

· 综述 ·

关联规则在方剂数据挖掘中的应用述评

毛斌 吴宏赞 胡志强

【摘要】从配伍规律研究、方—药—证对应研究、类方研究等方面回顾了关联规则挖掘算法在中医方剂学中的应用。肯定了其在传承专家经验、研究古今医案、发现配伍规律、中医药标准化等方面取得的成果,分析了该方法用于方剂学研究的优点与不足,展望了关联规则在方剂数据库研究中的发展前景,认为方剂数据标准化与关联规则算法创新是该领域未来的研究方向。

【关键词】关联规则; 方剂学; 数据挖掘; 综述

【中图分类号】R289 【文献标识码】A doi:10.3969/j.issn.1674-1749.2013.10.021

Commentary: Application of association rules in prescriptions data mining MAO Bin, WU Hongyun, HU Zhi-qiang. Department of Encephalopathy, Traditional Chinese Medicine Hospital of Peng Lai, Penglai 265600, China.

Corresponding author: HU Zhi-qiang, E-mail: fengchengfengsha@163.com

【Abstract】The application of association rules mining algorithm in prescriptions is reviewed from studies of compatibility law, correspondence of prescription-medicine-syndrome, classified prescriptions, etc. Its achievements in experience inheritance, studying ancient and modern medical cases, exploring compatibility law, standardization of TCM, etc. are affirmed. The advantages and disadvantages for prescription studies are analyzed. The prospects for the development of the association rules in the prescription database studies are forecasted. It is believed that prescription data standardization and association rules algorithm innovation is the interest of future research in this field. Thirty-two articles were referenced.

【Key words】Association rule; Traditional medical formulae; Data mining; Summary

关联规则作为数据挖掘技术的方法之一,最早由 Agrawal 等^[1]于一次国际会议上提出。该方法早先用于经济领域,用作发现大型超市购物清单数据中有趣的关联,故又称之为“购物篮分析”,近年来该方法受到各领域学者的广泛关注,并在中医方剂数据集中得到一定的探索与应用。

1 关联规则在方剂数据挖掘中的应用

方剂,是在辨证的基础上,选择合适的药物,酌定用量,按照组方的要求配伍而成的中药组合。几千年来,中医学积累了极其丰富的中医方剂资料,

历代的医学实践表明,这些方剂具有良好的临床疗效,其中必然蕴藏了大量的中医学知识。然而,传统的统计学方法难以有效的发现这些隐藏的知识,而数据挖掘恰恰适合处理这种非线性的模糊资料。关联规则便是其方法之一。可以将每一个方剂视为一个集合,而方中的每味药物是其一个子集,同时可以有不同脉证、不同剂量、不同给药途径等多个维度,由是便将方剂知识发现的问题转换为关联规则频繁模式挖掘问题,从而发现方剂中的多重规律。目前,运用关联规则对方剂进行数据挖掘主要从以下几个方面展开:

1.1 药对、药组配伍规律的挖掘

近年来,众多学者运用关联规则对中医常见疾病的治疗方剂进行了数据挖掘,探讨了其用药规律。名医医案是方剂资料的重要来源,蕴含着海量的名老中医的宝贵经验,值得深入发掘和继承。吴荣等^[2]对 7 位名老中医治疗冠心病的经典医案进行整理,运用 Weka 3.4 软件对其方剂的配伍规律进

基金项目:山东省中医药科技发展计划项目(2011-039)

作者单位:265600 山东省蓬莱市中医医院脑病科(毛斌);山东中医药大学附属医院脑病科(吴宏赞、胡志强)

