

引用格式:吴地尧,章新友,张玉娇,等.分类算法在中药研究中的应用及其进展[J].科学技术与工程,2019,19(35):1-9  
Wu Diyao, Zhang Xinyou, Zhang Yujiao, et al. Application and progress of classification algorithm in traditional Chinese medicine research [J]. Science Technology and Engineering, 2019, 19(35): 1-9

## 医药、卫生

# 分类算法在中药研究中的应用及其进展

吴地尧 章新友<sup>\*</sup> 张玉娇 牛晓录

(江西中医药大学药学院,南昌 330004)

**摘要** 为了研究分类算法在中药研究中的应用及其进展,为中药领域数据挖掘的相关研究提供有益的参考和借鉴,利用知网、万方和维普等作为主要文献检索平台,整理文献建立研究对象文献数据库。从311 746篇文献中,整理得到345篇相关文献。结果显示分类算法的应用较多集中在中药药性研究、药物分析、药物安全性/毒性、方剂配伍规律和中药功效等5个研究子领域,且人工神经网络算法运用的范围最广,适用于中药科研的多个子领域。因此得出:分类算法在中药科研中应用种类多,且涵盖的研究范围广;不同数学模型的分类算法均有特色,应当结合中药数据的特点,优选不同的分类算法;通过改进完善传统的分类算法,更有利于充分发挥分类算法在中药传承和创新中的作用。

**关键词** 分类算法 中药研究 数据挖掘 文献分析法

**中图法分类号** R961 TP391; **文献标志码** A

近年来,数据挖掘技术在中药领域的科研中得到了广泛应用,大大推动了中药研究的现代化和信息化<sup>[1,2]</sup>。数据挖掘方法是通过一定的算法和技术从数据库中发现隐含的知识,常用的数据挖掘方法有六大类:分类分析、关联规则分析、回归分析、聚类分析、离群点检测分析、特征化与区分<sup>[3]</sup>。尤其是分类算法所包含的算法模型种类繁多,其应用所覆盖的中药科研子领域比较广泛,各种分类算法模型既可以单独用于分析预测目标样本的属性和特点,也可以和传统的实验研究相结合,能较好地为实验作前期指导或对繁冗数据的处理。鉴于分类算法在中药科研中的应用广泛、算法多样,且迄今为止,未见对分类算法在中药科研中的应用现状和进展研究的综述报道。为此,笔者在构建研究对象文献数据库的基础上,对常用的各种分类算法模型在中药研究中的应用及其特点进行比较分析,全面地论述分类算法在中药科研中的应用现状,并对分类算法的改进与应用前景,提出了建设性的意见和预测。

2019年4月15日收到 国家自然科学基金(81660727)资助  
第一作者简介:吴地尧(1984—),女,汉族,江西南昌人,博士研究生,讲师,主管药师。E-mail:53792068@qq.com。

\*通信作者简介:章新友(1962—),男,汉族,江西德安人,博士,教授,博士研究生导师。E-mail:xinyouzhang@163.com。

## 1 中药研究分类算法文献检索与数据库构建

以中国知网作为主要文献检索平台,并利用万方和维普网补充检索未收录在知网中的其他文献,整合形成研究对象文献数据库。检索时,先以“中药”为主题词搜索,获得文献311 746篇,再以具体的分类算法名称为主题词,逐个继续在结果中检索(在311 746篇文献中继续检索)下载相关文献,搜索时间从2000年1月~2018年6月,检索的算法名称是根据文献[3—5]综述总结得出的在各行业中较为常用的分类算法。检索的具体名称如下:“决策树”、“随机森林”、“贝叶斯分类”、“支持向量机”、“人工神经网络”、“KNN”(“K-近邻”作为补充搜索)、“K-Means”(“K-均值”作为补充)、“Boosting”、“bagging”。另外,在万方和维普网文献检索平台,以“中药”和某具体挖掘算法名称为关键词或题名进行检索,查找并下载未收入在知网中的其他文献进行补充。然后,再将单个文献库里的文献按照算法运用的研究子领域进行分类,方法流程如图1所示。对于子领域内文献总数超过10篇以上者,统计文献篇数超过3篇及以上的作为该算法主要应用的子领域,如研究子领域分栏里只有一两篇文献的则不作讨论,合并归类在“其他”项里;对于子领域内文献总数在10篇以下者,则

统计文献篇数超过2篇及以上的作为该算法主要应用的子领域,如研究子领域分栏里只有一篇文献的则不作讨论,合并归类在“其他”项里。最后,根据分类探讨各算法在中药研究各子领域中的应用现状,并进行比较分析。

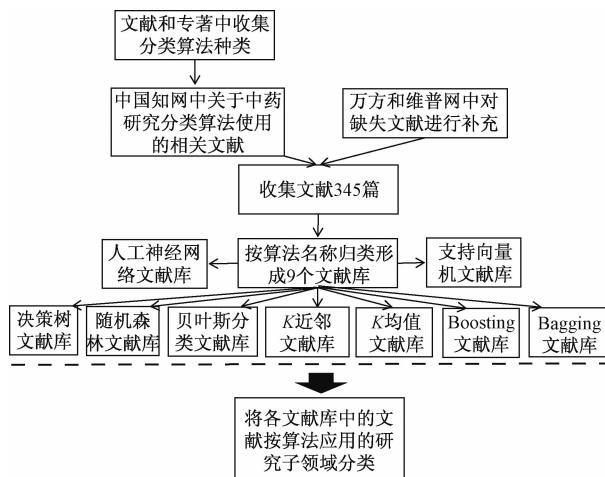


图1 文献收集及归类统计方法流程图

Fig. 1 Flowchart of statistical methods for literature collection and classification

