

数据挖掘算法在中药研究中的应用^Δ

吴地尧^{*},章新友[#],甘宇汾,于思婷(江西中医药大学药学院,南昌 330004)

中图分类号 R28 文献标志码 A 文章编号 1001-0408(2018)19-2717-06
DOI 10.6039/j.issn.1001-0408.2018.19.28

摘要 目的:为数据挖掘算法在中药研究中的进一步应用提供参考。方法:以“中药”“Apriori”“FP-growth”“层次聚类”“熵聚类”“决策树”“随机森林”“贝叶斯”“支持向量机”“人工神经网络”“logistic 回归”“线性回归”等为关键词,组合查询2000年1月—2018年5月发表于中国知网、万方数据、维普网相关文献,对数据挖掘算法在中药研究各子领域中的应用现状进行综述。结果:共检索到相关有效文献573篇。数据挖掘算法较常应用在方剂配伍规律、药物分析、中药药性研究、制剂工艺研究等中药研究子领域,但在医案研究、谱效关系、量化诊断标准等子领域的应用较少。在各子领域中,以在方剂配伍规律研究中运用的数据挖掘算法种类最多,包括Apriori、FP-growth、层次聚类、熵聚类、决策树、人工神经网络、贝叶斯分类、logistic 回归等,并以人工神经网络和支持向量机两种数据挖掘算法在中药研究各子领域中应用最广。结论:数据挖掘算法在中药研究各子领域中应用广泛,可为中药现代化研究提供有力的技术支持。

关键词 数据挖掘算法;中药研究;方剂配伍规律;药物分析;子领域

伴随着中药领域数据的暴涨^[1],中药数据挖掘应运而生。中药数据挖掘是在中医药理论指导下,对中药新药、中药组方规律、作用机制、有效成分和组效关系等多个方面进行深入挖掘的研究^[2]。数据挖掘算法是根据数据的需要,创建数据挖掘模型的一系列探索和计算的

方法。探讨数据挖掘算法在中药研究中的应用现状既可以直观地阐述中药研究领域各数据挖掘算法的使用现状,也能更详尽地了解数据挖掘算法应用的特点及领域,拓展数据挖掘算法在中药研究中的应用,为各算法的深入研究提供参考和借鉴。

异性分析[J].中国卫生事业管理,2016,33(8):575-577.

[2] 马彬,杨克虎,刘雅莉,等.英国医疗风险监管体系的循证评价及其对我国医疗风险管理的启示:关注病人安全,预防医疗差错,提高医疗质量[J].中国循证医学杂志,2006,6(7):514-522.

[3] LEAPE LL, BATES DW, CULLEN DJ, et al. Systems analysis of adverse drug events[J]. JAMA, 1995, 274(1): 35-43.

[4] RICHARD NK, STEVEN DW, JONATHAN C, et al. Prevalence and nature of medication administration errors in health care settings: a systematic review of direct observational evidence[J]. Annals of Pharmacotherapy, 2013, 47(2):237-256.

[5] 陈禾凤,许倍铭,徐蕾,等.医务人员对药品零加成政策认知情况的调查与分析[J].中国药房,2015,26(10): 1431-1434.

[6] 王存伟.自动发药机在我院门诊药房实践与体会[J].药学研究,2017,36(3):177-179.

[7] 李基,邵琼玲,王盛军.基于STM32与串口服务器的IP语音通信系统[J].计算机测量与控制,2017,25(12): 172-176.

[8] 鱼兆虎,王永艳,柴文政.排队叫号系统与HIS系统接口方式的探讨[J].电脑知识与技术,2015,11(11):97-98、100.

[9] 欧微,程岚,陈圣荣.基于C#的数据库表结构报表系统的设计与实现[J].微型机与应用,2013(21):7-9.

[10] WICKENS CD. Engineering psychology and human performance[M]. New York: Charles E.Merrill Publishing Company, 1984:121-123.

[11] SAMARANAYAKE NR, CHEUNG ST, CHENG K, et al. Implementing a bar code assisted medication administration system: effects on the dispensing process and user perceptions[J]. Int J Med Inform, 2014, 83(6):450-458.

[12] 徐帆,徐贵丽,夏伟,等.医疗机构门诊用药溯源管理系统的设计与应用[J].中国药房,2017,28(10):1379-1382.

