

DOI 10.19656/j.cnki.1002-2406.20221207

信息工程

关联规则在中医药数据挖掘中的应用

杨柳, 徐晴文, 黄汝佳, 朱姝静, 梁昊[✉], 彭清华

(湖南中医药大学, 湖南 长沙 410208)

【摘要】 关联规则作为数据挖掘的重要方法之一,在中医药领域的应用逐渐广泛与深入。笔者介绍了关联规则等基本相关概念,同时就关联规则的常用算法和软件进行了总结与对比分析;比较了购物篮数据与医学数据的异同;创新性地从不同维度联系的角度出发,对关联规则在中医药领域中症状、证素、方药等方面的应用进行归类和实例分析,分别归纳了关联规则在各领域应用的特点;并指出了关联规则在中医药数据挖掘领域中不足。

【关键词】 中医药;数据挖掘;关联规则

【引用格式】

杨柳,徐晴文,黄汝佳,等. 关联规则在中医药数据挖掘中的应用[J]. 中医药信息, 2022, 39(12): 35-40.

YANG L, XU Q W, HUANG R J, et al. Application of association rules in TCM data mining[J]. Information on TCM, 2022, 39(12):35-40.

关联规则是指两个或者多个变量的取值之间存在某种规律,该规律不涉及因果关系,但往往会存在“同时存在”或“从一对象可推出另一对象”的现象。关联规则作为数据挖掘的重要方法之一,在中医药领域的应用越来越广泛。但目前关联规则在中医药领域的应用仍较为粗浅生硬,处于“数据丰富,信息匮乏”的局面。而较复杂的规则和算法以及繁多的分析平台选择,依然会造成从业者的研究困扰以及结果的不准确性。本文从关联规则相关概念的介绍入手,在分析比较常用算法及软件的同时,着重阐释了目前关联规则在中医药领域的应用,促进医疗工作者对医疗数据之间的关联进行更好地理解及运用,在疾病的诊断与治疗中做出更高效、更准确的医疗决策,从而推动计算机辅助在中医药路线的发展。

1 关联规则方法简论

1.1 关联规则及购物篮数据概念

关联规则分析又称为购物篮分析。超市从大量商

品数据分析中,发现了顾客购买商品之间的规律,将经常同时被购买的商品安排到相邻货架,从而提高销售额,这一行为被称为“购物篮分析”,即从大量数据中发现隐藏的联系与规律。

1.2 关联规则相关概念

1.2.1 项集

所有事务中每一对象可以称为一个项,一个或多个项的组合称为项集。

1.2.2 支持度计数 σ

指项集在事务中出现的次数。

1.2.3 支持度 S

包含项集的事务 X 占有所有事务数量 N 的比例。

$$X \rightarrow Y: S = \sigma(X)/N = \text{count}(X)/N$$

1.2.4 频繁项集

如果对项集的支持度设定一个最小阈值,那么所有支持度大于这个阈值的项集就是频繁项集。

基金项目: 湖南省科技创新人才计划优秀博士后创新人才项目(2020RC2061);中国博士后科学基金项目(2020M682578);湖南省学位与研究生教学改革研究项目(2021JGYB117);湖南省教育厅科学研究项目(21B0366)

第一作者简介: 杨柳(2001-),女,湖南中医药大学学生,主要研究方向:中医数据挖掘。

✉通讯作者简介: 梁昊(1986-),男,博士,副教授,硕士研究生导师,主要研究方向:中医诊断学。

1.2.5 置信度 C

是指 Y 在包含 X 的事务中出现的频繁程度。

$X \rightarrow Y: C = P(Y|X) = s(X, Y)/s(X) = \text{count}(X, Y)/\text{count}(X)$

1.2.6 提升度

判断 Y 在整体个数据集出现的频率是否和 X 存在的频率等同。

$\text{Lift} = P(X \cap Y)/P(X) \cdot P(Y) = c(X \rightarrow Y)/s(Y) = P(Y|X)/P(Y)$

