

· 理论探讨与文献研究 ·

常见机器学习方法在中医诊断领域的应用述评

夏淑洁^{1,2}, 杨朝阳^{1,2}, 周常恩^{1,2}, 辛基梁^{1,2}, 张佳³, 杜国栋³, 李灿东^{1,2}

(1. 福建中医药大学中医证研究基地, 福建福州 350122; 2. 福建省中医健康状态辨识重点实验室, 福建福州 350122; 3. 厦门大学智能科学与技术系, 福建厦门 361005)

摘要: 对近年来数据挖掘技术中的常见机器学习方法应用于中医诊断的概况进行述评。近年来应用于中医诊断的常见机器学习方法有支持向量机、决策树、K近邻算法、人工神经网络、贝叶斯网络、AdaBoost算法等, 这些方法各有其优势与不足。今后在促进机器学习运用于中医诊断研究领域时, 需基于中医证的多维性、非线性和复杂性的特点, 针对中医诊疗中的各项问题制定统一的标准并规范收集数据, 选用合适的机器学习方法, 寻找中医大数据下的隐藏知识和规律, 以推动中医诊断客观化、规范化。

关键词: 支持向量机; 决策树; K近邻算法; 人工神经网络; 贝叶斯网络; AdaBoost算法; 中医诊断; 应用

中图分类号: R241

文献标志码: A

文章编号: 1007-3213(2021)04-0826-06

DOI: 10.13359/j.cnki.gzxbtcm.2021.04.032

Review of Application of Machine Learning Methods in the Field of Diagnostics of Traditional Chinese Medicine

XIA Shu-Jie^{1,2}, YANG Zhao-Yang^{1,2}, ZHOU Chang-En^{1,2}, XIN Ji-Liang^{1,2},
ZHANG Jia³, DU Guo-Dong³, LI Can-Dong^{1,2}

(1. Research Base of Traditional Chinese Medicine Syndrome, Fujian University of Traditional Chinese Medicine, Fuzhou 350122 Fujian, China; 2. Fujian Key Laboratory of Identification of Traditional Chinese Medicine Health State, Fuzhou 350122 Fujian, China; 3. Dept. of Intelligent Science and Technology, Xiamen University, Xiamen 361005 Fujian, China)

Abstract: The status of the application of common machine learning methods of the data mining technology in the field of diagnostics of traditional Chinese medicine (TCM) in recent years was reviewed. The common machine learning methods adopted in the field of the research of diagnostics of TCM in recent years included support vector machine, decision tree, K-nearest neighbor algorithm, artificial neural network, Bayesian network, and AdaBoost algorithm, and each of them has its advantages and insufficiency. It is suggested that the future research of diagnostics of TCM with the common machine learning methods should be based on the characteristics of multi-dimension, non-linearity and complexity of TCM syndrome, collecting the data of diagnostics of TCM should follow the unified criteria, the proper machine learning methods should be chosen for unveiling the hidden knowledge and regularity of TCM diagnostics, so as to promote the objectivity and standardization of TCM diagnostics.

Keywords: support vector machine; decision tree; K-nearest neighbor algorithm; artificial neural network; Bayesian network; AdaBoost algorithm; diagnostics of traditional Chinese medicine (TCM); application

收稿日期: 2020-03-12

作者简介: 夏淑洁(1991-), 女, 博士研究生; E-mail: xiashujie121@126.com

通讯作者: 李灿东(1964-), 男, 教授, 博士研究生导师; E-mail: fjzyled@126.com

基金项目: 国家自然科学基金联合基金项目(编号: U1705286); 国家中医药管理局标准化重大项目(编号: GZY-FJS-2017-64)

