

引用:易天霞,宋程,唐蔚,蒋益兰.数据挖掘技术应用于名老中医经验传承的研究概况[J].湖南中医杂志,2021,37(8):201-204.

数据挖掘技术应用于 名老中医经验传承的研究概况

易天霞¹,宋程²,唐蔚³,蒋益兰³

(1. 湖南中医药大学,湖南长沙,410208;2. 湖南省肿瘤医院,湖南长沙,410013;
3. 湖南省中医药研究院附属医院,湖南长沙,410006)

[关键词] 名老中医;临床经验;传承;数据挖掘;综述;学术性

[中图分类号] R2-031 [文献标识码] A DOI:10.16808/j.cnki.issn1003-7705.2021.08.065

中医药事业的生命力在于继承与创新,名老中医诊疗经验是前人的中医理论融合自己的临床实践,是一笔宝贵的精神财富,继承发展名老中医经验尤为重要。名老中医经验涵盖各个方面,但主要指其独特的学术思想、诊疗经验、治疗方法。目前名医经验主流传承方式有院校传承、师徒传承、成立名老中医工作室、著书立论、临床试验等方法,但这些方式普遍存在着传承效率低、创新性不足、受客观环境影响大、临床效果缺乏评价标准等问题。随着网络技术的快速发展,“大数据”和“精准医疗”的概念相继出现,数据挖掘技术开始应用于中医药事业。

数据挖掘技术是利用网络技术进行有目的、有探索性地获取数据中隐含的规律和知识,也被称为数据库知识发现,是指从大量、不完全、有噪声、模糊、随机的数据中提取隐含在其中的、人们事先不知道但又潜在着有用信息和知识的过程^[1]。本文介绍数据挖掘常用技术及进展,并分析数据挖掘技术局限及将来发展趋势,数据挖掘技术必将进一步推动中医药事业发展。

1 数据挖掘的常用技术

1.1 频数分析 频数分析可得出最常用的数据关系,是使用最广泛的算法。黄玉华等^[2]研究蔡连香教授从“冲任”论治妇科疾病,用频数分析得出高频药物 22 味,前 3 味为菟丝子、当归、白芍。梁秋雨等^[3]通过研究名老中医验方得出脂肪肝验方,用频数分析得到高频药物 26 味。唐晓菊等^[4]研究黄有荣教授治疗腰椎间盘突出经验,用频数分析得出常用药有当归、牛膝、甘草、杜仲等 33 味中药。

1.2 关联规则分析 关联规则分析是一种无向数据挖掘方法。它从大量数据项中寻找有意义的关联关系。李潇然等^[5]研究吕仁和教授治疗慢性肾衰竭经验得出药物核心关

联药对 31 组。王立国等^[6]通过研究 100 则名老中医肺癌的医案,利用关联规则得出医案中肺癌症状、证候、转移部位及年龄的内在联系。马林纳等^[7]研究中医治疗肺纤维化的文献,得到高频关联强度中药组合 6 组。

1.3 聚类分析 聚类分析是一种无监督数据挖掘方法。它基于观测之间的相似度或距离将观测分组。唐香香等^[8]通过对 179 例糖尿病下肢血管病变患者进行中医证候聚类分析,将该病分为 4 型:阴虚火旺、脉络瘀热证,寒湿困脾、气滞血瘀证,肝胃火盛、湿热瘀滞证,气血亏虚、筋脉失养证。刘安平^[9]通过研究刘吉善教授门诊病案,将使用频次大于 500 的前 40 味中药进行聚类分析,得出常用药的 4 类药物组合。陈振东等^[10]探索名老中医治疗反流性食管炎的用药规律,聚类分析得出 9 组常用药物。

1.4 因子分析 因子分析是一种变量简化技术,可用于找出支配相关关系的有限个潜在变量,以及找出关联紧密的药物。綦向军等^[11]研究名医治疗不孕症的验方,将使用频次超过 20 次的中药进行因子分析,提取出 21 个公因子。陈丽平等^[12]通过对 341 例以二陈汤为主方的名老中医医案进行数据挖掘,利用因子分析得到 9 组有意义的因子组,指出二陈汤主治证型为痰浊阻肺、脾虚湿困。李玲等^[13]研究周仲瑛教授诊疗类风湿性关节炎经验,应用因子分析方法得到药物核心组合 8 组。