如果该值等于 1, 说明两个条件没有任何关联。如果小于 1, 说明 X 与 Y 是负相关的关系, 意味着一个出现可能导致另外一个不出现。大于 1 则表示具有正相关的关系。一般在数据挖掘中当提升度大于 3 时, 挖掘出的关联规则是有价值的。

提升度是一种比较简单的判断手法, 实际中受零事务(既不包含 X 也不包含 Y 的事务)的影响比较大。所以如果数据中含有的零事务数量较大, 该度量则不适合使用。

1.3 关联规则相关算法

关联规则算法大致可分为宽度优先算法、深度优先算法、数据集划分算法、采样算法和增量式更新算法等^[1]。算法一般分为两个关键步骤: ①生成所有频繁项集; ②生成强关联规则, 即找出频繁项集中大于等于最小置信度阈值的关联规则^[2]。目前在相关领域关联规则较常用的算法主要是 Apriori 算法和 FP-growth 算法。

1.3.1 Apriori 算法

Apriori 算法是宽度优先算法的一种, 它需扫描数据库的次数等于最大频繁项集的项数。Apriori 算法有两个致命的性能瓶颈: ①产生的候选集过大(尤其是 2-项集), 算法必须耗费大量的时间处理候选项集; ②多次扫描数据库, 需要很大的 I/O 负载, 在时间、空间上都要付出很大代价^[1]。但王姣等^[3]根据支持的反单调性这一特性, 主要通过候选产生和候选前剪枝这两个关键步骤, 对 Apriori 算法进行了改进, 极大提高了计算效率。流程图见图 1。

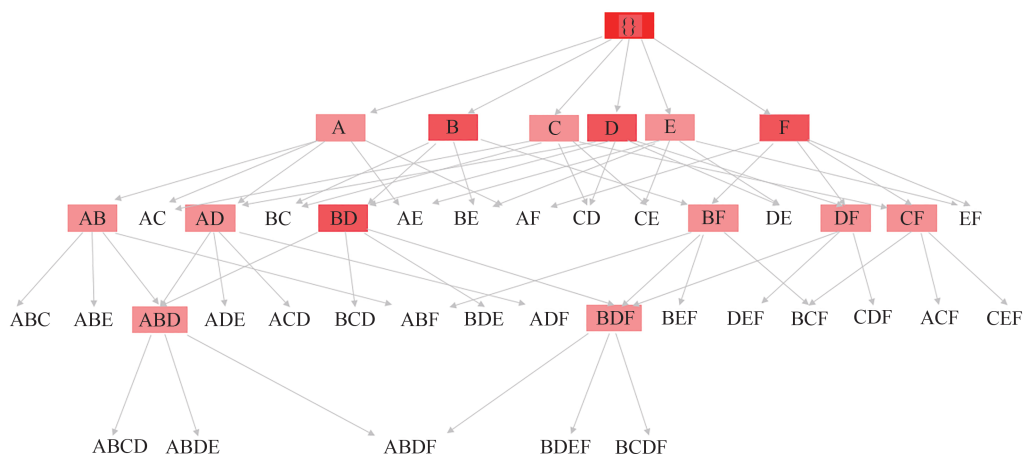


图 1 Apriori 算法流程图

1.3.2 FP-growth 算法

FP-growth 算法是深度优先算法中最新最高效的, 且从本质上不同于 Apriori 算法的经典算法。在算法中有两个关键步骤, 一是生成频繁模式树 FP-tree, 二是在频繁模式树 FP-tree 上挖掘频繁项集。流程图见图 2。

与 Apriori 算法相比, FP-growth 算法具有以下优点: ①FP-growth 算法只需扫描数据库 2 次, 避免多次扫描数据库; ②不需要产生庞大的候选项集, 在挖掘过程中大大减少了搜索空间, 在时间效率和空间效率上都有一个量级的提高^[1]。而杨云等^[4]更是在此基础上提出了 FP-growth 的改进算法, 进一步缩短了建立 FP-tree 的时间。