[13] 任正,李茜,顾中盛,等. RFID医药周转箱在我院住院药房的应用[J].中国药房,2017,28(19):2733-2736.

[14] 杨文字,陈磊,余靓平,等.全自动单剂量摆药机实施效果评价[J].药学实践杂志,2013,31(3):231-234.

[15] 焦蕾,胡扬,张岩,等.我院形似音似药品的安全管理模式和成效[J].中国药房,2015,26(4):509-512.

Δ 基金项目:国家自然科学基金资助项目(No.81660727)
* 讲师,主管药师,博士研究生。研究方向:中药信息管理与应用。电话:0791-87118630。E-mail:53792068@qq.com
通信作者:教授,博士生导师。研究方向:中医药信息管理。电话:0791-87118630。E-mail:xinyouzhang@163.com

(收稿日期:2018-04-24 修回日期:2018-07-02)
(编辑:刘 萍)

以数据挖掘任务为分类依据,对常用数据挖掘算法进行逐个统计和应用分析,数据挖掘能完成的任务总体可概括为两类——描述性任务和预测性任务^[2]。描述性任务即刻刻画目标数据中数据的一般性质;预测性任务即在当前数据上进行归纳从而做出预测。这两类任务可划分为六类:特征化与区分、关联规则分析、分类分析、回归分析、聚类分析、离群点检测分析^[3]。由于特征化与区分任务主要用于建立数据库及数据预处理,离群点检测任务大部分是由聚类分析方法完成^[2],故在本文不对这两种任务做详细介绍。完成中药数据挖掘任务常用数据挖掘算法见图1。

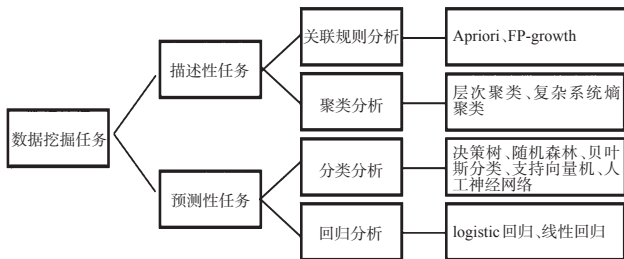


图1 完成中药数据挖掘任务常用数据挖掘算法

在本研究中,笔者以“中药”“Apriori”“FP-growth”“层次聚类”“熵聚类”“决策树”“随机森林”“贝叶斯”“支持向量机”“人工神经网络”“logistic回归”“线性回归”等为关键词,组合查询中国知网、万方数据、维普网等数据库中于2000年1月—2018年5月期间发表的数据挖掘算法在中药研究领域应用的相关有效文献,其中检索的算法名称根据 *Data mining: concepts and techniques* (第3版)一书以及相关文献^[4]总结得出。以数据挖掘算法分类建立11个文献库,再将单个文献库里的文献按照数据挖掘算法应用的中药研究子领域进行分类(该算法应用于中药研究同一领域的文献超过2篇及以上的,则可归入某子领域中,若同一领域只有1篇文献的则不单独分类讨论,合并归类在“其他”项里)。由于部分文献涉及到2个以上的挖掘算法,如一篇文献同时运用到决策树和随机森林两种算法,则该文献既归类到决策树文献库中,也归类到随机森林文献库中。最后,笔者根据检索的文献结果,按照算法应用的中药研究子领域进行分类,然后结合各算法的特点,综合探讨完成关联规则分析、分类分析、回归分析、聚类分析四类任务的常用数据挖掘算法在中药研究中的应用现状。

1 文献检索结果

共检索到相关有效文献573篇,纳入到11个文献库中,数据挖掘算法有效文献检索结果详见表1。

2 中药研究子领域分类及运用的数据挖掘算法

中药研究子领域有方剂配伍规律、药物分析、中药药性研究等,其具体分类详见表2。

由表2可知,数据挖掘算法主要应用在方剂配伍规律、药物分析、中药药性研究等中药研究子领域,在医案

表1 数据挖掘算法有效文献的检索结果

算法名称(文献库)	文献篇数
Apriori	39
FP-growth	14
层次聚类	28
复杂系统熵聚类	113
决策树	44
随机森林	22
贝叶斯分类	34
支持向量机	87
人工神经网络	129
logistic回归	51
线性回归	46
有效文献	573