1.5 贝叶斯网络 贝叶斯网络是一个有向无环图,由代表变量结点及连接这些结点的有向边构成^[14]。贝叶斯网络可以将具体问题中复杂的变量关系在一个网络结构表示,通过网络模型反映问题领域中变量的依赖关系,可用于构建中医专家诊疗系统,研究中医证型与疾病、症状与证候之间关系。如张霆等^[15]分析肺癌的中医证候,对 89 项中医症状

基金项目:中央引导地方科技发展专项资金项目(2017CT5029);湖南省名老中医药专家蒋益兰传承工作室建设项目(湘中医药函[2018]37号);湖南省长沙市科技计划项目(kq1907124,kq1901068)

第一作者:易天霞,女,2019 级硕士研究生,研究方向:中医内科学(肿瘤方向)

通讯作者:蒋益兰,女,主任医师,博士研究生导师,研究方向:中医药防治恶性肿瘤,E-mail:tianshangren624@163.com

通过贝叶斯网络处理,得出36个症状、舌象、脉象之间有较强的关联性,关系最密切的症状组成连续的症状链。甘小金等^[16]采取回顾性研究整理王老治疗子宫内膜异位症的门诊医案150例,运用贝叶斯网络方法对症状-证素信息做分类识别,得出150例医案均存在血瘀证,其中肾虚血瘀、气滞血瘀为最常见的证素组合。徐璿等^[17]收集835例冠心病患者的中医四诊证候信息,在专家辨证和先验知识的基础上,通过贝叶斯网络对四诊证候信息进行分类识别研究,结果显示基于贝叶斯网络建立的融合模型对心气虚、心阳虚、心阴虚、血瘀、痰浊5个证型的识别率,其中阳虚、血瘀识别率最高,为冠心病中医证候分类诊断提供了依据。

1.6 神经网络 神经网络也称作连接模型,是一种进行分布式并行信息处理的算法数学模型,主要通过调整内部大量节点之间相互连接的关系,从而达到处理信息的目的,与贝叶斯网络有相似之处,但神经网络对于数据处理更为灵活。罗悦等^[18]采用BP神经网络构建咳嗽疾病归经预测模型,通过对不同患者咳嗽的症状预测出咳嗽的病位,预测准确率达95.01%。吴燎等^[19]将真实的中医胃病临床数据通过数据清洗、数据编码、BP神经网络模型配置、模型训练,构建出BP网络模型,使每种病型的平均预测率达到90%以上,实现了中医胃病病型的辅助诊断。赵书颖等^[20]基于决策树及神经网络的方法,建立高血压病阴阳两虚证的诊断模型,决策树算法和神经网络建立高血压病阴阳两虚证诊断模型准确率都较高。

1.7 复杂系统熵方法 复杂系统熵方法是一种非监督的算法模式,它能自主地从海量数据中提取信息量最大的组合,特别适用于高度离散性类型的数据^[21]。该方法常用于研究疾病、证候、方剂的相关性,往往能得出一些隐含信息。如彭素娟等^[22]研究袁长津教授治疗咳嗽用药规律,通过无监督的熵层次聚类算法,进一步聚类得到治疗咳嗽的新方11个。刘铁钢等^[23]应用复杂系统熵方法挖掘技术研究孔光一教授治疗月经病的用药经验,得出常用药物核心组合27个,并分析关联最紧密的5个组合。

1.8 支持向量机 支持向量机通过将非线性数据赋予高维特征性,构造出最优分类超平面,可以弥补传统算法过拟合缺陷,主要用于小样本、非线性及高维模式识别的问题解决,对大样本及复杂数据模型处理有一定局限^[24]。张景芳等^[25]利用支持向量机构建了化合物神经毒性判别模型,模型的准确率、灵敏度、特异性均在80%以上,将该模型应用于山豆根神经毒性成分筛查,筛得13个潜在神经毒性化合物,其中4个已有文献验证。吴纯伟等^[26]基于支持向量机的5种大黄苷元治疗脑缺血的配伍研究,采用均匀设计结合SVM的模拟预测方法优选出了5种大黄苷元组分的最佳配比。