辨证论治是中医的精髓,精准辨证是有效治疗的前提。当前,中医辨证存在客观化、量化、标准化等方面的不足,限制了中医的进一步传承和发展,因此,迫切需要运用现代科学技术的理论和方法推进中医诊断的客观化与规范化研究^[1-2]。人体是一个复杂的有机整体,证候具有多维性、模糊性、复杂性的特点,难以简单用传统单一、线性的分析方法进行研究^[3-4]。而数据挖掘技术,尤其是机器学习^[5],利用海量临床数据,通过模拟中医临床诊断推理过程来挖掘临床诊断数据中繁杂的病-证-症关系,可克服人类认知和思维的局限性,充分照顾到中医整体观、模糊性、非线性的特点,从而寻找大数据下隐藏的中医知识和规律,助力中医诊断客观化的进程。以下对近年来常见机器学习方法在中医诊断领域的应用进行述评。

1 支持向量机

支持向量机(support vector machine, SVM)于1995年首次提出,是一种以统计学理论为基础的模式识别的机器学习方法^[6]。SVM是机器学习领域若干技术集大成者,它具有严格的理论和数学基础,可较好地实现结构风险最小化思想,能较为合理地解决小样本、非线性、高维数和局部最小等实际问题。SVM的关键在于针对样本数据不可分情况,利用核函数把一个复杂的分类任务映射,使之能转化成一个线性可分问题^[7]。因而,SVM主要的优势在于:(1)专门针对有限样本设计,目标是获得现有信息下的最优解,根据有限的样本信息在模型的复杂性及学习能力之间寻求最佳折衷^[8]。(2)算法将实际问题通过非线性变换映射到高维特征空间,并构建线性最佳逼近来解决原来空间的非线性逼近问题,这样既保证了机器学习取得最好推广能力(泛化能力),并且SVM的算法复杂性与数据维数无关,因而能较好解决维数灾难问题^[9]。

由于SVM不仅可以处理高维的特征向量,将非线性问题转化为线性可分问题,并且在样本数较少时也能获得较好分类效果,故SVM近年来在中医领域被大量使用。许明东等^[10]将支持向量机方法应用在高血压病中医证候诊断中,以常见症状、舌苔及舌体、脉象的量化数据为输入变量,

高血压病证型为输出变量,建立了基于SVM的高血压病中医证候诊断模型,该模型总体准确率为90.0%。Wang J等^[11]以SVM为基础构建了冠心病中医证候诊断模型,并得出8个不同证型中各症状条目的权重,可从深层次反映出诊治经验。Sun Z Q等^[12]通过采用非线性多元分析对症状信息降维处理,之后分别运用神经网络和多分类SVM对证候进行分类,发现中医症状信息本身具有非线性和多维性,运用多分类SVM比神经网络更具优势。但SVM算法对大规模训练样本难以实施。此外,在中医数据挖掘的实际应用中,一般要解决多类的分类问题,而SVM主要解决二分类问题,对多分类问题存在困难,此时可考虑通过多个二类支持向量机的组合或构造多个分类器的组合来解决^[13]。

2 决策树

决策树^[14](decision tree, DT)是一种相对简单的机器学习分类算法。其构建思想来源于人们的决策过程,在已知各种情况发生的概率基础上,通过构成决策树来评价项目风险,判断可行性的决策方法。因这种决策分支构成的图形很像一棵树干,故称为决策树。树中每个节点表示一个样本属性,每个分支则代表对该属性的判断,而每个叶子结点则对应最终的类别,通常将其看作一个预测模型,代表对象属性与对象值之间的一种映射关系^[15-16]。该模型易于理解和实现,用户即使未学习相关背景知识,也能发掘其简单直观的分类规则,只要通过适当的解释,用户就能理解决策树所表达的意义。其次,算法运行速度快,易转化为分类规则,分类准确率较高,只要沿着根结点向下一直到叶子结点,沿途分裂条件是唯一且确定的。但决策树在处理大样本集时,易出现过拟合现象,从而降低分类的准确性。

目前主要的决策树算法有ID3、C4.5、C5.0和CART算法等,常用于建立中医证的诊断模型、归纳名老中医辨证过程等。此外,基于大量决策树而形成的随机森林方法还可用于证候特征权重的确定。田艳鹏等^[17]采用C5.0、CRT、CHAID、QUEST决策树方法和神经网络方法提取痰湿壅盛证的诊断规律,发现C5.0决策树模型的准确率(93.7%)最高。苏翀等^[18]利用基于KL距离的决策树