作者简介:毛斌(1986-),硕士,医师。研究方向:脑血管疾病研究。E-mail:maobin198606@163.com

通讯作者:胡志强(1957-),女,本科,主任医师,硕士生导师。研究方向:头痛、脑血管病研究。E-mail: fengchengfengsha@163.com

行关联规则挖掘,发现活血药、化痰药、补虚药是其常用配伍组合,同时瓜蒌薤白类方、活血通脉剂、生脉散是名老中医治疗冠心病时的常用方剂。李秀娟等^[3]以《当代名老中医典型医案》为数据源,运用关联规则探索了治疗胃脘痛方剂的用药规律,发现白术、半夏、甘草所组成的药对是当代名中医治疗胃脘痛的常用药对。余俊英等^[4]运用关联规则分析了名医姚荷生治咳嗽医案,总结了其遣方用药规律。吴嘉瑞等^[5]借助中医传承辅助系统软件,挖掘了颜正华教授治疗胃脘痛的方剂用药规律,发现了 49 个核心药物组合与 29 个新处方。此后,该作者使用相同的方法对该教授治疗胸痹的方剂进行了挖掘,总结了用药规律^[6]。以上学者的研究显示了关联规则挖掘方法在名医经验传承上的优势,用规则的形式快速总结出隐藏于医案之中的名医组方配伍规律,使名医经验的表述更加客观、准确。

刘娟等^[7]检索到《中医大辞典·方剂分册》中治疗脾胃的方剂 1046 首,运用关联规则挖掘方法进行研究,该学者对挖掘结果的分析较为全面,将挖掘出的规则分为已知知识和未知知识进行分别处理,证实了茯苓—白术、木香—黄连等已知药对,并发现一些未知药对。宿树兰等^[8]通过检索《中医方剂大辞典》中治疗痛经的方剂,运用关联规则 Apriori 算法对使用频次较高的药物进行了关联分析,发现活血化瘀、温经散寒为古方治疗痛经的主要治法。李建等^[9]采用关联规则 Apriori 算法与熵方法分析了《中医方剂大辞典》中治疗肺痿方剂的用药规律。张欢等^[10]检索了该方剂辞典中治疗哮喘的方剂,运用关联规则进行了古今方剂对比研究,探讨了古今治疗哮喘病的用药异同,此后,该学者^[11]又运用关联规则探讨了其中十八反药对的应用,分别从朝代、反药组合、性味、剂型等方面进行了分析,发现在治疗哮喘中,乌头—半夏为最常见的反药配伍。这种古今对比的研究方法,能够更加清晰的发现所产生规则的特点,在涉及古方的关联规则研究中值得借鉴。刘岩等^[12]使用 SAS 软件的数据挖掘功能对古今治疗水气病的 1521 个方剂进行关联规则挖掘,最后筛选出符合中医理论的关联方 75 个。此外,有学者将关联规则方法与其他数据挖掘方法综合运用,对中医方剂的数据挖掘进行了有益的探索。袁楠等^[13]将聚类分析与关联规则相结合,进行了药对与剂量关系的分析,该方法为方剂量效关系的研究提供了一种新方法。运用关联规则对

方书进行挖掘,能够较好的总结中医疾病的遣方用药规律,为方剂学的研究提供了新的工具。

1.2 方—药—症—证规律研究

方剂是辨证论治的集中体现,挖掘方剂数据库中的知识,可以发现医家对辨证审因的认识、对临床表现与证候关系的认识、对治法治则与药物配伍的认识,是中医药现代化、标准化、诊断量化的必要手段。因此,运用关联规则进行方—药—症—证规律的研究已经成为近年来该领域的研究热点。尚景盛^[14]运用关联规则、因子分析等方法对历代半夏泻心方剂资料进行了挖掘,发现该方除治疗寒热错杂证外还常用于湿热证和肝胃不和证,并挖掘出常用的药物加减变化规律。该学者运用模糊数学方法对不同剂量组合进行了推理预测,具有较高的临床指导价值。时丽莎等^[15]借助个体化诊疗系统对 381 位冠心病病人的数据进行了关联分析,发现基于提升率的挖掘方法可提高效率。龚燕冰等^[16]以 2501 例 2 型糖尿病病历为数据源,运用关联规则分析了方—症—证直接的关联关系,发现了与生脉饮、六味地黄汤、四君子汤相关联的症状,并生成了可视化的网状图,使得挖掘的结果更加直观。李靖等^[17]运用 SQL Server 2005 软件对 201 例 IgA 肾病临床病例进行关联分析,发现了中医证候与现代病理学的相关性,为临床辨证提供了指导,这一研究扩展了关联规则在中医学的研究范围。刘广等^[18]收集了 1032 例胃炎病案,通过关联分析研究了症状之间、处方之间以及症状与处方之间的关联关系,是对中医证候诊断的标准化有益的探索。刘云涛等^[19]采用聚类分析、关联规则等数据挖掘方法,对名老中医防治病毒性肝炎的医案进行了挖掘,发现湿热蕴结、肝气郁滞是病毒性肝炎的常见证候,并挖掘出与黄疸、胁痛、疲乏、纳差、腹胀、肢肿等症状相关的药物,症—证—药相结合,使得对名医经验的继承更加系统化。喻玲等^[20]检索到《中华医典》中以“怒”为主证的方剂 229 个,运用 Clementine 数据挖掘软件进行关联规则挖掘,探讨了怒型与药物选择的关联关系。此外,通过对方剂数据的关联分析,能发现许多隐藏在方剂背后的中医基础理论知识。莫芳芳^[21]运用关联规则分析了“肺与大肠相表里”相关的方剂资料,发现了两条基本的辨证论治模式:(1)肺热多导致大肠热盛而出现多种大肠症状,其中泄泻相对较多,治疗以清热润肺为主,常配伍应用黄芩、生地、麦冬等。(2)肺气壅滞多导致大

