^[9]故姜半夏配伍茯苓,本研究表明其在止呕类方剂中,为相关性最强的药对,止呕类方中使用姜半夏即多配伍茯苓,使用茯苓即多配伍姜半夏,两者在止呕功效上起相辅相成的作用。

相关性名列第 4 的药对为姜半夏配生姜,置信度 Confidence(姜半夏→生姜) = 78.41%,不仅置信度为止呕方剂中最高,并且是支持度仅次于陈皮配生姜配伍的药对。当止呕方剂中用到姜半夏时,随症配伍多用生姜,《中药学》谓姜半夏“对痰饮或胃寒所致的胃气上逆呕吐尤宜,常配生姜通用”^[9]。生姜辛温为温散之品,能协助姜半夏以加强其燥湿化饮,降逆止呕之功,此外,生姜有解毒之功,能制姜半夏之悍,以解姜半夏之毒,此为相杀、相畏的配伍关系,故既为臣药又兼佐药之用^[9]。

故止呕类方剂中,药对茯苓配姜半夏、生姜配姜半夏为两组常用核心药对,具有紧密关联和相关性。

3.3 核心药组配伍分析 从表 5 得知,3 味药的核心药组为两组:姜半夏-生姜-茯苓和姜半夏-陈皮-茯苓。而其中 3 味药物配伍姜半夏、生姜、茯苓相关置信度为最高,故姜半夏、生姜、茯苓为呕吐类方剂关联度最强的主要配伍用药。

《金匱要略·痰饮病脉证并治十二》记载以生

姜、姜半夏组成的小半夏汤“治支饮呕吐不渴”,而以生姜、姜半夏、茯苓为主方的小半夏茯苓汤“治卒呕吐”^[10],可见数据挖掘知识发现与中医古代临床实践的记录不谋而合,医圣张仲景给后世留下的治疗呕吐的祖方、主方^[11],其经典意义已被现代数据挖掘技术所证明,大量的数据分析揭示中医止呕方剂配伍的核心药物即为生姜、姜半夏和茯苓。而在数据库知识发现中,陈皮也是非常重要的止呕药物,其与姜半夏、茯苓常常可以构成相须相使的增效作用,这一知识发现可为临床提供新的参考价值。

[参考文献]

- [1] 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):264.
- [2] 尚尔鑫,范欣生,段金殿. 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):218.
- [3] 张斌,张晶,史丽君. 正相关性指导下的关联规则剪枝算法[J]. 计算机工程,2011,37(10):1.
- [4] Zhou Zhongmei. Efficiently mining positive correlation rules[J]. Applied Mathematics & Information Science, 2011,5(2):39.
- [5] 李立,周奇,郑光,等. 基于文本挖掘技术分析中成药、西药对慢性胃炎的治疗规律[J]. 中国实验方剂学杂志,2011,17(24):229.
- [6] 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,2010,16(5):215.
- [7] 周滢,周梅,江玉,等. 邓中甲临床药对的配伍选析[J]. 中国实验方剂学杂志,2010,16(7):204.
- [8] 曾令明,唐常杰,阴小雄,等. 基于位图矩阵和双支持度的中药配伍挖掘技术[J]. 四川大学学报:自然科学版,2005,42(2):57.
- [9] 高学敏,王永炎,颜正华,等. 中药学[M]. 北京:中国中医药出版社,2007:55.
- [10] 郭延红,杜薇,王云峰,等. 小半夏加茯苓汤复方及拆方的总生物碱含量测定[J]. 中国实验方剂学杂志, 2011,17(4):64.
- [11] 冯果,刘文,张永萍,等. 结合药效学指标正交优选复方止呕颗粒提取工艺[J]. 中国实验方剂学杂志, 2011,17(18):32.

[责任编辑 邹晓翠]