

DOI: 10.13288/j.11-2166/r.2017.10.020

综 述

数据挖掘方法在名老中医用药规律研究中的应用现状

于林童, 曲文白, 余新波, 徐浩*

(北京中医药大学基础医学院, 北京市朝阳区北三环东路 11 号, 100029)

摘要 根据具体方法的不同, 分别从频数分析、关联分析、复杂网络分析、聚类分析等不同角度, 对近 10 年数据挖掘方法在名老中医用药规律中的应用加以综述。总结各数据挖掘方法在名老中医治疗疾病的用药频次、配伍规律、用药药性及药-症关联等方面研究中的应用方法及应用成果, 对方法以及研究目的适应关系做出相应评价, 提出对于不同研究目的较为常用的具体方法, 并认为数据挖掘可为发现名老中医用药的隐性知识提供思路。

关键词 数据挖掘; 名老中医经验; 用药规律

名老中医作为中医药学术发展的代表和传承的先驱, 其学术经验的总结和传承不仅能推动中医药学术的进步, 也对丰富中医药理论体系起到了巨大作用, 同时也是培养新一代中医人才的重要途径。中药的使用与配伍是中医治疗的核心内容之一, 其使用规律反映了名老中医多年临床经验以及对中医学理论的理解。因此对名老中医临床辨证和对症用药经验的传承具有重要价值。

在对名老中医用药规律的研究领域, 以数学方法为研究基础的数据挖掘技术发挥着重要作用。数据挖掘技术是通过多种算法对大量的数据进行分析, 并搜索隐藏于其中的信息, 加以提炼与分类, 有序地将规律性的信息体现出来。中药处方在用药频次、药物配伍和药证关联上都具有相当多且繁复的信息, 利用统计方法对二维数据之间的关联进行分析, 无疑会得到更加全面、客观和有效的信息。故借助于数据挖掘技术对名老中医用药配伍规律进行分析, 可以明晰其中暗含的经验与规律, 实现对名医用药经验的有效总结与传承。目前基于数据挖掘技术对名老中医用药规律的研究日益增多, 本文仅就近 10 年的研究文献进行综述, 希冀对相关研究中不同方法与不同研究对象的适用性方面提供思路。

1 频数分析

频数分析指用统计出单一对象在整体对象中出现的次数以及频率, 并对此进行分析, 得出所需信息的方法。在中医药研究领域这种分析方法多作为一种数据挖掘研究的基础出现, 统计某味药或某几味药在处方中出现的频次, 并挖掘出药物对于总体水平所起的作用大小。在此基础上, 也可取其中高频项目进一步进行关联分析或聚类分析, 从而得出数据之间的关联性 or 每类数据的共同属性特点。

张北华等^[1]利用频数分析方法, 对周仲瑛、蔡淦等 46 位中医药专家治疗肠易激综合征 (IBS) 的医案中症状、治法和用药频次进行了描述性统计, 结果显示, 在 116 例医案中, 针对 IBS 的肝郁病机用药以白芍、柴胡居多, 脾虚病机以白术、茯苓、党参常用。气滞病机以木香、枳壳常用, 湿邪病机以黄连、半夏、薏苡仁常用, 通过数据挖掘所得出的用药频率, 反证出肝郁、脾虚为 IBS 的基本病机。刘文军^[2]对薛伯寿教授用升降散验案的 100 例患者病历进行统计分析, 得出薛教授以升降散调理气机升降出入的遣方特点, 且用姜黄、大黄不在强通其便, 乃为凉降郁热, 引阴下行之意, 旨在和其阴阳, 调其升降的配伍认识。李求兵^[3]对张铁忠教授治疗老年病的处方进行数据挖掘, 根据频数统计结果得出其常用黄芪、山药、苍术、白术、泽泻、茯苓、女贞子等药, 认为老年病多以脾肾虚损

基金项目: 北京中医药大学大学生科研项目质量工程 (201610026021)