肠气滞而出现大便秘结症,治疗用宣肺降气以调畅肺肠气机,常配伍应用杏仁、紫菀、瓜蒌、郁金、桔梗等。关联规则挖掘方法为中医理论的研究开辟了新的路径,关联规则有望成为中医基础理论与临床用药新的连接点。

1.3 类方及其他方剂学相关研究:

类方,是指在药物组成上具有一定相似性的方剂的集合。而关联规则的数学模型恰恰是建立在频繁项集理论基础上的,因此在类方研究中显示了其独特的优势。

张昱等^[22]搜集四君子汤类方 967 首,采用关联规则的 FP-growth 算法挖掘了该类方的主治症状和配伍规律,证实了布尔关联规则用于挖掘类方主治症状与药物之间关联的可行性。叶亮等^[23]对古今治疗痛经的四物汤类方进行了关联规则挖掘,通过古今对比,探讨了该类方的药物组合规律。张文亮等^[24]检索了应用地黄饮子治疗中风病的文献,运用关联规则等方法对该类方的药物归经、功效、配伍等进行了挖掘。陈明等^[25]以《伤寒论》与《金匱要略》为研究对象,对桂枝汤类方的主治病症、桂芍比例、药对配伍等进行了全方位的挖掘。刘娟等^[26]筛选了《中医大辞典·方剂分册》中治疗脾胃病症的白术类方,运用关联规则对药物间及药证间的关联进行了分析。陈仁波等^[27]收集使用黄芪的临床医案 755 例,运用关联规则挖掘了黄芪类方的配伍及相关症状。冯石强^[28]运用关联规则、聚类分析等方法对柴胡类方数据进行了研究,发现了其药对、证候、药症等多方面的规律。以上研究表明关联规则用于类方研究对揭示辨证用药的规律、阐明方剂的主治病症具有重要的意义。此外,关联规则在其他方剂学相关领域的研究中也显示了良好的性能。雷蕾等^[29]运用关联规则对治疗肺癌的中药化学组分进行研究,发现了一些组分配伍的隐含知识。金锐等^[30]对《神农本草经》中 365 味中药的气—味—效三维立体数据进行了关联分析,探讨了中药药性与气味的关系。尚尔鑫等^[31]通过关联规则挖掘了存在配伍禁忌的中药药物属性的特点,初步探索了中药性味归经与禁忌的关系。