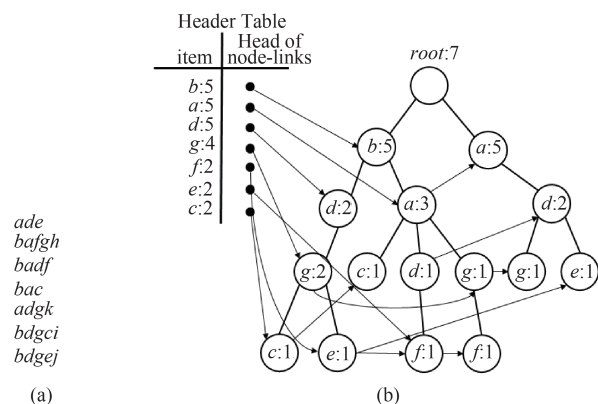


图 2 FP-growth 算法流程图

2 关联规则常用软件

探讨数据之间的关联规则往往需要一些辅助软件

进行分析,目前较为常见的有 SPSS Modeler、R 等,在中医药领域较为常用的还包括中医传承辅助平台或古今医案平台^[5],为有相关数据处理需求的人员提供了极大的便利。

2.1 中医传承辅助平台(TCMISS)

该平台主要应用于中医药研究领域,目前主要集中在名老中医经验研究、疾病用药规律研究、组方配伍研究等方面。软件将信息资料的录入、保存、提取、统计和分析等功能融为一体,优势在于方便非专业数据处理人员操作,同时拥有较丰富和全面的中医药领域信息数据等^[6]。但是应用范围较窄,目前仅用于药物配伍规律分析^[7]。TCMISS 一般仅提供简单粗糙的浅层分析,无法满足更深度的分析需求。

2.2 SPSS Modeler

在研究中,SPSS 公司推出的数据挖掘工具 SPSS Modeler 软件中关联规则分析运用得尤为广泛。作为一个开放式数据挖掘工具,其分析过程及结果非常容易与直观。在使用操作简单、不需要任何编程工作的基础上,拥有强大的数据预处理和转换功能,可提供丰富的数据挖掘模型和灵活的算法^[8]。但在使用过程中,该软件存在离群点检测、特征选择等节点预测性能不佳等问题^[9]。

2.3 R

R 是集统计分析于一体的、几乎可以满足所有科研需求的统计分析软件。它拥有多种数据分析方法,在具有强大的挖掘分析能力和扩展性的同时,可以通过 arulesViz 和 arules 等功能包,对数据进行更好的可视化输出^[10]。但是由于它无法在 R 语言的基础上开发互联网应用程序,其最终结果往往通过其他语言显示^[11]。

3 医学数据与购物篮数据的异同

关联规则起源并最早应用于购物篮数据,由于其自身高度的兼容性和可操作性,现如今也被广泛应用于医学数据特别是中医药的数据处理中。如在医学诊疗方面,可以提高诊断及治疗精准度,给人们带来更高效率和更高质量的医疗服务;在药物应用方面,可以不断地优化药物联用组合的同时提供开发新药的方向;此外,在公共卫生管理、医疗器械使用、医疗人员规模等一些方面均有不同程度的优化作用^[12]。

在面对数据处理时,医学数据与购物篮数据存在诸多共同点。如只要将患者与疾病比作不同商品,根据其呈现的置信度支持度等数据,便能得到患者与疾病之间的对应联系与规律,极大地提高了疾病诊疗效率。该流程可以应用于医学众多不同领域和数据中,如药物开发研究、医疗设备使用情况,再如中医证候、