## 2 中药研究中常用分类算法应用的现状分析

### 2.1 中药研究中常用的分类算法应用情况分析

利用上述方法检索并整理得到316篇相关文献,按算法名称进行统计分析其应用状况,所得结果见表1。

表1 中药研究中分类算法相关文献的检索结果

Table 1 Retrieval results of literature related to classification algorithm in traditional Chinese medicine research

算法名称	文献篇数
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
K近邻聚类	9
K均值聚类	4
Boosting	14
Bagging	2
总计	345

### 2.2 中药研究中常用各分类算法及其应用分析

#### 2.2.1 决策树及其应用分析

决策树是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大于等于零的概率,评价项目风险,判断其可行性的决

策分析方法,是直观运用概率分析的一种图解法<sup>[6]</sup>。由于这种决策分支画成图形很像一棵树的枝干,故称决策树<sup>[7]</sup>。检索基于决策树算法的中药研究相关文献,共收集文献44篇,依据每篇文献运用该算法的研究子领域,再将文献逐个分类,结果表明决策树较常应用在中药药性研究<sup>[8-12]</sup>等数个研究子领域,见表2。

表2 基于决策树算法的中药研究相关文献统计

Table 2 Literature statistics of traditional Chinese medicine research based on decision tree algorithm

研究子领域	文献篇数
中药药性研究	14
药物安全性评价	8
药物分析	6
药物疗效评价	5
方剂配伍规律	3
其他	8
总计	44

#### 2.2.2 随机森林及其应用分析

随机森林是一个包含多个决策树的分类器,并且其输出的类别是由个别树输出的类别的众数而定<sup>[13]</sup>。从直观角度来解释,即每棵决策树都是一个分类器,那么对于一个输入样本,N棵树会有N个分类结果。而随机森林集成了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<sup>[14]</sup>。检索基于随机森林算法的中药研究相关文献,共收集文献22篇,依据每篇文献运用该算法的研究子领域,再将文献逐个分类,结果表明随机森林较常运用于药物分析<sup>[15-19]</sup>等数个研究子领域,见表3。

表3 基于随机森林算法的中药研究相关文献统计

Table 3 Literature statistics of traditional Chinese medicine research based on random forest algorithm

研究子领域	文献篇数
药物分析	5
中药药性研究	3
药物靶点预测	3
药物安全性评价	3
其他	8
总计	22

#### 2.2.3 贝叶斯分类及其应用分析

贝叶斯分类是基于贝叶斯定理的统计学分类方法。其定义表达式为 $P(H|X) = [P(X|H)P(H)] / P(X)$ <sup>[3]</sup>。贝叶斯定理可表述为对于给出的待分类项,求解此项在各个类别中出现的概率,概率最大的项即认为此待分类项属于该类别<sup>[20]</sup>。检索基于贝叶斯分类算法的中药研究相关文献,共收集文献34篇,依据每篇文献运用该算法的研究子领域,再将文

文献逐个分类,结果表明贝叶斯分类算法较常运用于中药药性研究<sup>[21-25]</sup>等数个研究子领域,见表4。

**表4 基于贝叶斯分类算法的中药研究相关文献统计**  
**Table 4 Literature statistics of traditional Chinese medicine research based on Bayesian classification algorithm**

研究子领域	文献篇数
中药药性研究	9
方剂配伍规律	7
中药功效研究	4
药物疗效评价	3
药物成本-效果评价	3
其他	8
总计	34

## 2.2.4 支持向量机及其应用分析

支持向量机是针对二值分类问题提出的,旨在从样本中找到一个支持向量,能够建构出最好的分类超平面,这是支持向量机的核心内容<sup>[26]</sup>。检索基于支持向量机的中药研究相关文献,共收集文献87篇,依据每篇文献运用该算法的研究子领域,再将文献逐个分类,结果表明经分类统计支持向量机较常运用于药物分析<sup>[27-33]</sup>等数个研究子领域,见表5。

**表5 基于支持向量机算法的中药研究相关文献统计**  
**Table 5 Literature statistics of traditional Chinese medicine research based on support vector machine algorithm**

研究子领域	文献篇数
药物分析	25
制剂工艺研究	17
中药药性研究	12
中药鉴定	7
活性成分研究	5
中药毒性研究	5
中药功效研究	3
其他	13
总计	87

## 2.2.5 人工神经网络及其应用分析

神经网络是一类模式匹配算法,能够有效处理复杂系统中的数据,从中寻找出相应的模式、规律,归纳隐含在信息单元之间的关联规则的有效方法。它由大量处理单元相互连接构成,通过内部大量节点相互之间连接关系的调整,依靠网络系统的复杂程度,来实现信息处理的目的<sup>[34]</sup>。检索基于人工神经网络的中药研究相关文献,共收集文献129篇,依据每篇文献运用该算法的研究子领域,再将文献逐个分类,结果表明人工神经网络较常运用于药物分析<sup>[35-41]</sup>等数个研究子领域,见表6。

## 2.2.6 K-近邻聚类及其应用分析

K-近邻(KNN,K-nearest neighbors)聚类算法是

**表6 基于人工神经网络算法的中药研究相关文献统计**  
**Table 6 Literature statistics of traditional Chinese medicine research based on artificial neural network algorithm**