总之,关联规则作为一种数据挖掘技术,已渗透到方剂学研究的各个领域,且显示了其良好的适用性。

2 评述与展望

关联规则被提出仅仅 20 年,而用于中医学领域

不过近 10 年,近年来,中医药领域的专家学者对数据挖掘的研究热情不断高涨,基于关联规则进行方剂学研究的文献数量呈上升趋势,这与关联规则自身特点有关。与传统统计学相比,关联规则所处理的数据具有海量性、非线性、模糊性等特征,而这些恰恰是方剂资料所具备的特征。传统统计学用样本估测总体,得到的结果是具体的数值;而关联规则挖掘的是数据库总体中所存在的频繁模式,得出的是规则。正因如此,关联规则技术在传承专家经验、研究古今医案、发现配伍规律、中医药标准化等方面被广泛应用。目前,关联规则作为数据挖掘的方法之一,业已成为方剂学研究中必不可少的数学工具之一。

但总体来说该技术在中医学方剂研究中仍处于起步阶段,尚存在诸多问题。首先,关联规则是基于数据库的。数据库中数据的质量直接决定了规则的质量,虽然中国已建立了众多较为全面的方剂数据库,但由于方剂数据自身的特点,仍存在药物名称不统一、大量模糊数据(如药物用量“一枚”、“少许”、“不拘多少”等)、数据缺失与冗余并存等诸多问题,这在古代文献的数据库中更为突出。如何消除数据库中这些噪声,使清洗后的数据更为符合关联规则挖掘的要求已成为该项技术的难点。其次,关联规则是基于频次的。某种药物出现的频次多少直接决定了其支持度和置信度的大小,那些大众化的药物会获得较高强度的关联,从而产生大量无趣的垃圾规则,同时不可避免的会丢失稀有数据(例如某些仅针对罕见症状所选用的疗效确切的单味药物),虽然通过提升度、置信率、正态卡方、信息差等指标可对关联规则进行剪枝,最大限度的剔除垃圾规则,同时通过设定也可得到部分研究者感兴趣的稀有数据,但仍然存在两大风险:相信了实际不存在的规则和抛弃了实际有意义的规则。最后,关联规则离不开人的分析。挖掘出的规则究竟有没有实际意义还需要中医理论的指导,对药物组合的解释、方—药—证三者之间的内在联系等等都是研究者对规则的主观评价的过程,往往说服力不够强,尤其是那些隐藏的不为先前经验所知的规则,常常需要后续大量的临床及实验研究的支持。

因此,关联规则在方剂学研究中未来的方向,一方面应放在数据准备阶段,着力提高数据库的质量,解决方剂数据标准化问题;另一方面应放在算

法创新上,将关联规则的新算法与方剂学相结合,找到更为适合中医方剂资料的算法,使得出的规则更加客观、合理,更加符合中医学的需要。

总之,方剂学研究与数据挖掘、计算机技术相结合是大势所趋,是中医药现代化、标准化、规范化的必由之路。方剂数据具有复杂性及信息的海量性、效应多靶点性等特征,这决定了数据挖掘的重要性^[32]。作为数据挖掘重要手段的关联规则,如何更好的应用于方剂数据库挖掘,将成为该领域的研究重点。