注:有部分文献涉及2个及以上算法,因此会出现重复归类现象

表2 中药研究各子领域中运用的数据挖掘算法

中药研究子领域	文献篇数	数据挖掘算法
方剂配伍规律	189	Apriori、FP-growth、层次聚类、熵聚类、决策树、人工神经网络、贝叶斯分类、logistic回归
药物分析	83	层次聚类、随机森林、决策树、人工神经网络、线性回归、支持向量机
中药药性研究	48	Apriori、决策树、随机森林、人工神经网络、贝叶斯分类、支持向量机
制剂工艺研究	36	支持向量机、人工神经网络
中药安全性评价	25	决策树、随机森林、logistic回归、支持向量机
中药功效研究	20	贝叶斯分类、支持向量机、人工神经网络
中药鉴定	14	支持向量机、人工神经网络
药物疗效评价	13	logistic回归、决策树、贝叶斯分类
中药活性成分研究	10	支持向量机、人工神经网络
医案研究	4	Apriori
谱效关系	3	线性回归
量化诊断标准	3	logistic回归
专利研究	3	层次聚类
药物靶点预测	3	随机森林
药物成本-效果评价	3	贝叶斯分类
药物剂量研究	3	线性回归
药房调剂研究	2	FP-growth
其他	145	

注:有部分文献涉及2个及以上算法,因此会出现重复归类现象

研究、谱效关系、量化诊断标准等子领域的应用较少。在方剂配伍规律的研究中运用到的数据挖掘算法种类最多,人工神经网络和支持向量机算法应用的范围最广,适用于多个中药研究子领域。

3 应用于中药研究各子领域中的数据挖掘算法

3.1 关联规则分析

关联规则分析主要用于发现隐藏在大型数据集中有意义的联系。其主要包含两个阶段:第一阶段是从数据集中找出所有的频繁项集,第二阶段是由频繁项集产生关联规则^[5]。在中药研究领域,完成关联规则分析任务常用的算法有Apriori算法、FP-growth算法。

3.1.1 Apriori算法 Apriori算法是发现频繁项集的常用基本算法,该算法使用逐层搜索的迭代方法扫描数据库,收集满足最小支持度的项,找出频繁项集,并直接产生强关联规则^[6]。由检索的39篇相关文献可知,Apriori算法主要应用于方剂配伍规律、医案研究、中药药性研究等几个研究子领域。(1)方剂配伍规律研究。如针对

某具体疾病收集临床常用处方、名医名方或经典古方,经规范化处理后建立数据库,利用Apriori算法挖掘出数据库中的核心药对、高频药物组合,结合药物功效特点,在中医理论指导下阐述组方配伍规律,其分析结果对临床用药或者新药开发均有重要的指导意义^[6]。(2)医案研究。如利用Apriori算法对典型病例的中医医案中的理法方药进行关联规则挖掘,为中医临床治疗、中医药教学及中成药的研制提供参考^[7]。(3)中药药性研究。如利用Apriori算法挖掘出药性与功效间的有机联系,用现代科学理论阐释中药药性理论,促进中药现代化发展^[8]。

3.1.2 FP-growth算法 FP-growth算法是Apriori算法的优化处理,其解决了Apriori算法在运行过程中会产生大量候选集的问题^[3],且在大规模方剂数据集筛查挖掘中具有较强的泛化性和“鲁棒性”^[9],更适合用于大样本的方剂数据库挖掘。

由检索的14篇相关文献可知,FP-growth算法主要运用于药房调剂研究、方剂配伍规律研究。(1)药房调剂研究。如利用FP-growth算法统计常见的合并发放的药物、小包装中药饮片的使用原则以及优化药物仓位配置管理,从而提高药房调剂处方的效率及准确度^[10]。(2)方剂配伍规律研究。如利用FP-growth算法挖掘数据集中的频繁项集,从而找出方剂库中的核心药对、高频药物组合,从科学的角度阐明方剂配伍规律^[11]。

3.2 聚类分析

聚类分析是把数据集划分成若干子数据集的过程,使得数据集中的对象彼此相似,且区别于其他数据集中的对象。聚类分析主要包括划分聚类、层次聚类、基于密度的聚类分析、基于网格的聚类分析、复杂系统熵聚类^[12]。由于划分聚类、基于密度的聚类分析、基于网格的聚类分析的应用报道较少,因此本文着重探讨复杂系统熵聚类及层次聚类的应用。