研究子领域	文献篇数
药物分析	41
制剂工艺研究	19
方剂配伍规律	9
中药功效研究	8
中药鉴定	7
中药药性研究	5
活性成分研究	5
其他	35
总计	129

一种基于实例的分类方法。该方法就是找出与未知样本 $x$ 距离最近的 $k$ 个训练样本,看这 $k$ 个样本中多数属于哪一类,就把 $x$ 归为那一类<sup>[42]</sup>。检索基于K-近邻算法(KNN)的中药研究相关文献,共收集文献9篇,依据每篇文献运用该算法的研究子领域,再将文献逐个分类,结果表明K-近邻较常运用于药物分析<sup>[43]</sup>研究子领域,见表7。

**表7 基于人工神经网络算法的中药研究相关文献统计**  
**Table 7 Literature statistics of traditional Chinese medicine research based on artificial neural network algorithm**

研究子领域	文献篇数
药物分析	5
其他	4
总计	9

## 2.2.7 K-均值聚类及其应用分析

K-均值聚类(K-Mean)是指通过迭代逐次移动各类别的中心,直至得到最好的聚类结果为止,即类别中所有对象与中心的欧式距离最小。K-均值聚类以类别内的高相似性和类别间的低相似性为目标<sup>[44]</sup>。检索基于K-均值聚类算法的中药研究相关文献,共收集文献5篇,依据每篇文献运用该算法的研究子领域,再将文献逐个分类,结果表明K-均值聚类算法较常运用于方剂配伍规律<sup>[45]</sup>和药物分析研究<sup>[46]</sup>子领域,见表8。

**表8 基于人工神经网络算法的中药研究相关文献统计**  
**Table 8 Literature statistics of traditional Chinese medicine research based on artificial neural network algorithm**

研究子领域	文献篇数
方剂配伍规律	2
药物分析	2
其他	1
总计	5

### 2.2.8 Boosting(提升)算法及其应用分析

Boosting 算法是一种基于一系列基础分类器的组合分类模型算法,基础分类器可以选择任意单分类模型(如决策树等),单分类模型间不存在强依赖关系、可同时并行操作的方法<sup>[47]</sup>。检索基于 Boosting 算法的中药研究相关文献,共收集文献 14 篇,依据每篇文献运用该算法的研究子领域,再将文献逐个分类,结果表明 Boosting 算法较常运用于中药疗效评价<sup>[48]</sup>、药物分析<sup>[49]</sup>等研究子领域,见表 9。

**表 9 基于 Boosting 算法的中药研究相关文献统计**

**Table 9 Research literature of Chinese medicine based on Boosting algorithm statistics**

研究子领域	文献篇数
中药疗效评价	5
药物分析	4
其他	5
总计	14

### 2.2.9 Bagging(装袋)算法及其应用分析

Bagging 算法是一种基于一系列基础分类器的组合分类模型算法,基础分类器可以选择任意单分类模型(如决策树等),单分类模型间存在强依赖关系、必须串行生成序列化算法<sup>[50]</sup>。检索基于 bagging 算法的中药研究相关文献,共收集文献 4 篇,因文献数过少,不具备统计意义。

## 3 中药研究中常用分类算法的应用子领域分析

### 3.1 分类算法在中药研究中的主要应用领域

经统计分析,分类算法的应用较多集中在中药药性研究、药物分析、药物安全性/毒性、方剂配伍规律和中药功效研究等 5 个主要子领域,各子领域中可运用的分类挖掘算法统计分析结果,见表 10。

**表 10 分类算法应用较多的中药研究子领域**

**Table 10 Sub-fields of traditional Chinese medicine research in which classification algorithm is widely used**

研究子领域	可运用的分类算法
中药药性研究	决策树、随机森林、贝叶斯分类、人工神经网络
药物分析	决策树、随机森林、支持向量机、人工神经网络、K-近邻聚类、Boosting、K-均值聚类
药物安全性/毒性	决策树、随机森林、支持向量机、人工神经网络
方剂配伍规律	决策树、贝叶斯分类、人工神经网络、K-均值聚类
中药功效研究	贝叶斯分类、支持向量机、人工神经网络

### 3.2 分类算法在中药研究主要领域应用分析

#### 3.2.1 中药药性研究领域应用分析

中药药性理论是中药理论的核心,内容包括四气五味、归经、升降沉浮和毒性。中药药性既是对各

种药物性质及作用规律的概括,也是指导中医临床辩证用药的理论依据。由于中药药性理论大多来自古籍记载,导致出现记录不统一、记载缺省等现象,亟待更正及补充<sup>[51]</sup>。另外,还有新的药物组分或化合物,需要预测药性指导临床应用。因此,建立分类模型是最简便、直观、快速的方法。在中药药性预测的相关研究中,决策树、随机森林、贝叶斯分类和人工神经网络四种分类算法都可被运用于该子领域。

#### 3.2.2 药物分析领域应用分析

数据挖掘算法在药物分析中的运用主要集中在药物含量测定、指纹图谱的建立及模式识别的研究中。中药材因种类繁多,成分复杂,命名混乱,道地药材鉴定困难,中药指纹图谱的建立使中药质量评价体系有了革命性的发展<sup>[52]</sup>,结合分类算法和指纹图谱方法可深入研究图谱信息,发现和解析指纹图谱中的潜在信息。或者利用分类算法优化操作条件,与中药内在信息建立模式识别平台,实现对中药的快速鉴定和测定。在药物分析的相关研究中,包括决策树、随机森林、贝叶斯分类、人工神经网络、K-近邻聚类、Boosting 和 K-均值聚类的 7 种分类算法都可被运用于该子领域。