* 通讯作者: xuhaotcm@hotmail.com

升降失常所致，故用药多以补脾肾补气阴为主。赵宇明等^[4]基于频数分析和关联分析对王庆国教授 4937 首处方进行了数据挖掘并与刘渡舟教授常用方对比，发现王教授多用柴胡、黄芩、半夏、炙甘草、黄芪、白芍、当归、生晒参等药物，善用柴胡剂与泻心汤类方，说明其传承了刘渡舟教授重视和解少阳枢机的论治理论。田琳等^[5]在对现代中医药治疗头痛自拟处方 171 首的用药频数分析中发现，川芎的使用率高达 96.88%，且常与其他药物配伍用于治疗不同类型的头痛，应和了李东垣所言“头痛须用川芎”的用药经验。许金海等^[6]对施杞教授治疗腰椎间盘突出症典型医案 378 例进行了数据挖掘，根据频次分析结果发现其常用中药为党参、黄芪、白芍、柴胡、川芎、当归、熟地黄等，客观体现了施杞教授以圣愈汤为基础进行加减医治损伤的用药规律，与“以气为主，以血为先”的石氏伤科学术思想一脉相承。高宇^[7]对古今妇科崩漏、痛经、闭经共 890 例名医医案中用药规律进行分析，发现与古代相比，近现代治疗妇科疾病常用药在用药频次上存在差异，体现了近现代对妇科疾病演变规律的进一步认识。

综上，频数分析在研究中具有找出整体中关键项目的作用，尤善于在分析名老中医处方时找出值得关注的几味药物，因此也是分析名老中医用药规律较为适用的分析方法。

2 关联分析

关联分析可以挖掘存在于数据库中的项目或属性间的关系，通过在数据库中利用最小支持度门槛筛选高频项目组，再在高频项目组中由最小信赖度门槛筛选关联规则，其中支持度取决于项目组出现频率，信赖度取决于项目组项目间的一致性。崔松等^[8]同时利用关联分析与频数分析两种方法对何立人教授治疗心悸的 150 例 595 诊次医案进行数据挖掘，将症状与用药进行统计对比，从用药中找出高频项目，再由高频项目中产生关联规则，以此提炼出何立人教授治疗心悸的主要方药以炒党参、灵芝、白果、三七、生地黄、炒白术、丹参、脱力草、苦参等为主，药物频次和高频用药组合结果可体现何立人教授治疗心悸气血兼顾、补泻并施、以和为贵的治疗大法。朱蕾蕾等^[9]利用关联规则和网路图对蒋健教授诊治胃痛有效医案 100 例进行药物分析，结果显示出芍药-甘草和白芷-肉桂等药的强关联，由此得出蒋健教授治疗胃痛的基本方。

唐伟^[10]收集张炳秀教授临证多年治疗胃脘痛的医案，分析处方中所出现的 227 味中药的配伍关联规则，总结出脾胃虚弱是胃脘痛的主要病因，用药多选人参、黄芪、大枣、山药等。将脾胃气虚从各证型中剥离出来，单拟一型，此点与李东垣重补土主张一脉相承。

由上述研究可知，在中医药领域中关联分析方法以频数分析为基础，得出名老中医治疗某一疾病共同出现频率高的两种药物，进而从中分析出现其中一味药物时同时出现另一味药物的条件概率，由此两点作为判断是否能称为药物配伍的统计依据。利用其对项目组在数据库中出现的频次及组内项目出现时另一项目出现的条件概率的筛选，从而能有效判断两项目甚至多项目之间的相关程度，为研究名老中医常用药对、药物配伍和证药关系提供便利。

3 聚类分析

聚类分析又称群分析，它是研究分类问题的一种统计分析方法，同时也是数据挖掘的一个重要算法。聚类分析以相似性为基础，在一个聚类中的模式之间比不在同一聚类中的模式之间具有更多的相似性。依据研究对象的特点，自动对其分类，以减少研究对象的数目，将相近事物归入一类，在此基础上分析每类数据的共同属性或特点。张薇君^[11]采用聚类分析的方法对周仲瑛教授 62 例哮喘缓解期病案进行数据挖掘，分析其药物归类、用药频次和关联程度，结果发现其治疗哮喘以半夏、南沙参为核心药物，以补益肺肾、祛风化痰为治疗原则。伊璠^[12]对张艳教授治疗慢性心力衰竭 186 例患者病案进行分析，在频数分析的基础上加以聚类分析，最后得出 242 首处方中常用的 6 个用药类别，居于前两位的为补气药和活血化瘀药，印证了心气虚、心血瘀阻为慢性心力衰竭的主要病机。谢立栋等^[13]以中医药治疗偏头痛 228 篇相关文献为基础，采用频数分析找出高频药物，聚类分析得到高相关度药物及药对，结果显示治疗偏头痛的主要用药为疏散外风药、活血化瘀药、补虚药、平肝熄风药。