3.2.1 复杂系统熵聚类 复杂系统熵聚类是基于香农提出的信息熵理论,通过计算某变量与其他变量间的关联度系数来判断变量之间是否相关^[12]。该算法也是一种非监督的模式发现算法,通过计算熵值来完成聚类任务,特别适合具有高度离散性的中药方剂数据。由检索的113篇相关文献可知,复杂系统熵聚类主要运用中医传承辅助平台软件进行挖掘分析,利用复杂系统熵聚类的算法挖掘方剂核心组合,分析方剂配伍规律,也可作为发现中药新药处方的挖掘工具^[13]。

3.2.2 层次聚类 层次聚类即把数据划分成不同层次上的组群。由检索的28篇相关文献可知,层次聚类算法主要运用于方剂配伍规律、药物分析、专利研究等几个中药研究子领域。(1)方剂配伍规律。如利用层次聚类算法可形成单一病证以及单一方剂下的药物组合,或者根据性味形成聚类,在此基础上继续进行关联规则分析

挖掘药物的核心组合^[14]。(2)药物分析。如利用层次聚类方法可对中药的多种活性成分或者分子骨架进行合理分类,分析化学成分的远近亲疏关系^[15]。(3)专利研究。如利用层次聚类方法可揭示中药复方专利发展的规律及专利如何驱动行业发展的内在因素^[16]。

3.3 分类分析

分类分析是一种重要的预测数据的分析形式,包括两个阶段:构建模型的学习阶段和将构建的模型用于预测数据特性的分类阶段^[3]。完成分类分析任务的常用算法有决策树、贝叶斯分类、随机森林、支持向量机和人工神经网络。

3.3.1 决策树 决策树是一种类似流程图的树结构,在已知各种情况发生概率的基础上,通过构成决策树来求取净现值的期望值 ≥ 0 的概率,评价项目风险,判断可行性进行决策分析^[3]。由检索的44篇相关文献可知,决策树算法主要应用于中药药性研究、中药安全性评价、药物疗效评价、药物分析和方剂配伍规律等几个中药研究子领域。(1)中药药性研究。如总结已知药物的性效规律后,再利用决策树算法来预测未知或缺失的药性,或利用决策树研究药性规律、性效关系^[17]。(2)中药安全性评价。如将流行病学指标、用药相关因素或者中药化学成分等作为决策树分析的变量,用于预测不良反应发生的概率或者药物是否有脏器毒性^[18]。(3)药物疗效评价。如以患者中医四诊信息和疾病诊断指标为基础,利用决策树模型探索临床检测指标、疾病信息与诊疗效果间的关系,可为临床治疗提供参考^[19-20]。(4)药物分析。如利用决策树建立挖掘模型,预测对指纹图谱评价或定量分析影响较大的特征参数,或者利用决策树建立指纹图谱的多维多息特征的数字化评价系统,从而指导优化试验操作条件,提高药物分析结果的准确性^[21]。

3.3.2 随机森林 随机森林是利用“多棵树”对样本进行训练并预测的一种分类器,其基本单元为决策树。从直观角度来解释,每棵决策树为一个分类器,针对输入的样本给出相应的分类选择,并进行投票。随机森林集成所有的分类投票结果,将投票次数最多的类别指定为最终的输出结果^[22]。由检索的22篇相关文献可知,随机森林算法主要应用于药物靶点预测、药物分析、中药药性研究和中药安全性评价等几个中药研究子领域。(1)药物靶点预测。如利用一系列蛋白质数据库建立药物成分-靶点的随机森林模型,用于预测和识别中药有效成分的作用靶点,该法相比传统的药理试验方法,模型识别不仅效率高且操作简便^[23]。(2)药物分析、中药药性研究、中药安全性评价。这3个子领域的随机森林应用过程皆与决策树类似^[24-26]。但随机森林建立的模型泛化能力更强,实用性更高^[27]。