#### 3.2.3 中药安全性研究领域应用分析

中药制剂因为品种繁多成分复杂,个体化用药差异等情况造成中药安全性研究困难重重,中药不良反应已成为掣肘中医药国际化的影响因素之一<sup>[53]</sup>。当前,数据挖掘技术是国内外药物安全事件分析与评价的主要方法,可结合以往的不良反应报告、药物信息、患者因素、处方剂量等作为分析属性,用分类算法建立检测和预警系统,有效开展安全性评价,预防和减少不良反应的发生。在中药安全性的相关研究中,决策树、随机森林、贝叶斯分类和人工神经网络四种分类算法都可被运用于该子领域。

#### 3.2.4 方剂配伍规律研究领域应用分析

中药方剂配伍规律是中药方剂研究的核心问题之一<sup>[54]</sup>,数据挖掘可从方剂用药频次、功效组合等多个角度与层次,为科学合理高效的分析目标方剂数据所蕴含的配伍规律提供可靠的方法。单用或联用挖掘算法可对药组、药对进行组方分析,以及对历代名医遣药组方规律的发掘。在方剂配伍规律的相关研究中,决策树、贝叶斯分类、人工神经网络和 K-均值聚类四种分类算法都可被运用于该子领域。

#### 3.2.5 中药功效研究领域应用分析

中药功效是依据中医理论高度概括的药物治疗作用,也是中药理论的重要组成内容。由于中药功效大多来自古籍记载,导致出现功效记录不统一、名称不规范的现象<sup>[55]</sup>,或者由于药物本身具有多个临

床作用,使得同一药物在不同文献中分属不同类别,因此结合中药药性特点、药理指标等属性,利用分类算法建立归类模型,可以避免传统经验带来的误差,为中药功效的研究提供了新思路。在中药功效的相关研究中,贝叶斯分类、支持向量机和人工神经网络三种分类算法都可被运用于该子领域。

## 4 中药研究中常用各分类算法模型的比较分析

### 4.1 决策树与随机森林算法模型的分析比较

相对于其他挖掘算法,决策树更易于被人理解,可以列出决策问题的全部可行方案和可能出现的各种自然状态,以及各方法在不同状态下的期望值,在运用于方案选择或因素分析时,更易于直观地预测或表述结论。其次,执行效力较高,能够短时间内对大量数据源做出可行且效果良好的预测分析,例如可处理大量的中药化合物集或数据库中大量的处方样本<sup>[56]</sup>,同时,如数据量较小时,也能做出较准确的分析。该算法的缺点在于容易产生过拟合的现象,使得建立的分类模型虽能与当前数据集完美匹配,但无法适应其他数据集,对未知的数据集的预测效果差。随机森林与决策树算法的分析过程大致相同,虽然单棵树容易产生过拟合的现象,但因为随机森林存在多棵树,属性广度的增加也可消除掉过拟合现象。因此,基于随机森林建立的模型泛化能力更强,实用性更高。另外,两者共同存在的缺点即当数据预处理工作量比较大,当类别太多或当分类属性有缺失值,树的性能将会降低。

### 4.2 决策树与贝叶斯分类算法模型的分析比较

就决策树或随机森林而言,分类的归属一般由结点测试的结果决定。但如果结果以数值型呈现,或者没有证据和经验可循时,则无法得出有效的分类预测。例如,文献[57]报道以生理生化为指标建立决策树模型预测药物的寒热药性,树模型某一通路最终的测试结点为红细胞数,结果为5.57,仅凭数值尚无法断定药性为寒或热,则该预测模型无法建立。可见,决策树或随机森林的设置需要有较为准确的经验、知识或证据。贝叶斯分类算法和决策树、随机森林算法相比,更适合用于处理不确定性推理的复杂问题。在贝叶斯分类的运行步骤中,每一步都是用概率大小来推进后续分类,以复杂的联合概率来确保预测的准确性。因此,常常将贝叶斯分类过程中得出的条件概率作为指导,用于建立树模型。因此,首先建立了贝叶斯模型,利用训练集得出概率表,再根据概率建立决策树。如在上例中提到当某一通路最终结点——红细胞数为5.57时,根据

概率分布可得知药物是寒性的概率为87.5%,因此可将该通路的最终分类结果输出为“寒性”。

### 4.3 支持向量机与人工神经网络算法模型的分析比较

相比较其他算法,人工神经网络的在中药研究中的应用报道较多,呈现多元化、多方向发展的特点。这是因为该算法线性拟合能力强,分类准确度高,具有强大的自我学习能力和联想记忆的功能。科学界正在兴起一种学说认为科学主体是非线性的,它更接近感觉领域,也更接近世界的真实面目<sup>[58]</sup>。中医药研究作为科学研究的一部分,也正面临着这样的挑战,这就使得神经网络在分析具有非线性特征的中药数据时,更加客观准确,预测结果拟合度较高。但该算法同样有缺点,在设计过程中它需要大量的参数,且无法显示推理过程,可能导致输出结果接受度低。支持向量机在完成分类任务时也能自动寻找出对分类有较好区分能力的支持向量,由此构造出的分类器可以最大化累与类的间隔<sup>[59]</sup>。综合归纳两种算法的特点,对于分析结果输出为多分类的研究问题适合用人工神经网络算法解决,而多变量小样本的二分类研究问题适合用支持向量机算法解决。