在中医药研究领域常针对治疗某一疾病的处方中的药物进行聚类分析，通过数据库记录的相似性归纳为若干类别，以不可知的分类标准、分类目标进行程序化归类，得出分类后进行总结，发现类别中的共同点，如药物性味、功能主治等，在很大程度上降低了人为分类产生的主观影响，同时能从系统聚类中总结出药物与药物之间未发现或未设想

的共同点,可为药物属性、名老中医用药规律等研究提供思路。

4 复杂网络分析

复杂网络是一种表现为自组织、自相似、吸引子、小世界、无标度等全部或部分特性的网络,其复杂性表现在结构、网络进化、连接多样、节点多样等几个方面,通过不同特性分为无尺度网络、无向网络、无权网络等几类。复杂网络通过其集聚程度以及幂律分布的特点,对其中数据进行分析,可得出数据间的相关性、集聚程度和分布特征等信息。在中医药研究中,进行数据挖掘时多采用此方法作为挖掘名老中医用药中的核心药物和核心药物组合,进而以药物为基础阐述名老中医用药的指导思想。

崔慧娟^[4]基于张代钊教授 145 例 324 诊次的肺癌诊治病例,采用复杂网络中无尺度网络对数据进行挖掘,发现其治疗肺癌在注重益气养阴、软坚散结的基础上,咳嗽重加枇杷叶、川贝母;咯血多加藕节、白及;音哑加玄参、川芎;胸痛加乳香、没药、瓜蒌皮。史成和^[5]利用高忠英教授治疗慢性胃炎的 730 首处方,通过复杂网络分析得出位于核心的药物组合有太子参、白术、天花粉、浙贝母等,由此总结出高教授健脾、润燥的治疗思路。李爱峰^[6]采集聂丽芳教授诊治原发性肾病综合征 161 例患者病历,基于复杂网络进行数据挖掘,结果显示,聂教授治疗原发性肾病综合征组方首选茯苓、芡实、地黄、丹参、金银花、冬瓜皮等药,体现了注重养阴益气、活血利水等治疗方法的综合运用。

以上研究在操作中多将名老中医治疗某一疾病的处方中所有药物作为网络中的节点,以出现在同一处方中作为节点间的链接,进而生成网络进行数据挖掘,得出组方中与其他药关联度最高的几味药,这几味药的产生不仅源自于与其链接的药物的多样化程度,在某种程度上也反映了其自身出现的频次。因此,这些药物不仅代表着名老中医治疗某种疾病时对该疾病患者身体状况及病因的整体认识,同时也反映了其对社会人群体质如“阳常有余,阴常不足”等概括的认识。