2 数据挖掘技术的优势

2.1 研究内容丰富 数据挖掘技术应用于中医药研究二十余年,之前由于缺乏研究数据,对算法操作复杂等原因,数据挖掘技术主要用于研究名老中医用药经验。随着数据

的多样化及算法的简化,研究疾病广泛,研究内容涉及中医、中药、针灸、证候、治则治法等各个方面,研究对象除了经验丰富的临床医家,亦可研究古代典籍,临床文献等,也可以研究将某疾病领域专家用药规律集中起来,研究其中异同。该算法也在不断发展和完善,现在研究往往采用两种以上算法,并根据研究目的灵活组合算法。对于解决问题来说,算法只是提供一种方法,并不是解决某种特定问题的唯一方法,也不存在所谓优方法,大部分算法之间也并不互相排斥^[27]。

2.2 操作简化 随着大量数据的获取,数据重复、冗余,算法操作复杂,非专业人士难以熟练操作等问题,数据挖掘软件便是顺利进行数据挖掘的工具。目前国内有代表性的中医辅助诊疗系统有中医传承辅助平台、中医临床科研信息共享系统、中医诊疗大数据智能分析系统古今医案云平台、中医医案知识服务与共享系统等,这些平台可用于总结名老中医用药经验,发现最佳配伍,研究疾病辨证规律,如证类、证素研究、症状分析、病机研究、病位分析等。根据研究需要选取合适的软件,如中医传承辅助平台主要有频数分析、关联分析、频数分析、熵方法,主要用于研究名老中医临床常用方及挖掘新方,古今医案云平台在病案输入支持手工、语音、OCR、批量医案结构化录入,研究方法包括贝叶斯网络,复杂网络,支持向量机,方剂树,还有案例推理技术,可以快速找到相似医案。

2.3 数据可视化 中医药数据属于知识密集型数据,往往较为复杂抽象,如果能将分析内容“多维度”呈现,则能更好地分析事物之间的关系。偏序结构图不仅可以从宏观共性中发现事物存在的普遍性,还可以从微观个性中发现事物存在的独有性。重要的是偏序结构图可处理的数据范围广、数量大,但形成的结构图并不复杂且无交叉,是表征事物间联系的有效的可视化技术手段^[28]。如张茜^[29]通过研究李可教授书中典型的危重病例,将患者各项信息生成数据库,再使用偏序结构图生成工具将分表格生成属性偏序结构图,根据图中可视化信息研究对象普遍性与特异性之间的关系,挖掘李可医案中的症治规律,分析李可辨治用药的特点。

2.4 与现代实践融合 中医药现代研究成分复杂,往往将药物提纯后研究,但不符合中医的特点,网络药理学基于“疾病-基因-靶点-药物”相互作用网络,通过网络分析,系统综合地观察药物对疾病网络的干预与影响^[30]。它体现的是多分子协同作用,与中医整体观及中药复方相似,故近些年网络药理学研究很多。赵信科等^[31]基于网络药理学分析半夏厚朴汤的重要成分与作用机制,通过筛选共挖掘出半夏厚朴汤主要活性成分44个,其相应的作用靶点218个,并筛选出100个关键作用靶点以及10条重要的信号通路。

3 数据挖掘技术的不足

3.1 数据不够真实准确 大数据时代后,“以数据为导向”是真实世界中医临床科研范式的前提与技术关键^[32]。一方面由于中医临床术语不统一或地域习惯,往往导致数据存