症状、中药配伍之间的关联等。

然而医学数据和超市商品数据也存在着不同。一方面,超市商品数据关联可来源于日常经验,医学数据关联需要大量演算分析和专家经验。超市数据的关联规则的逻辑常建立于家庭这个整体,如一个家庭同时存在需要喝啤酒的成年人、使用尿布的婴儿,且牛奶与面包常常是固定的早餐搭配等,其数据逻辑简洁清晰。而医学数据常常作用于单独个人,结果常具有不可控性,如一位胃癌患者卷入一场意外中导致失明,而胃癌与失明并没有直接上的联系。因此医学数据往往需要更大量的数据证明其中的关联不是偶然^[13]。

另一方面,超市商品数据关联为单维联系,而医学数据为多维联系。超市商品数据仅仅是商品之间的单维联系,一旦撤去中间载体,数据关联则无法解释。如前文中显示,如果不用家庭作为载体,则无法解释啤酒与尿布等商品之间的关系。而医学数据不同,如中医常利用关联规则探究名老中医用药经验规律时,常常从理法方药各方面分析,一旦形成症-证-药等完整逻辑链条后,任意两点又可组合并拥有合理的解释。

4 关联规则在中医药数据挖掘的具体应用

关联规则在中医药领域的应用方式,最主要是通过从历代所流传下的医案典籍和当代中医临床实践经验中,获取大量可靠数据加以分析,总结出理法方药之间的一系列规律。在中医药数据挖掘中对关联规则的使用,不管是对医学临床事业还是中医学术经典的传承和改进,都起了极大的推动作用。笔者将从症状、证素、药物本身的一维关联及彼此之间相互联系的二维关联进行如下探讨。

4.1 一维关联

一维关联指的是同一性质的事物中点与点构成的线性联系。通过发现一个事件上两个或多个点之间的联系,从而可以推出该事件中绝大多数点之间的联系与规律。具体在中医药数据上,一般是局限在症-证-药的某一个层级内部。层次图见图 3。

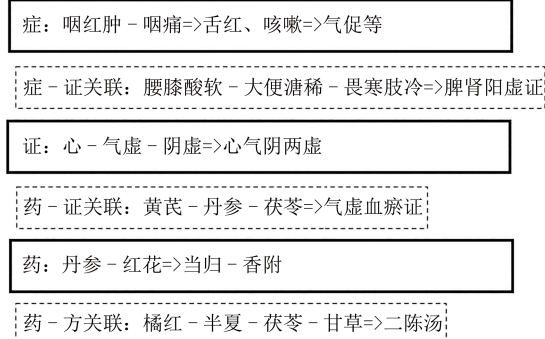


图3 症-证-药层次图

4.1.1 症状关联

目前在中医研究症状之间的关联上,研究方式往往仅是简单基于某种病或证的前提下,然后就症状或症状组合与其研究内容关联,如构成症-证、症-药关联等,然后进行进一步分析。而症-证关联往往是单纯作为其中一环,对所研究病证进行机理分析,增强其症状联系等方面的认识。如艾军等^[14]在研究小儿肺炎的证候病机时,从热、郁、痰三个角度出发,通过关联规则分析得到各自核心症状组合,如咽红肿痛-舌红、咳嗽-气促、咯痰-肺湿啰音等,然后进一步从核心症状组合分析其机理,进一步增强对小儿肺炎证候病机的理解。

目前症状关联相关研究领域仍较为浅层,除了作为中间分析媒介,它同时可以作为独立主题来讨论。如通过探讨症状与症状之间的联系规律,研究症状之间的共通机制,预防并发症的发生以及为后续治疗方法提供新思路都具有借鉴意义。

4.1.2 证素关联

证素是指构成证候的基本要素,可以分为病位证素和病性证素。但其三阶双网证素辨证体系和“降维升阶”等思想,在利于证候特点归纳的同时,也提供着证候-证素-证名组合千变万化的可能性^[15]。