参 考 文 献

- [1] Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases [C]//ACM SIGMOD Record. ACM, 1993, 22(2): 207-216.
- [2] 吴荣,刘晔,王阶,等.基于关联规则的名老中医冠心病用药规律研究[J].中国中药杂志,2007,32(17):1786-1788.
- [3] 李秀娟,张天嵩,张素,等.基于数据挖掘探索名老中医治疗胃脘痛的用药规律[J].福建中医药,2012,43(3):1-3.
- [4] 余俊英,查青林,黄利兴.姚荷生治疗咳嗽医案关联规则分析[J].江西中医学院学报 2011,23(2):17-20.
- [5] 吴嘉瑞,张冰,杨冰.基于关联规则和复杂系统熵聚类的颜正华教授治疗胃脘痛用药规律研究[J].中国实验方剂学杂志,2012,18(20):1-5.
- [6] 吴嘉瑞,张冰,杨冰,等.基于关联规则和熵聚类算法的颜正华治疗胸痹用药规律研究[J].中国实验方剂学杂志,2013,19(2):348-351.
- [7] 刘娟,蒋永光,任玉兰,等.关联规则在中药药对挖掘中的应用[J].时珍国医国药,2006,17(4):492-493.
- [8] 宿树兰,尚尔鑫,叶亮,等.治疗痛经方药的关联规则分析[J].南京中医药大学学报,2008,24(6):383-385.
- [9] 李健,卢朋,张瑞贤,等.《中医方剂大辞典》中治疗肺痿方剂的用药规律分析[J].中国实验方剂学杂志,2012,18(10):1-5.
- [10] 张欢,范欣生,王崇骏,等.古今哮喘方用药规律对比的关联规则研究[J].中国中医药信息杂志,2009,16(3):94-96.
- [11] 张欢,范欣生,陶静,等.基于关联规则等方法的古今哮喘方中十八反药对的应用研究[J].南京中医药大学学报,2010,26(2):89-92.
- [12] 刘岩,范昕,李康.关联规则分析在中医古代医籍方剂筛选中的应用[J].中国医院统计,2010,17(4):294-297.
- [13] 袁楠,金晖,田玲,等.基于聚类和模糊关联规则的中医药对量效分析[J].计算机应用研究,2009,26(1):59-61.
- [14] 尚景盛.半夏泻心汤配伍规律数据挖掘方法初探[D].北京:北京中医药大学,2005.
- [15] 时丽莎,鲁明羽,李丹.关联规则在冠心病中医临床证型诊断中的应用[J].广西师范大学学报:自然科学版,2008,26(3):206-209.
- [16] 龚燕冰,罗曾刚,高思华,等.基于关联规则的 2 型糖尿病方症关系的横断面研究[J].中医杂志,2010,51(1):64-67.
- [17] 李靖,王硕仁,徐冰,等.基于关联规则的 IgA 肾病中医证候与病理相关性的探讨[J].北京中医药,2011,30(9):653-655.
- [18] 刘广,吴磊.基于关联规则的中医胃炎诊疗与处方间的挖掘研究[J].江苏中医药,2011,43(11):72-74.
- [19] 刘云涛,郑丹文,罗翌.数据挖掘技术在名老中医防治传染病经验传承中的应用体会[J].临床医学工程,2012,19(8):1343-1345.
- [20] 喻玲,林蓉,吴丽丽.从中医传统方剂初步分析中医七情之怒临床用药特点[J].武警医学院学报,2010,19(4):319-321.
- [21] 莫芳芳.基于古代医案分析的“肺与大肠相表里”理论的应用研究[D].北京:北京中医药大学,2012.
- [22] 张昱,陈云慧,王燕,等.布尔关联规则挖掘四君子汤类方药物配伍规律研究[J].江苏中医药,2008,40(7):67-68.
- [23] 叶亮,范欣生,卞雅丽,等.古今治疗痛经的四物汤类方关联规则研究[J].南京中医药大学学报,2008,24(2):94-96.
- [24] 张文亮,张斌,赵耀武,等.地黄饮子类方治疗中风恢复期的用药规律及研究[J].甘肃中医学院学报,2010,27(1):23-27.
- [25] 陈明,陆建峰,赵国平.基于双向强关联规则的仲景桂枝汤类方研究[J].南京中医药大学学报,2009,25(4):249-251.
- [26] 刘娟,蒋永光,胡波,等.白术类方的药证关联分析[J].成都中医药大学学报,2004,27(4):55-56.
- [27] 陈仁波,张明明,张志强,等.基于数据挖掘的黄芪临床用量及相关应用分析[J].环球中医药,2011,4(6):438-441.
- [28] 冯石强.基于数据挖掘对柴胡类方方证规律的研究[D].北京:北京中医药大学,2012.
- [29] 雷蕾,崔蒙,秘仲凯.关联规则挖掘在治疗肺癌组方中药发现中的应用研究[J].中国中药杂志,2010,35(16):2192-2195.
- [30] 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):794-803.
- [31] 尚尔鑫,范欣生,段金庵,等.基于关联规则的中药配伍禁忌配伍特点的分析[J].南京中医药大学学报,2010,26(6):421-424.
- [32] 范欣生,尚尔鑫,王崇骏,等.方剂研究中数据挖掘方法的适用性探讨[J].南京中医药大学学报,2008,24(6):379-382.

(收稿日期:2013-04-01)

(本文编辑:蒲晓田)