在重复、歧义、缺省、多义、含糊等问题;另一方面数据挖掘处理的对象是随机的交易数据,而挖掘医案都是经过人工挑选的经典案例,都对数据挖掘质量有一定程度的影响。中医药领域数据多为广义小样本数据,研究者多为高校学生,研究多为某一医家诊疗经验,往往会出现临床病历研究中样本数有限,但包含的维度(如使用的中药种类、症状、证候等维度)则可能远大于样本数。这说明经典的数据挖掘算法不一定适用于广义小样本的中医药数据,使用之前需要进行评估^[33]。

3.2 挖掘软件相对局限 田少磊等^[34]使用文献计量学方法研究近些年发表的中医数据挖掘相关文献,得出目前研究方法多为关联规则、聚类分析、神经网络、频数分析、因子分析;传统的数据挖掘方法往往给出的是相关性的大小,而缺乏对因果关系的解释。且目前数据平台也有着规模小、建设重复、缺乏统一标准、检索功能不完善、处理数据方式单一,研究内容同质化等不足,当前数据挖掘技术有待形成共识性操作规范,建议组建多学科团队开展相关工作,完善优化技术方法,制定科学规范的技术标准^[35]。

3.3 结果缺乏深度解读 数据挖掘采用不同的技术往往能得到多样的结果,目前对研究结果的解读往往侧重基础分析,而缺乏深层次的分析。如王静等^[36]认为单一的研究方法存在不能全面真实地反映和传承名老中医学术经验等不足,需将定性访谈、数据挖掘等方法综合运用才能实现对名老中医临床经验的全面继承。对于数据挖掘结果解读不够深刻是个普遍认识,一方面需要从数据挖掘每个阶段把控质量,另一方面应对结果进行临床验证、机制分析等,对研究结果去伪存真。

4 数据挖掘技术的未来趋势

4.1 数据规范化 中医数据存在重复、歧义、缺省、多义、含糊等问题,中医数据还有待细致加工,首先要把控的就是医案文献的医案文献的标准化、规范化,无论是古书典籍还是临床医案,四诊信息、症状指标、患者症状都依赖于医师和患者的主观诊断,且典籍中往往有患者信息缺失,而患者的失访或接受其他治疗均会影响数据采集。另外由于中医数据的多义、含糊,规范中医临床术语标准,利用数据库实现中医临床文献的自动标注和统计分析,数据开放共享,最终会建立一个面向全国开放的中医数据平台,从采集、录入、统计、挖掘各流程确保数据质量。

4.2 系统专业化 数据挖掘系统专业化一直是我们追求的目标,而神经网络、贝叶斯网络、决策树等将目标一步步变为现实。但目前研究范围小,诊断精准度仍不够,更重要的是不具备中医诊疗思维,且中医专家系统依赖于数据挖掘、图像处理、模式识别等方面。李本岳等^[37]提出要构建中医专家系统,应搭载图像识别模块实现“望诊”,搭载语言识别模块实现“闻诊和问诊”,搭载自然语言处理模块实现“问诊”。专家系统未来应在四诊客观化、证候规范化和辨证诊断智能化上多加研究,以期在技术上有更多突破,进而提高

挖掘的效率和准确性。

4.3 数据挖掘与多组学整合研究 数据挖掘与多组学整合研究联合是目前趋势,两者相辅相成。数据挖掘可用于挖掘核心药物组合,多组学整合研究一方面可用于验证数据挖掘结果,一方面可有效预测与分析药物的潜在作用与起效机制,有利于新方新药的研发。现在已有相关理论研究,如樊启猛指出若将贝叶斯网络与超分子化学结合,使超分子化学解决中医药理论研究中的定性问题,而贝叶斯网络则解决其中的定量问题^[38]。

5 小结

中医是一个数不尽的宝库,还有许多值得研究的东西,如性味归经、三焦辨证等,计算机技术必然会让中医药焕发新的光彩。总之,机遇与挑战并存,数据挖掘的结果要能够指导临床实践,而临床经验也能通过数据挖掘系统而全面地展现,最终有望挖掘出可以超越个体的、适用于中医临床的优秀群体经验。