5 讨论

数据挖掘技术作为一种以数学方法为基础的统计学研究手段,应用于名老中医用药规律的研究不但极大地提高了研究工作的水平和质量,更能利用

数学方法的特点达到免除研究者的主观影响,进而使研究结果更加符合实际情况。

目前关于名老中医用药规律的数据挖掘方法主要包括频数分析、关联分析、复杂网络分析、聚类分析等。从近 10 年以数据挖掘方法研究名老中医用药规律的文献中所使用的方法与研究对象进行分析,可得出特定方法与特定研究对象之间存在较为强烈的亲和性与适应性,如作为基础的单味药用药频率研究与频数分析,两者皆能作为研究的初步,进而有助于下一步的研究。关联规则与用药配伍研究及药物症状关联研究中,关联规则注重在筛选出高频次的项目后,对项目组合的关联性进行剖析,进而得出项目间的关联,恰恰与药物-药物、药物-症状关联的研究相吻合,关联分析中的筛选高频项目保证了药-药、药-症研究有足够的研究对象和足够的研究必要性,项目组内的条件概率筛选规避了仅从项目组出现频率得出结论所造成的虚假关联现象,保证了研究的准确性。聚类分析中所给出的以随机标准进行的分类,既存在有关药物药性的聚类,也有关于药物对症的聚类,在聚类分析过程中,产出进行研究前所未注意或未知方向的分类,进而为研究提供方向,保证了研究的客观性。因此,各种数据挖掘方法皆能在研究名老中医用药规律的不同方面起到有效的辅助作用。作为数学手段,其严谨性、客观性、宏观性和便捷性是中医药科研中必不可少的特点,不但为研究提供了方法,也为研究的可靠性提供了保障。

但需要提出的是,目前的数据挖掘研究常存在研究与方法相脱离的现象,即研究者只熟悉研究中专业相关的过程与目的,对于以数学为基础的研究方法理解不到位,而造成数据挖掘研究多在数据库构建完成后,直接套用现成的数据挖掘算法程序,在文献中无法给出方法以及方法与目的对应方面较明确的说明。因此,进行数据挖掘研究前应加强相关方法基础知识的掌握,在充分理解数据挖掘方法理论与特点后再进行方法的选择,从而更有效地利用分析工具,使研究结果更加科学有效。

近年来越来越多的中医药研究者致力于利用数据挖掘方法研究中医药尤其是名老中医用药规律,利用科学的研究手段研究传统医学,使传统经验升华为科学规律,势必能使中医药在走向国际化、现代化的道路上迈出新的一步,也能为中医药的发展带来新的血液。

(下转第 900 页)

和外表形态,同时列举了相应的治疗原则,是中医学的精髓,有着悠久的历史,为人民群众所接受,具有良好的群众基础。癌症告知可通过“五态人格测验表”对患者进行分型,对不同类型的患者用不同类型的告知方法,做到因人制宜,以达到患者癌症真实病情告知的最佳效果,最终达到提高癌症患者生存质量的目的。

参考文献

- [1] 曾倩,崔芳芳,宇传华. 等. 中国癌症发病、死亡现状与趋势分析 [J]. 中国卫生统计, 2016, 33(2): 321-322.
- [2] WHO News. The world can not defeat cancer only by treatments [EB/OL]. (2015-01-40) [2016-06-25]. <http://www.who.int/media/centre/news/releases/2014/cancer-report-20140203/zh>.
- [3] 郑雪. 人格心理学 [M]. 广州: 暨南大学出版社, 2007: 9.
- [4] 杨秋莉,徐蕊,于迎,等. 五态人格、体质类型与抑郁症的中医证型的关系探讨 [J]. 中医杂志, 2010, 51(7): 655-657.
- [5] LIN CC, TSAI HF, CHIOU JF. et al. Changes in levels of hope after diagnostic disclosure among Taiwanese patients with cancer [J]. Cancer Nurs, 2003, 26(2): 155-160.
- [6] 郭振洁. 探讨癌症患者家属是否告知患者坏消息的主要因素 [J]. 心理医生, 2015, 21(22): 238-239.
- [7] 郭振洁. 晚期癌症患者病情告知策略研究 [J]. 医学信息, 2014, 27(4): 540.
- [8] 李菁. 我国癌症患者知情同意的现状与实施方式研究 [J]. 经济与社会发展, 2012(11): 69-73.
- [9] 曾铁英,张琪,杨笑梅,等. 癌症患者对重症疾病告知策略要求的调查研究 [J]. 护理研究, 2008, 22(34): 3126-3128.
- [10] 孙菲菲,王娜,苏静,等. 我国癌症患者病情告知的影响因素研究进展 [J]. 中国医药指南, 2016, 14(6): 32-34.
- [11] SULLIVAN RJ, MCNAPACE LW, WHITE RM. Truth-telling and patients diagnoses [J]. J Med Ethics, 2001, 27(3): 192-197.
- [12] FIGG WD, SMITH EK, PRICE DK, et al. Disclosing a diagnosis of cancer: where and how does it occur? [J]. J Clin Oncol, 2010, 28(22): 3630-3635.
- [13] DAUGHERTY CK, HLUBOCKY FJ. What are terminally ill cancer patients told about their expected deaths? A study of cancer physicians' self-reports of prognosis disclosure [J]. J Clin Oncol, 2008, 26(36): 5988-5993.
- [14] 李秀华. 护士在肿瘤病人行使知情权中的作用 [J]. 实用护理杂志, 2002, 18(1): 72-73.
- [15] 魏乾伟,刘宇,陈征. 患者癌症诊断被告知的现状及策略研究 [J]. 中华现代护理杂志, 2014, 20(7): 745-749.
- [16] 肿瘤患者告知与同意的指导原则 [J]. 医学与哲学(临床决策论坛版), 2008, 29(10): 7-8.