建立慢性阻塞性肺病中医诊断模型,以F-Measure、G-Mean、ROC曲线下面积以及精度召回率曲线下面积作为评价指标,结果表明采用改良的决策树较传统决策树可取得更好的预测效果。刘广等^[19]基于C4.5决策树算法对经预处理后的800例中医胃炎患者的辨证数据进行分类,并提取其分类规则,该规则基本符合中医胃炎的辨证规律以及名老中医诊治胃炎的诊疗经验。吕航等^[20]运用CHAID算法,构建基于中医人格、体质类型预测2型糖尿病(T2MD)患者合并冠心病的患病风险模型,结果显示以少阴人格、阴寒质、血瘀质为预测变量构建的糖尿病患病风险模型预测准确度为93.6%,所得的3条预测冠心病患病风险的规则也符合中医理论。蔡晓路^[21]采用随机森林方法实现了类风湿关节炎证型判别模型的构建,并对特征症状进行权重计算,其权重计算结果与现行的证候特征相比,将有助于证候表述的完善。通过以上分析,可以看出决策树模型具有较好的解释性,故在中医诊断领域应用较广,但该方法存在过拟合现象,这将是今后研究过程中需要解决的问题。

3 K近邻算法

K近邻(K-nearest neighbor, KNN)分类算法假设一个样本在特征空间中的K个最相似(即特征空间中最近邻)的样本中大多数属于某一类别,则该样本也属于这个类别^[22]。该算法有3个基本要素:K值、距离度量和分类决策规则。其中K值的选择与结果密切相关,K值较小意味着只有与输入实例较接近的训练实例才会对预测结果有作用,但容易发生拟合;而K值较大时,学习的估计误差减少,但近似误差增大,此时与输入实例较远的训练实例也会对预测起作用,可导致预测发生错误。一般来说,K值常选用一个较小数值,通常采用交叉验证方法选择最佳K值^[23]。该算法的特点为:首先,无需参数估计与训练,比较简单有效,精度高,且对噪声不敏感;其次,因KNN依靠周围有限的邻近样本,因而对类域交叉或重叠较多的待测样本来说,KNN相对更理想,也更适合多分类问题。但KNN分类算法也存在着解释性较差、计算量大、当样本不均衡时可能导致结果偏差等问题。

近年来,针对本算法的不足之处,研究者不

断进行改进,并衍生出了一系列的算法,如K-D树KNN算法、快速KNN算法、ML-KNN算法等。其中,ML-KNN模型为近年来典型的多标签学习算法,具有简单、易行、错误率较低等优点,是在传统的单标签KNN模型基础上结合贝叶斯算法发展起来。中医的临床表征(证候)往往不会单一出现,而是彼此联系交织,即涉及到证型往往不是单一的,而是兼夹的。ML-KNN算法可将与诊断相关的多个标签作为一个整体处理,保留了各标签之间的相关关系,因而也更符合中医核心的整体思维。王刚^[24]对ML-KNN算法进行了改良,并将该改良的方法用于401例帕金森病患者的病例数据分析,分析结果提示该方法在对帕金森的中医诊断中取得了良好的效果。辛基梁等^[25]研究中医健康状态辨识中的多标记分类方法,与单标记方法KNN相比,ML-KNN性能更佳。Liu G P等^[26]通过设计中医冠心病问诊标准化量表,并利用ML-KNN技术构建了基于中医问诊的诊断模型,并与RankSVM、BPMLL及KNN相比较,前者具有更好的预测效果,并且基于ML-KNN算法建立的冠心病诊断中医辨证预测模型,可同时解决多证素兼夹的标记问题。可见,与单标签KNN相比,基于多标签的ML-KNN算法更适合中医证候诊断过程中存在的多标签即证素、证型等兼夹问题。