参考文献

- [1] 周光华,李岳峰.数据挖掘技术在卫生统计信息工作中的应用研究[J].中国卫生信息管理杂志,2012,9(6):82-86.
- [2] 黄玉华,李国白,杨阳,等.蔡连香从“冲任”论治妇科疾病辨证用药规律研究[J].中国中医药信息杂志,2017,24(3):91-95.
- [3] 梁秋雨,李瑞敏,相安,等.基于名老中医验方数据分析谈脂肪肝辨治思路与方法[J].四川中医,2017,35(9):23-25.
- [4] 唐晓菊,张磊,刘汝专,等.基于数据挖掘的黄有荣教授治疗腰椎间盘突出症用药规律分析[J].中国中医急症,2020,29(2):221-225.
- [5] 李潇然,王世东,李佳玥,等.基于关联规则算法和复杂系统熵聚类的吕仁和教授治疗慢性肾衰竭经验研究[J].现代中医临床,2016,23(3):9-13.
- [6] 王立国,邵爱贤.基于关联规则的名老中医医案中肺癌转移特点的数据挖掘分析[J].中医临床研究,2016,8(13):5-7.
- [7] 马林纳,王瑞,苗明三.基于数据挖掘的治疗肺纤维化中药用药规律分析[J].中草药,2020,51(6):1406-1411.
- [8] 唐香香,牟新,付国春.179例糖尿病下肢血管病变患者中医证候聚类分析[J].浙江中西医结合杂志,2020,30(4):333-335.
- [9] 刘安平,刘吉善,沈绍武,等.基于名老中医刘吉善门诊病案的数据挖掘分析[J].湖南中医药大学学报,2020,40(1):75-79.
- [10] 陈振东,安彬,田旭东.基于数据挖掘探索名老中医治疗反流性食管炎的用药规律[J].国医论坛,2020,35(2):27-29.
- [11] 葛向军,张兆萍,章浩展,等.基于数据挖掘的名老中医治疗不孕的用药规律分析[J].中国中药杂志,2019,44(7):1491-1496.
- [12] 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.
- [13] 李玲,周学平,李国春,等.运用数据挖掘技术探讨周仲瑛教授诊疗类风湿性关节炎经验[J].中国中西医结合杂志,2013,33(12):1705-1707.
- [14] 吴嘉瑞,唐仕欢,郭位先,等.基于数据挖掘的名老中医经验传承研究述评[J].中国中药杂志,2014,39(4):614-617.