研究证素关联,首先就在于归纳一组或多组症状的特性,然后与证素相关联。其次再是研究疾病或证候中不同证素之间的关联。前者如虞海虹等^[16]在进行基于教材挖掘中医证素诊断规则研究时,就挖掘出如腰膝酸软-肾、月经量少-胞宫等病位病素诊断规律,也同时得到了如口苦、舌苔黄、尿短黄-热证等病性证素诊断规律。而后者如杨继等^[17]从 2 831 例新型冠状病毒肺炎患者中,通过关联规则分析,得出新型冠状病毒肺炎的病性证素以湿、热为主,病位证素则多与肺、脾肾相关。进一步得到 4 个主要证候类型,其中就包括以湿、热、肺-湿热郁肺证为重要组合,为中医研究相关领域提供的重要参考资料。

证素作为症与证之间的媒介,可以推动对症-证关联更好的理解。通过关联规则研究不同证素之间的组合规律,更有利于证候的明确诊断及相似证候之间的鉴别。同时,通过研究不同证候之间相同的证素,可以为研究证候之间的联系提供新的思路。

4.1.3 药物关联

目前关于药物关联的研究,大多是在治疗某一种病证的基础上,对药物配伍进行病-药关联分析。而药物之间的关联,可以根据单药或药对进行配伍规律总结,然后就所应用病证方面进行分析。如崔一然等^[18]为了探讨丹参-红花药对的用药规律,在《中医方

剂大辞典》中提取相关方剂,发现丹参红花-香附、丹参红花-白芍等核心药对多归于心、肝经,且丹参红花常与行气活血等药物联用等,这一规律对指导临床遣方用药具有重要意义。而胡慧明等^[19]在探讨山楂的组方药对规律时,不仅得出了其核心组合药对、性味归经的相关结论,同时在得到了主治疾病分布的情况下,分析了山楂在具体病证中所发挥的独特功效与优势,进一步明确了其在中医临床中的合理应用。此类研究对传统中药的作用机理及其临床使用提供了新的思路,对开发新药也起到了极大的推动作用。

4.2 二维关联

二维关联指的是不同性质事物之间构成的线与线联系。事件在选取具有代表性的点组成不同的线性关系后,通过与另一事件所选出的线性关系组合,从而发现两个或多个事件之间的联系规律。

一般二维关联的构建,往往是通过发现两个事件 a 与 b 之间的项集联系。首先在利用支持度设定阈值之后,在事件 a 和 b 中收集项集。两个事件各自提取项集组合形成集合,同时组成的不同集合又构成一个大型项目集。从大型项目集随机抽取子集,并在设定支持阈值后进行算法的计算,此时生成的项目集包括:仅包括 a 的集合、仅包括 b 的集合和同时包括 a 和 b 的集合。为探讨 a 与 b 的关联,选用同时包括 a 与 b 的集合。最后通过置信度、支持度等值进行降序排列找到具有代表性的集合,再从中分析 a 与 b 的关系^[13]。

4.2.1 症-证的关联

中医证候往往会由不同的症状组成,如何快速抓住核心症状进行证候诊断,提高诊疗速度与质量,利用关联规则探讨症-证之间的关联显得十分必要。如张欣欣等^[20]通过关联规则分析得出心力衰竭相关证候与其相对的核心症状,如心肾阳虚证对应的核心症状是双下肢水肿-心悸,而气虚血瘀证的对应核心症状是胸闷-乏力以及舌质紫暗-乏力等,该结果对心力衰竭的证候分析起到了极大的作用。

研究症-证关联不仅有利于提高疾病相关证候的鉴别能力,同时对研究相关证候本身有一定意义。如林犷^[21]在研究慢性肾小球肾炎的相关证型及症状时,关于脾肾虚证的分型及表现就曾提到,脾肾阳虚[虚寒]证的核心症状是腰膝酸软-大便溏稀-畏寒肢冷等;而脾肾两虚[亏虚]证的核心症状却是舌苔薄-乏力-腰痛-脉濡等。这可能是由于构成证候的病性、病位不同,其所对应的具体症状也会有所改变。通过几组相似证候中不同核心症状的鉴别分析,有利于更好地提高证型分辨的效率。