### 4.4 中药研究中组合分类各算法模型的比较分析

组合分类把单个分类模型组合在一起,创建一个改进的复合分类模型。装袋、提升和随机森林都是组合分类方法。提升与装袋相比,由于提升过度关注模型的错误率,因此存在复合模型对数据过分拟合的危险,而装袋则不太收到过分拟合的影响。尽管与单个模型相比,两者都能够显著提高准确率,但是提升往往得到更高的准确率。随机森林的准确率可与提升相媲美,但是对错误和离群点更鲁棒,随着森林的个数增加,森林的泛化误差也逐渐收敛。随机森林的准确率依赖于个体分类器的实力和它们之间的依赖性。由于随机森林在每次划分时只考虑很少的属性,因此它们在大型数据库上非常有效,效率更高于装袋和提升<sup>[1]</sup>。

## 5 结论与展望

依据研究结果可以得出:一是分类算法在中药科研中应用种类多,且涵盖的研究范围广;二是不同数学模型的分类算法均有特色,应当结合中药数据的特点,优选不同的分类算法;三是数据挖掘技术在中药领域的应用,大大地推动了中药研究现代化、科学化、规范化的进程。但由于中药领域数据的特殊性,挖掘算法的使用仍然还在不断的探索与拓展当中。分析现有的文献显示,研究者经常反复单独使

用一种算法研究类似的问题，导致难以取得突破性的进展。探讨的分类算法虽种类较多，但在子领域内进行各算法优劣比较的报道较少，笔者认为运用策略模式的理念<sup>[60]</sup>，即尝试用各种算法来分析相同的问题，或是用同一种算法来解决不同的问题，如此一来，既可优选结果，也可综合比较各算法对同一问题或不同算法对同一问题的结果，从而优选最佳方案得到更精准的结果。同时，也可将某种算法推广到未曾探索的新领域，使该方法得到最大化的应用，这必将促进挖掘算法的发展。笔者认为数据挖掘在中药研究中具有广泛的应用前景，且在中药研究的新兴应用领域更具有创新性和挑战性，应当结合中药数据的特点致力于改进完善传统的算法，并秉持开放的态度不断引入新的技术和新算法，充分发挥数据挖掘在中药传承和创新中的作用，定当为中药现代化研究和信息化、国际化提供有力的技术支持。

## 参 考 文 献

- 1 王 鑫，何丽云，白 艳. 中药信息学的研究方法与应用荟萃 [J]. 医学信息, 2015(14) : 346-347  
Wang Xin, He Liyun, Bai Yan. Meta-analysis of research methods and applications of TCM informatics [J]. Medical Information, 2015 (14) : 346-347
- 2 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  
Tong Xu, Xie Qingyu, Meng Qinggang. On the scientific value of TCM data integration analysis under the background of big data era [J]. Journal of Chinese Medicine Information, 2015, 22(8) : 1-3
- 3 韩家炜, Micheline Kamber, 裴 健. 数据挖掘: 概念和技术 [M]. 3 版. 北京: 机械工业出版社, 2017  
Han Jiawei, Micheline Kamber, Pei Jian. Data mining—concepts and techniques [M]. 3rd ed. Beijing: China Machine Press, 2017
- 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  
Zou Zhiwen, Zhu Jinwei. Research and review of data mining algorithms [ J ]. Computer Engineering & Design, 2005, 26 ( 9 ) : 2304-2307
- 5 周牒嵒, 陈 琳, 向 华. 数据挖掘算法研究[J]. 现代电子技术, 2011, 34(20) : 75-78  
Zhou Dielan, Chen Lin, Xiang Hua. Research on data mining algorithm [J]. Modern Electronic Technology, 2011, 34(20) : 75-78
- 6 王楚亮, 郭保辉, 许 林. 概率发生时点不确定情况下的决策树分析[J]. 现代经济(现代物业中旬刊), 2009, 8(5) : 142, 27  
Wang Chuliang, Guo Baohui, Xu Lin. Decision tree analysis under uncertainty of probability occurrence time point [ J ]. Modern Economy ( Modern Property Mid-month ), 2009, 8(5) : 142, 27
- 7 吴嘉瑞, 秦 丹, 张 冰, 等. 基于决策树算法的七情配伍相使相恶药性规律研究[J]. 中国医药指南, 2014 (3) : 157-158  
Wu Jiarui, Qin Dan, Zhang Bing, et al. Study on the law of the compatibility of seven emotions and their adverse effects based on decision tree algorithm [ J ]. Chinese Medical Guide, 2014 (3) : 157-158
- 8 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  
Yao Meichun, Zhang Yanling, Yuan Yuemei, et al. Prediction of efficacy classification of traditional Chinese medicine by quantitative method [ J ]. Journal of Beijing University of Chinese Medicine, 2004, 27(4) : 7-9
- 9 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  
Cui Yingying, Wu Jiarui, Zhang Dan, et al. Epidemiological characteristics of adverse reactions of heat-clearing TCM injections based on data mining [ J ]. Journal of Pharmacopidemiology, 2017, 26(10) : 675-682
- 10 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  
Zha Qinlin, He Yiting, Yu Jianping, et al. Relationship between syndrome information and curative effect of rheumatoid arthritis based on decision tree analysis [ J ]. Chinese Journal of Integrated Traditional Chinese and Western Medicine, 2006, 26 ( 10 ) : 871-876
- 11 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  
Wu Jiarui, Guo Weixian, Zhang Xiaomeng, et al. Research on clinical medication law of Yan Zhenghua, a Chinese medicine master based on data mining [ J ]. Journal of Chinese Medicine Information, 2014, 21(8) : 16-18
- 12 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  
Hui Na. Application of near infrared spectroscopy in quality control of dangshen and compound danshen tablets [ D ]. Lanzhou: Lanzhou University, 2011
- 13 贺建章, 王海波, 季知祥, 等. 基于随机森林理论的配电变压器重过载预测[J]. 电网技术, 2017(8) : 201-205  
He Jianzhang, Wang Haibo, Ji Zhixiang, et al. Weight overload prediction of distribution transformer based on random forest theory [ J ]. Power Grid Technology, 2017(8) : 201-205
- 14 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  
Zhou Tianning, Ming Dongping, Zhao Rui. Land cover classification based on parameter optimization random forest algorithm [ J ]. Science of Surveying and Mapping, 2017, 42(2) : 88-94
- 15 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  
Yu Yayun, Liu Yongguo, Jiang Yu, et al. Prediction of drug-target interaction based on fingerprint similarity [ J ]. Chinese Journal of Traditional Chinese Medicine, 2017, 42 ( 18 ) : 3578-3583
- 16 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  
Xia Bohou, Hu Yuzhen, Xiong Suhui, et al. Application of random forest algorithm in fingerprint analysis of traditional Chinese medicine: a case study of fingerprint analysis of Xia Sangju granules from different brands [ J ]. Chinese Journal of Traditional Chinese Medi-