3.3.3 贝叶斯分类 贝叶斯分类是基于贝叶斯定理的

一种统计学分类方法,即将给出的待分类项,利用贝叶斯定理理解此项在不同类别中出现的概率,将概率最大的项定义为此待分类项所属的类别^[3]。由检索的34篇相关文献可知,贝叶斯分类算法主要应用于中药药性研究、方剂配伍规律、中药功效研究、药物疗效评价、药物成本-效果评价等几个中药研究子领域。(1)中药药性研究。如将收集的临床药理数据、生理生化指标、基原要素等药物相关信息作为网络节点(也称变量)来建立网络拓扑图和条件概率表,以此预测中药或其组分的四气五味以及功效^[28]。(2)方剂配伍规律。如收集方剂、证候、药味的出现次数、有效率等关键数据并进行统计,然后建立贝叶斯分类预测药物疗效,或是与其他算法结合研究方剂组方规律,分析“病-证-方-药”的关系^[29]。(3)中药功效研究。如将收集的药理药效相关指标,建立贝叶斯模型来预测中药及组分的功效^[30]。(4)药物疗效评价。如在贝叶斯理论上建立网状Meta分析,收集药物基本功效、适应证、安全性、评价现状等指标,对药物疗效、安全性进行复杂的综合比较分析^[31]。(5)药物成本-效果评价。如在药物经济学理论的基础上设计贝叶斯混合处理比较法,为临床合理用药提供决策依据,并为药物经济学评价提供方法上的参考^[32]。

3.3.4 支持向量机 支持向量机是一种二分类模型,旨在从样本中找到一个支持向量,构建出最好的分类超平面对样本进行分割,以实现分割间隔最大化^[33]。由检索的87篇相关文献可知,支持向量机算法主要应用于药物分析、制剂工艺研究、中药药性研究、中药鉴定、中药活性成分研究、中药安全性评价、中药功效研究等几个中药研究子领域。(1)药物分析。如利用支持向量机与红外光谱结合建立快速药物检测模型,为产品质量的实时监测与质量控制提供新方法^[34]。(2)制剂工艺研究。如利用支持向量机为药物提取时间、溶剂量、提取率等制剂条件建立预测模型,从而得到最优工艺参数^[35]。(3)中药药性研究。如以元素或化学成分的含量统计结果作为药性分类的特征指标,利用支持向量机建立药性识别模型,从而深入阐明药性的内在机理^[36]。(4)中药鉴定。如将支持向量机与光谱技术结合,对中药进行无损快速鉴别,提高了中药鉴别的速率^[37]。(5)中药活性成分研究。如利用支持向量构建化学成分与药效关系,建立组效关系模型,从而准确地预测化合物的活性,所得结果对新药的深入研究具有极其重要的意义^[38]。(6)中药安全性研究。如基于毒性标志物或相关物理化学性质建立毒性判别的支持向量机模型,可为中药安全性的研究提供新的方法^[39]。(7)中药功效研究。如利用支持向量机建立功效分类模型,对不同的复方组合进行药效预测,可为中药复方的二次开发提供参考^[40]。

3.3.5 人工神经网络 人工神经网络是根据生物学中

神经网络的基本原理,以网络拓扑知识为理论基础,模拟人脑的神经系统对复杂信息处理机制的一种数学模型^[41]。由检索的129篇相关文献可知,人工神经网络主要应用于方剂配伍规律、药物分析、制剂工艺研究、中药功效研究、中药鉴定、中药药性研究、中药活性成分研究等几个中药研究子领域。(1)方剂配伍规律。如利用人工神经网络将不同配比的药方与药效指标建立关联预测,优选出最佳疗效的组分配比药方^[42]。(2)药物分析。如利用人工神经网络或与红外光谱结合建立快速药物检测模型,为产品质量的实时监测与质量控制提供新方法^[43]。(3)制剂工艺研究。如以药物提取时间、溶剂量、提取率等制剂条件为基础,建立预测模型,利用人工神经网络优化工艺参数^[44]。(4)中药功效研究。如将收集的化合物、药性、药理作用等指标,利用人工神经网络对新化合物或未知成分的药效进行预测分类^[45]。(5)中药药性研究。如将收集的药物相关信息,利用人工神经网络建立药性识别模型从而实现对未知药物药性的预测^[46]。(6)中药鉴定。如将人工神经网络与光谱技术结合,实现对中药的快速鉴别^[47]。(7)中药活性成分研究。如利用人工神经网络构建化学成分与药效关系模型,从而准确预测化合物的活性,促进新药的发展^[48]。在药物分析、中药药性研究等几个研究子领域中,人工神经网络的运用方法、步骤,目的均和支持向量机类似。综合两种算法的特点可知,分析结果输出为多分类的研究适合采用人工神经网络算法来解决,而多变量小样本的二分类研究适合采用支持向量机算法解决。