4.2.2 药-方的关联

由于中医诊治需辨证论治,临床使用某首方剂时药物组成往往会有所不同,或者同种药物在不同的方剂使用时可以发挥不同的功用。为了研究相关方剂的核心传承以及促进临床药物的灵活使用,关联规则对于探讨中药与相关方剂之间的联系显得尤为必要。

其中最直白的联系表现为,同样的方剂在面对不同临床表现时所进行的药物加减变化,如陈丽萍等^[22]在探讨名老中医应用二陈汤规律后,发现虽然经常有所加减,但橘红-半夏-茯苓-甘草为二陈汤最核心的药对,另外通过分析最常用的配伍药物及功效,如莱菔子-紫苏子-白芥子合用降气化痰、白术-山药补气健脾等,有利于指导二陈汤的临床运用。

而方-药关联的内容,除了具体方剂的药物组合分析以外,还可以从发现新处方、归纳方剂效用、量效关系研究等多个方面阐释分析^[23]。对相关方剂的关联规则进行分析,不仅有利于经典药方更好地传承,对指导临床临证加减也具有极大的意义。

4.2.3 药-证的关联

通过关联规则探讨药-证之间联系,主要是针对某种证型找出其关联度最高的药物或药对。一般来说,目前大多数研究多是从病-证-药的大前提出发进行药-证对应的,这是因为相同的证型所对应的疾病往往不同,根据具体病证的病位等不同,所归纳的药物规律也会有所差异。

比如同样是研究治疗气虚血瘀证的用药规律,乔利杰等^[24]在基于文献研究的基础上统计治疗冠心病心力衰竭的单味药物时,发现治疗的主要药物为黄芪、丹参、川芎;任风英等^[25]通过研究李跃华的用药规律发现,其治疗慢性脑供血不足使用频次前3名药物分别是葛根、川芎、炙黄芪;而申鑫惠等^[26]在研究了当代医家用药规律后,发现在治疗糖尿病周围神经病变时,使用频次较高的药物为黄芪、当归、川芎。通过对比发现,这些治疗的核心药物在存在差异的同时,也存在共同药物。黄芪-川芎这一共同药对可能是治疗气虚血瘀证的核心药对。研究药-证之间关联,能为中医证型和药物的现代化提供相互促进的作用,为证型的相关临床研究以及药物的药理分析等提供研究方向。

4.2.4 症-药的关联

由于中医临床症状成因往往因人而异,当下大多数症状与药物的联系往往以病-证-症-药的方式呈现,而不是直接的联系。然而也存在不同情况。

第一,当某些病理现象既是病又是症时,那么就可以构建直接联系。如张聪等^[27]研究治疗缓慢性心律失

常的症药关联时,当缓慢性心律失常作为一种病时,它需要通过缓慢性心律失常(病)-乏力、胸闷(症)-附子、甘草(药)这样的组合间接构建关联;而作为一种症时,缓慢性心律失常却可以与附子、丹参-甘草、黄芪-桂枝-甘草这样的药对构建直接关联。同时也可以进一步深挖在症-药对应关系下,如药物分类、归经等关联,如同样是探讨骨质疏松症的用药规律,郑升鹏等^[28]在发现配对药物后,进一步得出治疗骨质疏松的中药多为补阳药、活血化瘀药,而归经多在肝肾两经的结论。

第二,从总结某名家医案典籍出发,总结医家治疗经验。张琴^[29]基于《伤寒论》总结了症状与药物的对应规律,如治疗发热时常用的药物为桂枝、甘草、芍药、大黄、半夏等,这些药物的药性多是辛温、甘温,涉及病种多为三阳病,符合病在三阳多实热少虚寒的特点。再如与渴、身痛-猪苓、泽泻以及大便难、发热-厚朴、枳实等多组证候组及其相关药对,提示该类药物对相应症状(组)或具有特异性的治疗作用,对指导临床上具有借鉴意义。