- cine, 2017, 42(7): 1324-1330
- 17 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(9): 1833-1837  
Lei Lei, Wang Xinzhou, Zhang Li, et al. QSAR study on cardiotoxicity of Chinese medicine chemical components in rats [J]. World Science and Technology-Modernization of Chinese Medicine, 2015 (9): 1833-1837
- 18 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  
Wu Siyuan, Hu Youfen, Liu Xiaowei, et al. Random forest model study on the classification of cold and hot properties of traditional Chinese medicine [J]. Software Guide, 2014, 13(10): 71-74
- 19 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  
Qiu Yihui, Lin Chengde. Abnormal sample detection method based on random forest method [J]. Journal of Fujian University of Engineering, 2007, 5(4): 392-396
- 20 陈红松, 王 钢, 宋建林. 基于云计算入侵检测数据集的内网用户异常行为分类算法研究[J]. 信息网络安全, 2018(3): 1-7  
Chen Hongsong, Wang Gang, Song Jianlin. Research on the classification algorithm of abnormal behaviors of Intranet users based on cloud computing intrusion detection data set [J]. Information Network Security, 2018(3): 1-7
- 21 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  
Qi Fang, Rong Rong, Xue Fuzhong. The application of bayesian network model in the analysis of the characteristics of Chinese traditional medicine[J]. Journal of Shandong University (Medical Science Edition), 2011, 49(5): 147-152
- 22 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  
Tian Ye, Li Shu. Data mining analysis on the medication regularity of zhanlibin in the treatment of malignant tumors [J]. Journal of Traditional Chinese Medicine, 2014, 55(8): 657-660
- 23 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  
Liu Ying, Li Jiang, Wang Yun, et al. Application of bayesian network in predicting the efficacy of traditional Chinese medicine in promoting blood circulation and removing blood stasis [J]. Journal of Beijing University of Chinese Medicine, 2008, 31(4): 229-231
- 24 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  
Sun Jijia, Zhang Lei, Shao Jianhua, et al. Efficacy evaluation of traditional Chinese medicine in the treatment of post-hepatitis b cirrhosis based on bayesian method [J]. Journal of Mathematical Medicine & Pharmacy, 2012, 25(4): 436-440
- 25 师 宇, 饶惠霞, 翁开源. 5类抗类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  
Shi Yu, Rao Huixia, Weng Kaiyuan. Cost-effective-ness analysis of 5 types of anti-rheumatoid arthritis drugs—Based on bayesian MTC method [J]. Journal of Guangdong Pharmaceutical University, 2016, 32(6): 790-796
- 26 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  
Qin Yuping. Research on text classification algorithm based on support vector machine [D]. Dalian: Dalian University of Technology, 2008
- 27 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  
Zhang Fuqiang, Tang Xiangyang, Wang Junquan, et al. Cluster analysis of salvia miltiorrhiza by infrared spectroscopy based on machine learning [J]. Computer & Applied Chemistry, 2010, 27 (9): 1301-1303
- 28 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  
Guo Liwei, Li lingjuan, Dong Jie. Research on Chinese medicine film process based on computer chemistry [J]. Membrane Science and Technology, 2011, 31(3): 196-204
- 29 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  
Liu Jin, Deng Jiagang, Qin Jieping, et al. Study on the identification of the medicinal property of traditional Chinese medicine based on infrared spectral data [J]. Shi Zhen Journal of Traditional Chinese Medicine and Traditional Chinese Medicine, 2010, 21 (3): 561-563
- 30 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  
Li Feiyan, Xu Manfei, Qiao Yanjiang. Origin differentiation of gasterodia elata based on near-infrared spectroscopy [J]. World Science and Technology-Modernization of Traditional Chinese Medicine, 2015, 17(7): 1405-1408
- 31 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  
Xu Mingling. Study on the structure-activity relationship of traditional Chinese medicine for promoting blood circulation and removing blood stasis [D]. Beijing: Beijing University of Chemical Technology, 2009
- 32 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  
Zhang Jingfang, Jiang Ludi, Zhang Yanling. Application of support vector machine in the study of nephrotoxicity of traditional Chinese medicine [J]. Chinese Journal of Traditional Chinese Medicine, 2015, 40(6): 1134-1138
- 33 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  
Wu Huimin, Ye Shaozhen. Efficacy analysis of traditional Chinese medicine for osteoarthritis based on rough set and SVM [J]. Journal of Fuzhou University ( Natural Science Edition ), 2013, 41 (3): 311-316
- 34 修春波. 人工智能[M]. 北京: 机械工业出版社, 2011  
Xiu Chunbo. Artificial intelligence [M]. Beijing: China Machine Press, 2011
- 35 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]. 中国药学杂志, 2016, 51