4 人工神经网络

人工神经网络(artificial neural network, ANN)是模仿生物神经系统中神经元的一种数学处理方法。ANN由大量处理单元(神经元)互联组成非线性大规模自适应动态系统,依靠系统的复杂程度,通过调整大量单元间相互连接关系,从而达到处理信息的目的^[27]。ANN模型的性能主要与神经元的特性、神经元之间相互连接形式以及为适应环境而改善性能的学习规则有关。ANN构成原理与功能更接近人脑,擅于适应环境、总结规律以处理模糊复杂、推理规则不明确的问题,不用考虑各变量之间是否独立及满足正态分布等条件,并且ANN可给出结构参数,这也与传统的统计分析不一样^[28]。而在此基础上进行算法拓展,进一步兴起了深度学习,它运用分层抽象的思想,被应用在图像识别、语音识别等领域,具有更高的识别精度^[29]。然而,因ANN具有黑箱运行的特点,

ANN在解释推理过程和推理依据及其存储知识的意义时存在一定困难。目前,常见的神经网络主要有向前神经网络、BP神经网络、卷积神经网络、Hopfield神经网络等。

临床上,证候之间、证候与诊断结果之间、证候与方药之间的关系具有非线性、复杂性、模糊性、非定量的特点,因而ANN比较符合中医辨证理论和中医数据处理的运用要求^[30]。辛基梁^[31]采用“中医证素辨证系统”作为病例采集工具,采用基于神经网络的多标记分类算法对1146例病例进行结果预测,其平均精度可达到0.79。岳桂华等^[32]研究高血压中医证候与分级的关系时,运用基于共轭梯度算法的BP神经网络技术建立模型,该模型对高血压分级的平均识别率可达75%,说明BP神经网络技术应用于高血压中医证候与分级研究具有方法学上的可行性。王娟等^[33]应用多元数据统计和神经网络挖掘方法构建基于多系统理化指标信息的慢性心力衰竭(CHF)血瘀证诊断模型,筛选出有统计学意义的指标信息20项,测试样本的准确率达到82.4%。由此可见,ANN表现出极大的灵活性和自适应性,因此人工神经网络技术可运用于智能中医辨证系统构建过程中^[34]。但人工神经网络技术具有难解释性等缺陷,难以显现隐藏在网络中的知识。为进一步适应医学领域,可以使用决策树、规则提取等技术对人工神经网络技术进行改进,提高系统的可理解性。

5 贝叶斯网络

贝叶斯网络(Bayesian networks)又称贝叶斯信念网络,是由Pearl J^[35]在1988年提出,现已成为近些年的研究热点。贝叶斯网络作为一种更高级、应用范围更广的贝叶斯分类算法,其主要思想是将概率统计应用到复杂领域中以进行不确定性推理及数据分析,并充分考虑变量间的依赖关系,能更好地提高分类正确率,故被认为是表达不确定知识和推理的最有效理论模型之一^[36]。贝叶斯网络的结构为一个有向无环图,主要由节点和连接这些节点的有向边构成。其中,节点表示已知或未知的随机变量;节点间的有向边记录了两节点间的条件概率,以此可类推表示所有节点间的关系^[37]。但因贝叶斯网络的建造是一个复杂的任务,故需要该领域专家的参与并不断完善。另外,因

各节点间关系复杂,条件概率表存在着难以计算且不便使用的问题^[38]。

贝叶斯网络能充分利用先验知识,并用图表来显示变量之间的关系和解释因果链接关系,可以处理不完全、不精确、有噪声的信息,因而也被多次运用到中医诊断问题上^[39]。吕胤^[40]分别运用决策树、支持向量机和贝叶斯网络3种数据挖掘分类方法对高血压病的临床病例进行分类处理,发现贝叶斯网络分类算法准确率最高,耗时最短。徐全壹等^[41]比较分析贝叶斯网络和聚类分析在构建肾虚证数学诊断模型中的差异,发现聚类分析能大致构建出肾虚证症状的主要结构系统,但可能会丢失很多重要的特性,而贝叶斯网络分析结果提供了一个清晰的整体肾虚证症状系统,并成功地检测到4种不同的肾虚证症状类型,故认为贝叶斯网络可更有效地分析出肾虚证症状系统的不同层次。徐璿等^[42]在专家辨证和先验知识的基础上运用贝叶斯网络对冠心病患者四诊证候信息进行分类识别,发现该模型对本病5个不同证型的识别率较高,因而认为该方法在冠心病中医证候分类客观化研究中具有较好的应用前景。可见,运用贝叶斯网络可以揭示症状之间及症状与证型之间的关系,并归纳出该证型具有代表性的症状,这将不仅有助于证候诊断的标准化和规范化,还可辅助诊疗提供参考。