5 小结

尽管关联规则在中医数据挖掘中的应用逐渐广泛、深入,为中医药的发展研究做出了重大贡献,并不断提供新的思路,然而关联规则在中医药领域的应用依然存在局限。

首先,许多研究过于粗浅,只是依赖古今医案云平台或中医传承辅助平台进行粗浅的挖掘。一者,中医辅助平台往往只能帮助非数据分析专业者进行数据处理,无法进行更深入的分析操作。二者,大多数研究在利用关联规则上,往往仅是对某些病症进行中药配伍规律分析,大部分结果都是挖掘出的共识或常识,难以发现一些隐性的知识或规律。

其次,许多研究对关联规则参数理解不深,仅仅依靠支持度和置信度进行数据支撑,而对提升度的参数不重视,这往往会导向一个并不准确的结果。如在分析治疗某种病症的核心药物时,若仅靠置信度和支持度,结果往往指向使用频次最高的药物,然而频次高的药物却不一定是起核心药效的药物,这就会导致后续研究内容的偏差,这时往往需要提升度等指标加以佐证。

再次,正如前文提及,现如今许多研究往往是中医利用中医辅助平台进行分析,而缺乏中医与计算机学者的交叉合作,虽然便于非专业人士进行数据分析的操作,但也导致大多关联规则的研究过于粗浅,仅仅是简单的一对一关联规则分析,而对于多维和多级关联规则分析等更高级的方法在中医药数据挖掘的应用还不够。

尽管关联规则在中医数据挖掘应用上仍存在问题,但依然对中医药的发展起到了不可忽视的作用,存在着广阔的应用前景。关联规则、网络模型可视化可以应用到更多领域,如结合网络药理学揭示中药作用于疾病靶点的具体机制,进一步提高中医药对疾病的诊疗水平,从而推动中医药事业现代化的进程。

【参考文献】

- [1] 王爱平,王占凤,陶嗣干,等. 数据挖掘中常用关联规则挖掘算法[J]. 计算机技术与发展,2010,20(4):105-108.
- [2] 崔妍,包志强. 关联规则挖掘综述[J]. 计算机应用研究,2016,33(2):330-334.
- [3] 王姣,马静雅,谷丰强,等. 基于关联规则的数据挖掘的研究与应用[J]. 粘接,2020,41(5):95-98.
- [4] 杨云,罗艳霞. FP-Growth算法的改进[J]. 计算机工程与设计,2010,31(7):1506-1509.
- [5] 梅泰中,许吉,张洋,等. 2012~2017年中医药数据挖掘研究状况分析[J]. 数理医药学杂志,2020,33(1):70-72.
- [6] 唐仕欢,申丹,卢朋,等. 中医传承辅助平台应用评述[J]. 中华中医药杂志,2015,30(2):329-331.
- [7] 刘凯,王艳国. 中医传承辅助平台在中医药研究中应用现状[J]. 山东中医杂志,2015,34(5):392-394.
- [8] 王玲玲,付桃芳,杜俊英,等. 基于SPSS Clementine的关联规则分析在中医药数据挖掘中的应用优势和局限[J]. 云南中医学院学报,2016,39(6):98-102.
- [9] 张钰莎,蒋盛益. Clementine软件功能缺陷分析[J]. 信阳师范学院学报(自然科学版),2015,28(3):450-453.
- [10] 陈荣鑫. R软件的数据挖掘应用[J]. 重庆工商大学学报(自然科学版),2011,28(6):602-607.
- [11] 张春龙. 统计分析工具和R语言在科学研究中的应用比较[J]. 科技创新导报,2017,14(31):113-114.
- [12] 刘慧冉,马继炎,谯治礁. 关联规则在医学数据中的应用[J]. 中国当代医药,2015,22(30):18-21,25.
- [13] DODDI S, MARATHE A, RAVI S S, et al. Discovery of association rules in medical data[J]. Informatics for Health and Social Care, 2001,26(1):25-33.
- [14] 艾军,汪爱传,戴铭,等. 基于症状关联规则的小儿肺炎中医证候