- (6) : 454-458
- Wu Chunwei, Guo Jiawen, Chen Chao, et al. Study on optimizing the compatibility of naomaitong in the treatment of ischemic stroke based on artificial neural network [J]. Chinese Pharmaceutical Journal, 2016, 51(6) : 454-458
- 36 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
- Hou Liqiang, Wang Shuang. Machine learning algorithm for finger-print recognition of traditional Chinese medicine [J]. Inner Mongolia Journal of Traditional Chinese Medicine, 2014, 33(22) : 97-98
- 37 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
- Li Yu, Li Xiao, Xue Fuzhong, et al. Study on the property of traditional Chinese medicine based on artificial neural network [J]. Journal of Shandong University (Medical Edition), 2011, 49(1) : 57-61
- 38 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
- Liu Xianlei, Sun Zhihai. Application of artificial neural network in the research of traditional Chinese medicine preparation [J]. China Modern Distance Education of Traditional Chinese Medicine, 2012, 10(3) : 128-130
- 39 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
- Yong Xiaoja. A study on the method of using knowledge discovery technology to judge the efficacy of prescription [D]. Chengdu: Chengdu University of Traditional Chinese Medicine, 2006
- 40 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
- Yuan Sihui, Su Weiwei, Wu Zhong. A new method for identification of traditional Chinese medicine: Computer model classification and its application [J]. Chinese Materia Medica, 1997 (8) : 424-427
- 41 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
- Han Shengnan. Study on the antitumor active components of essential oil from zanthoxylum bungeanum based on the relationship between Chinese traditional medicine group and effect [D]. Tianjin: Tianjin University, 2014
- 42 潘丽芳, 杨炳儒. 基于簇的 K 最近邻(KNN)分类算法研究[J]. 计算机工程与设计, 2009, 30(18) : 4260-4262
- Pan Lifang, Yang Bingru. Research on K-nearest neighbor (KNN) classification algorithm based on cluster [J]. Computer Engineering and Design, 2009, 30(18) : 4260-4262
- 43 杜建卫. 基于小波变换的中药光谱图谱与色谱图谱分类特性研究[J]. 北京石油化工学院学报, 2016, 24(1) : 59-63
- Du Jianwei. Study on the classification characteristics of Chinese traditional medicine spectrum and chromatogram based on wavelet transform [J]. Journal of Beijing Institute of Petrochemical Technology, 2016, 24(1) : 59-63
- 44 张安定. 遥感原理与应用题解[M]. 北京: 科学出版社, 2016
- Zhang Anding. Remote sensing theory and application [M]. Beijing: Science Press, 2016
- 45 刘 广, 孙艳秋. 基于 K-means 聚类算法的消渴方剂研究[J]. 中华中医药学刊, 2017(1) : 173-178
- Liu Guang, Sun Yanqiu. Research on thirst-quenching formula based on K-means clustering algorithm [J]. Chinese Journal of Traditional Chinese Medicine, 2017(1) : 173-178
- 46 Li P F, He M X, Xu Z, et al. Terahertz spectral clustering of traditional Chinese medicine based on the first derivative[J]. Journal of Measurement Science and Instrumentation, 2017(4) : 371-377
- 47 付 彬, 王志海, 王中锋. Boosting 算法中基分类器权重的动态赋值[J]. 广西师范大学学报(自然科学版), 2009, 27 (3) : 85-88
- Fu Bin, Wang Zhihai, Wang Zhongfeng. The weight of the base classifier in the Boosting algorithm dynamic assignment [J]. Journal of Guangxi Normal University (Natural Science Edition), 2009, 27 (3) : 85-88
- 48 谢红艳, 谢春光, 高 泓. 中药复方对 2 型糖尿病血管病变患者 hs-CRP, IL-6 的影响[J]. 中国实验方剂学杂志, 2014, 20 (10) : 209-212
- Xie Hongyan, Xie Chunguang, Gao Hong. Effect of Chinese herbal compound on hs-crp, il-6 in patients with type 2 diabetic vascular disease [J]. Chinese Journal of Experimental Formulae, 2014, 20 (10) : 209-212
- 49 陈 昭, 吴志生, 史新元, 等. Bagging 偏最小二乘和 Boosting 偏最小二乘算法的金银花醇沉过程近红外光谱定量模型预测能力研究[J]. 分析化学, 2014, 42(11) : 1679-1686
- Chen Zhao, Wu Zhisheng, Shi Xinyuan, et al. Bagging and Boosting of least squares algorithm of least squares honeysuckle alcohol sinking process of the near infrared spectrum quantitative models predict ability research [J]. Journal of Analytical Chemistry, 2014, 42 (11) : 1679-1686
- 50 赵 苏, 李 秀, 刘文煌. 基于分类器性能评价的 Bagging 文本分类算法[J]. 计算机工程, 2008, 34(1) : 61-63
- Zhao Su, Li Xiu, Liu Wenhua. Bagging text categorization algorithm based on classifier performance evaluation [J]. Computer Engineering, 2008, 34(1) : 61-63
- 51 周 密, 王 耘, 乔延江. 利用数据挖掘方法预测中药缺失药性的初步研究[J]. 中国中医药信息杂志, 2008, 15(6) : 93-94
- Zhou Mi, Wang Yun, Qiao Yanjiang. A preliminary study on the prediction of the absence of medicinal properties of traditional Chinese medicine by data mining [J]. Chinese Journal of Information on Traditional Chinese Medicine, 2008, 15(6) : 93-94
- 52 蔡宝昌. 中药指纹图谱的研究[J]. 科学中国人, 2005 (3) : 21-22
- Cai Baochang. Research on fingerprint of traditional Chinese medicine [J]. Chinese Science, 2005 (3) : 21-22
- 53 王仲霞, 鄢 丹, 王立福, 等. 中药注射剂临床合理用药监测与不良反应应急系统的临床应用[J]. 中国中医药科技, 2013, 20(1) : 81-82
- Wang Zhongxia, Yan Dan, Wang Lifu, et al. Clinical monitoring of rational drug use of TCM injections and clinical application of adverse reaction emergency system [J]. China Science and Technology of TCM, 2013, 20(1) : 81-82
- 54 张海英, 薛 洁. 中药复方配伍规律的研究进展[J]. 新疆中医药, 2006, 24(3) : 69-71
- Zhang Haiying, Xue Jie. Research progress on compatibility of TCM compounds [J]. Xinjiang Traditional Chinese Medicine, 2006, 24 (3) : 69-71