6 AdaBoost算法

AdaBoost算法是一种集成算法,即将多个机器学习算法构成集成分类器来完成学习任务^[43]。该算法的理论基础是假设存在“弱”分类器,这种弱分类器的预测能力(分类正确性大于0.5)仅比随机猜测准确一点,当其个数趋于无穷个数时,最终形成一个预测错误率很低的“强”分类器^[44]。作为一种算法框架,AdaBoost可用于绝大多数的机器学习算法以提高原算法的预测精度,故被评为数据挖掘十大算法之一。它的自适应性源于前一个基本分类器分错的样本会得到加强,加权后的全体样本再次被用来训练下一个基本分类器。同时,在每一轮中加入新的弱分类器,直到达到某个预定的足够小的错误率或指定的最大迭代次数^[45]。AdaBoost算法具有理论扎实、不需要先验知识、能够显著改善子分类器预测精度等优点^[46-47],其在解

决各行业应用问题中都获得了较好的结果。但在运用AdaBoost集成算法的过程中还需要解决如何得到若干个个体学习器和如何选择一种合理的结合方法的关键问题。

目前, AdaBoost算法在中医领域的应用研究不多,主要集中在舌诊、面诊等方面。胡申宁^[48]根据传统的中医理论,将AdaBoost算法融合到舌色、苔色的判别中,先在HSV[色调(H),饱和度(S),明度(V)]颜色空间中对舌象进行特征提取、降维,提取有效的舌象特征,而后通过AdaBoost算法把一系列弱分类器提升为强分类器,实现了具有较好鲁棒性的舌体自动分割。吴墩华^[49]在中医面诊规范化研究中,运用AdaBoost和FCM聚类算法对人脸特征进行粗定位,并在这些矩形区域中利用聚类获得它们的初始区域,在此基础上以面色脏腑分属图上特征点的颜色作为面色特征,获得了84.6%的识别率。此外,也有将AdaBoost算法运用在中医证候规范化研究。佟旭^[50]尝试运用AdaBoost算法构建糖尿病肾病患者的多标签的证型分类模型,其分类准确率接近98%,得出该方法的算法结构和原理对于解决中医多维度数据具有良好的优势。可见,该方法目前主要应用于中医望诊研究领域,其特点是可以综合多种算法的优势,但具体如何针对中医数据特点实现更好地结合亦需进一步研究。

中医诊断的基本原理为“司外揣内”,通过望、闻、问、切方式获得患者的症状、体征,从而推测导致疾病发生的原因及当前阶段的证候特点,又称之为“因发知受”^[51-52]。中医诊断过程是一个多源信息的获取和处理、整合的过程,由于传统中医诊法主要凭借医生的个人经验来辨析与判别病症,因而诊断的结果不可避免地受到医生主观意识的影响。近年来,对中医四诊的客观化和规范化研究方兴未艾^[53]。随着人工智能的发展和多学科融合交叉并广泛应用于中医证候诊断的研究,中医诊断客观化、规范化研究得到了有力的推动,通过机器学习可寻找中医大数据下的隐藏知识和规律。然而面对中医领域存在的数据种类繁多、结构多样并且缺乏统一标准的问题,如何保证数据集的质量是目前面临的巨大挑战^[54]。今后在促进机器学习运用于中医研究领域时,需针对中医诊疗中的各项问题制定统一的标准并规

范数据收集。

参考文献:

- [1] 夏淑洁, 李灿东. 中医诊断现代化与健康状态辨识[J]. 中医杂志, 2019, 60(4): 277-279.
- [2] 王忆勤. 中医诊断技术发展及四诊信息融合研究[J]. 上海中医药大学学报, 2019, 33(1): 1-7.
- [3] MA T, TAN C, ZHANG H, et al. Bridging the gap between traditional Chinese medicine and systems biology: the connection of cold syndrome and NEI network [J]. *Molecular Biosystems*, 2010, 6(4): 613-619.
- [4] 王永炎. 珍视中医原创思维, 调整学科方向[J]. 中国中医基础医学杂志, 2015, 21(11): 1337-1338.
- [5] DEO R C. Machine learning in medicine [J]. *Circulation*, 2015, 132(20): 1920-1930.
- [6] CORTES C, VAPNIK V. Support-vector networks [J]. *Machine Learning*, 1995, 20(3): 273-297.
- [7] 黄啸, 浦小祥. SVM核函数的研究及其在语音激活检测中的应用[J]. 苏州大学学报(工科版), 2008, 28(3): 56-59.
- [8] 王文剑, 门昌骞. 支持向量机建模及应用[M]. 北京: 科学出版社, 2014.
- [9] 冯帆, 倪中新. 基于支持向量机的高频金融时间序列预测[J]. 应用数学与计算数学学报, 2017, 31(3): 265-274.
- [10] 许明东, 马晓聪, 温宗良, 等. 支持向量机在高血压病中医证候诊断中的应用[J]. 中华中医药杂志, 2017, 32(6): 171-174.
- [11] WANG J, HE Q, YAO K W, et al. Support vector machine (SVM) and traditional Chinese medicine: syndrome factors based an SVM from coronary heart disease treated by prominent traditional Chinese medicine doctors [C]. *International Conference on Natural Computation*, IEEE: 2009.
- [12] SUN Z Q, XI G C, YI J Q. Differentiation of syndromes with SVM [A] / *Advances in Neural Network [C]*. ISNN, 2006: 786-791.
- [13] 任冷, 周维民. 针对非平衡多分类问题SVM算法的优化研究与应用[J]. 电脑知识与技术, 2016, 12(5): 218-220.
- [14] QUINLAN J R. Induction on decision tree [J]. *Machine Learning*, 1986, 1(1): 81-106.
- [15] PODGORLEC V, ZORM M. Decision tree learning [M]. Springer Science Business Media New York, 2013. https://doi.org/10.1007/978-3-642-27737-5_117-2.
- [16] 马梦羽, 沈璐, 文天才, 等. 数据挖掘技术在中医诊疗数据分析中的应用[J]. 中国中医药信息杂志, 2016, 23(7): 132-136.
- [17] 田艳鹏, 丁学义, 朱羽硕, 等. 基于决策树和神经网络的高血压病痰湿壅盛证诊断模型研究[J]. 中华中医药杂志, 2018, 33(8): 373-378.
- [18] 苏翀, 任瞳, 王国品, 等. 利用决策树建立慢性阻塞性肺病中医诊断模型[J]. 计算机工程与应用, 2019, 55(3): 230-