- 病机学研究[J]. 北京中医药大学学报,2016,39(2):140-143.
- [15] 叶桦,冯全生,温川颀,等. 证素辨证的基本原理与问题分析[J]. 中华中医药学刊,2019,37(12):2840-2842.
- [16] 虞海虹,黄文金,杨雪梅. 基于教材挖掘中医证素诊断规则研究[J]. 山东中医药大学学报,2017,41(1):29-31.
- [17] 杨继,张垚,张春阳,等. 2831例新型冠状病毒肺炎患者中医证素特点的文献调查分析[J]. 江苏中医药,2020,52(4):74-80.
- [18] 崔一然,刘欣,申丹,等. 《中医方剂大辞典》含丹参、红花药对组方规律数据挖掘分析[J]. 中国中药杂志,2016,41(3):528-531.
- [19] 胡慧明,翁家俊,朱彦陈,等. 基于数据挖掘的《中医方剂大辞典》含山楂组方用药规律研究[J]. 中国现代应用药学,2021,38(21):2713-2720.
- [20] 张辛欣,焦华琛,李运伦. 心力衰竭中医证型与症状的关联规则分析[J]. 山东中医药大学学报,2020,44(3):252-260.
- [21] 林犷. 慢性肾小球肾炎的中医症状-证候-药物关联规则挖掘的研究[D]. 成都:电子科技大学,2016:46-50.
- [22] 陈丽平,李建生,蔡永敏. 基于数据挖掘的名老中医应用二陈汤规律分析[J]. 中国实验方剂学杂志,2017,23(12):201-207.
- [23] 刘梦玲,章新友,丁亮,等. 数据挖掘方法在中药配伍规律研究中的应用与进展[J]. 中国中药杂志,2021,46(20):5233-5239.
- [24] 乔利杰,彭广操,卫靖靖,等. 基于文献研究分析冠心病心力衰竭气虚血瘀证的用药规律[J]. 中国实验方剂学杂志,2021,27(9):176-183.
- [25] 任凤英,鲍婷婷,王博深,等. 基于中医传承辅助平台分析李跃华治疗慢性脑供血不足气虚血瘀证的用药规律[J]. 辽宁中医杂志,2020,47(10):128-132,223.
- [26] 申鑫惠,杨宇峰,石岩. 基于中医传承辅助平台探析当代医家治疗糖尿病周围神经病变的用药规律[J]. 中华中医药学刊,2021,39(8):45-49.
- [27] 张聪,侯平. 中医治疗缓慢性心律失常症药关系研究[J]. 中医杂志,2016,57(14):1247-1250.
- [28] 郑升鹏,曾志奎,姜自伟,等. 基于数据挖掘的骨质疏松症用药规律研究[J]. 中华中医药学刊,2018,36(2):341-344.
- [29] 张琴. 基于关联规则的《伤寒论》症-药关系研究[D]. 北京:北京中医药大学,2017:23-26.

(收稿日期:2022-04-20)

Application of Association Rules in TCM Data Mining

YANG Liu, XU Qingwen, HUANG Rujia, ZHU Shujing, LIANG Hao[✉], PENG Qinghua
(Hunan University of Chinese Medicine, Changsha 410208, China)

【Abstract】 As one of the important methods of data mining, association rules are widely and deeply applied in the field of TCM. This study introduced some basic concepts of association rules, and summarized and compared the common algorithms and software of association rules. The similarities and differences between shopping basket data and medical data were compared. Innovatively, it classified and analyzed the application of association rules in the fields of TCM, from the perspective of different dimensions, including symptoms, syndromes and prescriptions. It summed up the characteristics of the application of association rules in various fields, and also pointed out the shortcomings of association rules in the field of TCM data mining.

【Key words】 TCM; Data mining; Association rules