- 55 梁新武. 中药功效表述混乱的原因、规范化研究的方法及意义 [J]. 国医论坛, 2007, 22(2) : 43-44  
 Liang Xinwu. Causes of confusion in TCM efficacy expression, methods and significance of standardized research [J]. Traditional Chinese Medicine BBS, 2007, 22(2) : 43-44
- 56 叶立, 王新洲, 朱永亮, 等. 利用树模型预测中药成分的肝毒性[J]. 中国药学杂志, 2014, 49(18) : 1583-1588  
 Ye Li, Wang Xinzhou, Zhu Yongliang, et al. Prediction of hepatotoxicity of traditional Chinese medicine components by tree model [J]. Chin J Pharmacol, 2014, 49(18) : 1583-1588
- 57 赵兴业. 中药寒热药性生理生化评价指标的初步研究[D]. 北京: 北京中医药大学, 2007  
 Zhao Xingye. A preliminary study on the physiological and biochemical evaluation indexes of the chilling-febrile property of traditional Chinese medicine [D]. Beijing: Beijing University of Traditional Chinese Medicine, 2007
- 58 孟庆刚, 王连心. 中医药研究别忘了非线性[J]. 中国中医药信息杂志, 2005, 12(9) : 5-6  
 Meng Qinggang, Wang Lianxin. Non-linearity in TCM research [J]. Chinese Journal of Information on TCM, 2005, 12(9) : 5-6
- 59 陈凯, 朱钰. 机器学习及其相关算法综述[J]. 统计与信息论坛, 2007, 22(5) : 105-112  
 Chen Kai, Zhu Yu. Overview of machine learning and its related algorithms [J]. Statistics and Information BBS, 2007, 22 (5) : 105-112
- 60 罗山水, 章新友, 张春强, 等. 基于策略模式的中药数据挖掘研究与系统设计[J]. 世界科学技术·中医药现代化, 2015, 17 (5) : 929-933  
 Luo Shanshui, Zhang Xinyou, Zhang Chunqiang, et al. Research and system design of traditional Chinese medicine data mining based on strategic model [J]. World Science and Technology-Modernization of Traditional Chinese Medicine, 2015,17(5) : 929-933

## Application and Progress of Classification Algorithm in Traditional Chinese Medicine Research

WU Di-yao, ZHANG Xin-you<sup>\*</sup>, ZHANG Yu-jiao, NIU Xiao-lu

(College of Pharmacy, Jiangxi University of Traditional Chinese Medicine, Nanchang 330004, China)

**[Abstract]** In order to study the application and progress of classification algorithm in traditional Chinese medicine (TCM) research and to provide useful reference for data mining in TCM. CNKI, Wanfang and Weipu was used as the main literature retrieval platform to establish the literature database of the research object. The result showed 345 related articles were collected from 311 746 articles. The application of classification algorithm is mainly concentrated on five research sub-fields, including the study of Chinese medicine property, drug analysis, drug safety/toxicity, prescription compatibility rules and the efficacy of Chinese medicine. Moreover, the artificial neural network algorithm has the widest application range in subfields of traditional Chinese medicine research. It is concluded that the classification algorithm is widely used in the research of traditional Chinese medicine. The classification algorithms of different mathematical models have their own characteristics and should be optimized according to the characteristics of traditional Chinese medicine data. By improving the traditional classification algorithm, it is more conducive to give full play to the role of classification algorithm in the inheritance and innovation of traditional Chinese medicine.

**[Key words]** classification algorithm      research on traditional Chinese medicine      data mining      literature analysis