3.4 回归分析

分类和回归是预测问题的两种主要类型,其区别在于输出变量类型的不同。分类是定性输出,用于预测离散变量;回归是定量输出,用于预测连续变量。用于完成回归任务的常用算法有线性回归、logistic回归。

3.4.1 线性回归 线性回归是利用数理统计中回归分析,来确定两种或两种以上变量间相互依赖的定量关系的一种统计分析方法。由检索的46篇相关文献可知,线性回归算法主要应用于药物分析、谱效关系、药物剂量研究等几个中药研究子领域。(1)药物分析。如利用回归算法建立回归方程获得相关系数和线形图,从而反映药物浓度与仪器响应值的关系,主要用于验证分析方法的准确性^[49]。(2)谱效关系。如利用回归算法研究指纹图谱中所含信息与药效之间的关系,揭示其相关性^[50-51]。(3)药物剂量研究。如利用回归算法优化处方中每味药物的剂量配比而剂量研究是处方配伍研究中的重要内容^[52]。

3.4.2 logistic回归 与线性回归不同,logistic回归的因变量是分类变量(二分类或者多分类),其主要用来解决分类问题^[3]。由检索的51篇相关文献可知,logistic回归

算法主要应用于方剂配伍规律、药物安全性评价、药物疗效评价、量化诊断标准等几个中药研究子领域。(1)方剂配伍规律。如文献中的研究大多是利用logistic回归算法对疾病分型用药建立模型,或是统计药物性味归经等因素建立多元回归模型,为医师用药提供参考依据^[53]。(2)中药安全性评价。如利用logistic回归算法分析中药诱发不良反应或毒性的相关因素及发生规律,促进临床合理用药^[54]。(3)药物疗效评价。如利用logistic回归分析影响疗效、并发症或是依从性的相关因素^[55]。(4)量化诊断标准。如统计患者的中西医临床指标、生理相关因素,再利用logistic回归分析建立疾病证型分类模型用于疾病的预测^[56]。

4 讨论

通过分析研究发现,在中药研究领域,数据挖掘算法的运用种类较多,研究范围涵盖方剂配伍规律、药物分析等十几个研究子领域。尤其在方剂配伍规律分析中,挖掘算法的运用相当成熟,已形成标准模式化操作,且邹锦等^[57]和董俊龙^[58]的研究也报道运用多种算法相结合对方剂配伍规律进行深入挖掘,促进了中药新处方的发现,为新药研发提供了新的手段。在药物剂量研究、中药活性成分研究、药物疗效评价等几个子领域中,数据挖掘算法也逐步开展,但相关应用报道较少,一般更倾向于用传统的临床试验或是药物试验作为研究分析手段。但是,仅靠烦琐复杂的试验去解决问题不仅耗材、耗力、效率低,且试验得出的大量数据也无法得到有效处理。笔者建议将数据挖掘方法与传统试验相结合,用人工智能算法得出的结果指导药物试验,从而减少不必要的损耗或者避免处理繁冗的数据。

总体而言,目前结合两种以上算法的应用报道较少,只出现于方剂配伍规律、药物分析、中药药性研究等几个子领域中。若能在处理中药数据时充分结合各算法的优点,促进数据挖掘技术在中药研究的应用。另外,在中药鉴定和制剂工艺研究方面,建议利用支持向量机和人工神经网络来对多个药物属性或工艺条件进行分类预测。

综上,本文研究了各数据挖掘算法在中药各子领域中的应用,具有一定的参考价值。建议在今后的研究中,拓展数据挖掘算法的应用领域,推动子领域内挖掘算法的应用,提升结果可信度,促进数据挖掘在中药研究中的应用,为中药现代化研究提供有力的技术支持。

参考文献

[1] 佟旭,谢晴宇,孟庆刚.论大数据时代背景下中医药数据集成分析的科学价值[J].中国中医药信息杂志,2015,22(8):1-3.

[2] 张博.基于关联规则的数据挖掘技术在中药方剂配伍中的应用研究[J].兰州文理学院学报(自然科学版),2011,25(1):82-86.

[3] HAN JIAWEI, KAMBER MICHELINE, PEI JIAN. *Data mining: concepts and techniques*[M]. 3rd edition.北京:机械工业出版社,2017:10-320.

[4] 邹志文,朱金伟.数据挖掘算法研究与综述[J].计算机工程与设计,2005,26(9):2304-2307.