- [15] 张霆,陈波,徐涛,等.基于贝叶斯网络的肺癌证候研究[J].中国中医药科技,2014,21(6):599-600,603.
- [16] 甘小金,陈艳,马秀丽.基于贝叶斯网络的王子瑜教授治疗子宫内异位症的辨证规律研究[J].世界中西医结合杂志,2019,14(10):1350-1352.
- [17] 徐瑾,许朝霞,许文杰,等.基于贝叶斯网络原理的835例冠心病病例中医证候分类研究[J].上海中医药杂志,2014,48(1):10-13.
- [18] 罗悦,杨威龙,周紫仪,等.基于BP神经网络的咳嗽疾病归经研究[J].时珍国医国药,2019,30(12):3048-3050.
- [19] 吴燎,陈思珏.BP神经网络在中医胃病病型诊断中的实现[J].科技风,2019(20):244.
- [20] 赵书颖,张新雅,李运伦.基于决策树及神经网络的高血压病阴阳两虚证诊断模型的研究[J].中华中医药学刊,2019,37(5):1120-1123,1288-1291.
- [21] 唐仕欢,陈建新,杨洪军,等.基于复杂系统熵聚类方法的中药新药处方发现研究思路[J].世界科学技术—中医药现代化,2009,11(2):225-228.
- [22] 彭素娟,吴彬才,杨柳,等.基于数据挖掘方法分析袁长津教授治疗咳嗽用药规律研究[J].中医药导报,2018,24(7):51-55.
- [23] 刘铁钢,陈建新,张望,等.复杂系统熵方法探讨孔光一教授治疗月经病经验[J].北京中医药大学学报,2013,36(5):341-344,363.
- [24] 许文杰,刘攀,燕海霞,等.528例冠心病患者中医脉象非线性动力学特征在证候诊断模型中的应用[J].中华中医药杂志,2014,29(5):1661-1665.
- [25] 张景芳,蒋芦荻,张燕玲.支持向量机在中药神经毒性成分筛查中的应用[J].中国中药杂志,2014,39(17):3330-3334.
- [26] 吴纯伟,宿爱山,陈超,等.基于支持向量机的5种大黄芩元治疗脑缺血的配伍研究[J].中草药,2015,46(22):3377-3381.
- [27] 高静,程铭.粗糙集理论辅助现代医疗诊断研究综述[J].科技与创新,2019(11):25-27,29.
- [28] 路双双.基于偏序结构表示原理的乐曲结构可视化及分类的研究[D].秦皇岛:燕山大学,2019.
- [29] 张茜.基于属性偏序结构图数据挖掘方法对李可急危重症疑难病医案的知识发现[D].广州:广州中医药大学,2018.
- [30] 陈海彬,周红光,李文婷,等.网络药理学——中药复方作用机制研究新视角[J].中华中医药杂志,2019,34(7):2873-2876.
- [31] 赵信科,孙少伯,刘凯,等.基于网络药理学探讨半夏厚朴汤作用机制[J].中医学报,2020,35(2):379-383.
- [32] 刘保延.真实世界的中医临床科研范式[J].中医杂志,2013,54(6):451-455.
- [33] 朱彦,朱玲,崔蒙.论中医药数据挖掘[J].中华中医药杂志,2016,31(8):2932-2935.
- [34] 田少磊,李敬华,童元元,等.基于文献计量学的中医遣方用药经验数据挖掘研究现状分析[J].中国中医药信息杂志,2020,27(5):97-101.
- [35] 刘兴方,韩学杰.数据挖掘技术在医案研究中的应用与讨论[J].中国实验方剂学杂志,2014,20(9):247-250.
- [36] 王静,夏桂成.名老中医学术思想和临证经验传承研究初探[J].江苏中医药,2015,47(8):18-20.
- [37] 李本岳,李伟荣,潘海峰,等.浅谈人工智能对中医诊断的影响[J].世界科学技术—中医药现代化,2020,22(5):1624-1628.
- [38] 樊启猛,贺玉婷,李海英,等.贝叶斯网络在中医药理研究的现状及与超分子化学联合应用前景[J].湖南中医药大学学报,2019,39(11):1407-1411.

(收稿日期:2020-12-01)

中暑后的急救误区

错误一:掐人中 中暑的根本原因是人体产热和散热功能严重失调,因此首先要降温、脱离热源。掐人中仅仅能起到刺激疼痛的作用,不能用来急救。尤其对于昏迷患者,掐人中时可能因为疼痛刺激而引起误吸窒息,反而存在病情加重的潜在风险。

错误二:猛喝凉水 中暑后需要合理补水,但不可大量喝凉水。大量出汗伴有电解质流失,如果只是大量喝凉水,会导致人体出现稀释性低钠、氯血症,即俗称的“水中毒”。此时喝水要少量、多次,适当加些食盐,会有更好的效果。

错误三:服藿香正气水 很多人认为藿香正气水是治疗中暑的“神药”,其实这是片面的。不推荐在中暑后使用藿香正气水的主要原因有二:首先,藿香正气水主要应用于解表化湿、理气和中,用于内伤湿滞或夏伤暑湿所致的“阴暑”,与现代医学的中暑(热射病)不同;其次,藿香正气水的乙醇含量为40%~50%,服用后可能引起脱水,中暑损伤会更严重。

错误四:服退热药 中暑发热的机制与细菌或病毒感染引起的发热不同。中暑导致的发热是因为人体的散热中枢受阻,造成的热量在体内积聚,不能通过出汗来散发热量。而退热药主要通过让身体发汗来达到降温目的,中暑后身体本来就缺水,再服用退热药迫汗外出则导致身体更加缺水,适得其反。对于中暑后的高热,一般情况下应首先选择传统物理方法,效果显著且安全。(http://www.cnctm.com.cn/news.html?aid=177574)