(收稿日期: 2017-01-19; 修回日期: 2017-02-25)

编辑: 黄 健

(上接第 888 页)

参考文献

- [1] 张北华,高蕊,李振华,等. 治疗肠易激综合征的专家经验挖掘分析 [J]. 中国中西医结合杂志, 2013, 33(6): 757-760.
- [2] 刘文军. 薛伯寿教授应用升降散方证规律及临床传承研究 [J]. 北京: 中国中医科学院, 2012.
- [3] 李求兵. 张铁忠补益脾肾、升降气机治疗老年病学术思想和经验总结 [J]. 北京: 中国中医科学院, 2012.
- [4] 赵宇明,尤海燕,刘哲,等. 基于数据挖掘的王庆国教授对刘渡舟教授用药传承规律的研究 [J]. 北京中医药大学学报, 2012, 35(5): 293-296, 302.
- [5] 田琳,朱建贵,王映辉,等. 基于数据挖掘的现代中医药治疗头痛自拟处方用药规律分析 [J]. 中国中医药信息杂志, 2012, 19(11): 22-24.
- [6] 许金海,王晶,叶洁,等. 基于数据挖掘的施杞教授治疗腰椎间盘突出症用药经验分析 [J]. 上海中医药大学学报, 2012, 26(6): 56-62.
- [7] 高宇. 基于数据挖掘的古今医案常见妇科疾病研究 [J]. 广州: 广州中医药大学, 2012.
- [8] 崔松,沈梦雯,车立娟. 基于数据挖掘的何立人教授治疗心悸病用药特色分析 [J]. 上海中医药大学学报, 2013, 27(3): 32-35.
- [9] 朱蕾蕾,孙继佳,金彩映,等. 基于数据库及数据挖掘的中医医案方法学研究: 蒋健教授治疗胃痛的临床经验总结 [J]. 中华中医药杂志, 2013, 28(10): 2888-2893.
- [10] 唐伟. 基于数据挖掘的张炳秀教授诊治胃脘痛临床经验研究 [J]. 南京: 南京中医药大学, 2011.
- [11] 张薇君. 基于数据挖掘的周仲瑛教授辨治哮喘缓解期病案的研究 [J]. 南京: 南京中医药大学, 2012.
- [12] 伊璠. 慢性心力衰竭的中医证治规律及临证经验数据挖掘与研究 [J]. 沈阳: 辽宁中医药大学, 2014.
- [13] 谢立栋,李海涛,胡志强. 中医治疗偏头痛用药规律分析 [J]. 湖北中医杂志, 2011, 33(10): 52-53.
- [14] 崔慧娟. 张代钊教授运用中医药治疗肺癌经验的整理和挖掘 [J]. 北京: 中国中医科学院, 2012.
- [15] 史成和. 高忠英教授学术思想与经验总结及运用数据挖掘方法对慢性胃炎用药规律的研究 [J]. 北京: 北京中医药大学, 2011.
- [16] 李爱峰. 基于复杂网络与互信息的聂丽芳教授治疗原发性肾病综合征经验研究 [J]. 北京: 中国中医科学院, 2012.

(收稿日期: 2016-10-27; 修回日期: 2016-12-24)

编辑: 崔京艳