[5] 周牒岚,陈琳,向华.数据挖掘算法研究[J].现代电子技术,2011,34(20):75-78.

[6] 潘碧琦,潘建科,刘军,等.基于关联规则和复杂系统熵聚类的痛风用药规律研究[J].中华中医药杂志,2014,29(6):2040-2043.

[7] 袁锋.基于数据挖掘的中医医案分析系统的设计与实现[D].济南:山东师范大学,2006.

[8] 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].结合医学学报,2011,9(7):794-803.

[9] 李德琳,魏本征,张诏,等.基于FP-growth算法的中医抗病毒方剂配伍规律探索[J].中华中医药学刊,2018,36(3):663-668.

[10] 刘闽碧.基于FP-growth算法的中药配方数据挖掘[J].医学信息,2009,22(12):2629-2631.

[11] 肖汉杰.基于数据挖掘的药材并发关系和调剂效率提升研究[D].昆明:昆明理工大学,2013.

[12] 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.

[13] 鲁可,陈腾宇,黄慈辉,等.基于中医传承辅助平台软件的治疗咳血方剂的组方规律分析[J].中国药房,2018,29(6):769-773.

[14] 秦桂超.焦中华教授治疗恶性肿瘤遣方用药规律数据挖掘研究[D].济南:山东中医药大学,2015.

[15] 陈军辉.西洋参活性成分及其指纹图谱研究[D].南昌:南昌大学,2005.

[16] 杨旭杰.基于统计方法模型分析的中药复方专利保护研究[D].北京:北京中医药大学,2012.

[17] 姚美村,张燕玲,袁月梅,等.中药药性量化方法对补虚药功效归类预测的研究[J].北京中医药大学学报,2004,27(4):7-9.

[18] 崔盈盈,吴嘉瑞,张丹,等.基于数据挖掘的清热类中药注射剂不良反应流行病学特点研究[J].药物流行病学杂志,2017,26(10):675-682.

[19] 查青林,何羿婷,喻建平,等.基于决策树分析方法探索类风湿性关节炎证病信息与疗效的相关关系[J].中国中西医结合杂志,2006,26(10):871-876.

[20] 吴嘉瑞,郭位先,张晓朦,等.基于数据挖掘的国医大师颜正华临床用药规律研究[J].中国中医药信息杂志,2014,21(8):16-18.

[21] 惠娜.近红外光谱分析技术在党参及复方丹参片质量控制中的应用[D].兰州:兰州大学,2011.

[22] 周天宁,明冬萍,赵睿.参数优化随机森林算法的土地覆盖分类[J].测绘科学,2017,42(2):88-94.

- [23] 于亚运,刘勇国,蒋羽,等.基于指纹相似度的药物-靶点相互作用预测[J].中国中药杂志,2017,42(18):3578-3583.
- [24] 夏伯候,胡玉珍,熊苏慧,等.随机森林算法在中药指纹图谱中的应用:以不同品牌夏桑菊颗粒指纹图谱分析为例[J].中国中药杂志,2017,42(7):1324-1330.
- [25] 雷蕾,王新洲,张黎,等.中药化学成分对大鼠心脏毒性的QSAR研究[J].世界科学技术-中医药现代化,2015,17(9):1833-1837.
- [26] 吴思媛,胡幼芬,刘晓伟,等.中药寒热药性分类的随机森林模型研究[J].软件导刊,2014,13(10):71-74.
- [27] 邱一卉,林成德.基于随机森林方法的异常样本检测方法[J].福建工程学院学报,2007,5(4):392-396.
- [28] 齐方,容蓉,薛付忠.贝叶斯网络模型在中药整体药性特征分析中的应用[J].山东大学学报(医学版),2011,49(5):147-152.
- [29] 田野,李舒.战丽彬治疗恶性肿瘤用药规律的数据挖掘分析[J].中医杂志,2014,55(8):657-660.
- [30] 刘颖,李江,王耘,等.贝叶斯网络在中药活血化瘀功效预测中的应用[J].北京中医药大学学报,2008,31(4):229-231.
- [31] 孙继佳,张蕾,邵建华,等.基于贝叶斯方法的中药治疗乙肝后肝硬化疗效评价[J].数理医药学杂志,2012,25(4):436-440.
- [32] 师宇,饶惠霞,翁开源.5类抗风湿性关节炎类药物的成本-效果分析:基于贝叶斯MTC法[J].广东药学院学报,2016,32(6):790-796.
- [33] 秦玉平.基于支持向量机的文本分类算法研究[D].大连:大连理工大学,2008.
- [34] 张福强,唐向阳,王俊全,等.基于机器学习的红外光谱丹参聚类分析[J].计算机与应用化学,2010,27(9):1301-1303.
- [35] 郭立玮,李玲娟,董洁.基于计算机化学方法的中药膜过程研究[J].膜科学与技术,2011,31(3):196-204.
- [36] 刘进,邓家刚,覃洁萍,等.基于红外光谱数据的中药药性识别研究[J].时珍国医国药,2010,21(3):561-563.
- [37] 李飞雁,徐曼菲,乔延江.基于近红外光谱技术天麻的产地区分[J].世界科学技术-中医药现代化,2015,17(7):1405-1408.
- [38] 徐明玲.活血化瘀类中药的构效关系研究[D].北京:北京化工大学,2009.
- [39] 张景芳,蒋芦荻,张燕玲.支持向量机在中药肾毒性研究中的应用[J].中国中药杂志,2015,40(6):1134-1138.
- [40] 吴惠敏,叶少珍.基于粗糙集与SVM骨性关节炎复方中药功效分析研究[J].福州大学学报(自然科学版),2013,41(3):311-316.
- [41] 周玉.模糊竞争神经网络在中医舌诊中的应用研究[D].广州:广东工业大学,2007.
- [42] 吴纯伟,郭嘉雯,陈超,等.基于人工神经网络优化脑脉通治疗缺血性脑中风组分配伍研究[J].中国药理学杂志,2016,51(6):454-458.
- [43] 侯立强,王爽.中药指纹图谱识别的机器学习算法研究[J].内蒙古中医药,2014,33(22):97-98.
- [44] 刘现磊,孙志海.人工神经网络在中药制剂研究中的应用[J].中国中医药现代远程教育,2012,10(3):128-130.
- [45] 雍小嘉.采用知识发现技术判断方剂功效的方法研究[D].成都:成都中医药大学,2006.
- [46] 李雨,李骁,薛付忠,等.基于人工神经网络的中药药性判别研究[J].山东大学学报(医学版),2011,49(1):57-61.
- [47] 元四辉,苏薇薇,吴忠.中药鉴定新方法:计算机模式分类技术及其应用[J].中药材,1997(8):424-427.
- [48] 韩胜男.基于中药组效关系的花椒挥发油抗肿瘤活性组分的研究[D].天津:天津大学,2014.
- [49] 冯军勤,周誉昌,吕华,等.运用近红外漫反射光谱技术检测中药水分含量[J].大众科技,2006(2):46-47.
- [50] 许良,毕开顺.多元线性回归分析法在蒙药森登-4汤谱效关系解析中的应用[J].计算机与应用化学,2008(10):1189-1192.
- [51] 刘吉楠.北豆根细胞毒活性谱效关系研究[D].哈尔滨:黑龙江大学,2012.
- [52] 李爱敏,王立堂,朱丽军.对现代六经辨证中运用中药配伍用量的数理统计分析[J].中医研究,1995(5):9-10.
- [53] 周保安,冯玉林,吴成.2型糖尿病的补益药证治用药规律及logistic多元回归分析[J].实用糖尿病杂志,2015,11(3):30-32.
- [54] 刘红杰,陈亮,李天昊,等.基于中医传统理论建立中药肾毒性的logistic回归预测模型[J].中药新药与临床药理,2016,27(4):571-577.
- [55] 林宁,余楚婷,范海媚,等.上呼吸道感染治疗用药对感染后咳嗽发病的影响[J].中国中医急症,2016,25(10):1904-1906.
- [56] 黄翌.慢性乙型肝炎中医辨证计量诊断及用药特点的研究[D].广州:暨南大学,2005.
- [57] 邹锦,杨学琴,张小勇,等.基于关联规则和熵聚类的中医经方治疗抑郁症用药规律[J].中国实验方剂学杂志,2017,23(10):211-215.
- [58] 董俊龙.关联规则和聚类分析在方剂配伍规律研究中的应用[D].沈阳:东北大学,2010.

(收稿日期:2018-04-16 修回日期:2018-07-05)

(编辑:唐晓莲)