- 235.
- [19] 刘广, 孙艳秋, 裴媛. 基于C4.5决策树算法的中医胃炎实验数据分类挖掘研究[J]. 中华中医药学刊, 2016, 34(12): 144-147.
- [20] 吕航, 杨秋莉, 杜渐. 基于决策树预测糖尿病合并冠心病患病风险的中医人格体质特征研究[J]. 南京中医药大学学报, 2017, 33(6): 639-642.
- [21] 蔡晓路. 基于随机森林的类风湿关节炎证型判别模型研究[D]. 北京: 北京中医药大学, 2016.
- [22] KELLER J M, GRAY M R, GIVENS J A. A fuzzy K-nearest neighbor algorithm [J]. IEEE Transactions on Systems Man & Cybernetics, 2012, 15(4): 580-585.
- [23] PRASATH V B S, ALFEILAT H A A, LASASSMEH O, et al. Distance and similarity measures effect on the performance of K-nearest neighbor classifier: a review [J/OL]. Research Gate: 2017. <https://www.researchgate.net/publication/319135278>.
- [24] 王刚. 多标签学习及其在帕金森中医诊断中的应用[D]. 南京: 南京大学, 2014.
- [25] 辛基梁, 张佳, 李绍滋, 等. 中医健康状态辨识中的多标记分类方法研究[J]. 中华中医药杂志, 2019, 34(9): 3952-3955.
- [26] LIU G P, LI G Z, WANG Y L, et al. Modelling of inquiry diagnosis for coronary heart disease in traditional Chinese medicine by using multi-label learning [J]. BMC Complementary and Alternative Medicine, 2010, 10: 37. DOI: 10.1186/1472-6882-10-37.
- [27] GUPTA N. Artificial neural network [A]// CIRP Encyclopedia of Production Engineering [G]. Berlin, Heidelberg: Springer, 2014. https://doi.org/10.1007/978-3-642-20617-7_100023.
- [28] ZHANG Q, YANG L T, CHEN Z, et al. A survey on deep learning for big data [J]. Inform Fusion, 2018, 42: 146-157.
- [29] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.
- [30] 刘旺华, 洪净, 李花, 等. 人工神经网络在中医诊断信息化中的应用[J]. 湖南中医药大学学报, 2017, 37(7): 809-812.
- [31] 辛基梁. 人工神经网络在中医临床辨证模型研究中的应用[D]. 福州: 福建中医药大学, 2017.
- [32] 岳桂华, 温宗良, 杨靖, 等. 基于BP神经网络的高血压病中医证候与分级关系建模研究[J]. 中国中医基础医学杂志, 2012, 18(3): 326-327.
- [33] 王娟, 赵慧辉, 陈建新, 等. 基于神经网络技术分析的慢性心力衰竭血瘀证诊断模型研究[J]. 世界中医药, 2018, 13(9): 39-43.
- [34] 褚娜. 基于混合智能的中医辨证系统研究[D]. 上海: 上海交通大学, 2012.
- [35] PEARL J. Probabilistic reasoning in intelligent systems: networks of plausible inference [M]. San Mateo, CA: Morgan Kaufmann, 1988.
- [36] KWAN M, CHOW K P, LAW F, et al. Reasoning about evidence using Bayesian networks [A]// IFIP International Conference on Digital Forensics 2008: advances in digital forensics IV [C]. 2008: 275-289.
- [37] 敬瑞星, 卢健康, 赵鹏飞, 等. 基于贝叶斯网络的系统可靠性分析平台[J]. 计算机工程与应用, 2013, 49(4): 71-76.
- [38] 马梦羽, 沈璐, 文天才, 等. 数据挖掘技术在中医诊疗数据分析中的应用[J]. 中国中医药信息杂志, 2016, 23(7): 132-136.
- [39] 蓝璠, 王晶, 尤圣富, 等. 贝叶斯网络技术在中医证候研究中的运用[J]. 中国中医骨伤科杂志, 2015, 23(2): 70-72.
- [40] 吕胤. 基于数据挖掘方法的高血压病中医证型分类数字化研究[D]. 长沙: 湖南中医药大学, 2017.
- [41] 徐全壹, 邓珂, 耿直, 等. 贝叶斯网络和聚类分析在肾虚证中的比较研究[J]. 时珍国医国药, 2015, 26(6): 1530-1531.
- [42] 徐璿, 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.
- [43] FREUND Y, SCHAPIRE R E. A decision-theoretic generalization of on-line learning and an application to boosting [A]// European Conference on Computational Learning Theory [C]. 1995: 23-27.
- [44] 王玲娣. Adaptive Boosting算法及组合分类器的构建研究[D]. 无锡: 江南大学, 2018.
- [45] 曹莹, 苗启广, 刘家辰, 等. AdaBoost算法研究进展与展望[J]. 自动化学报, 2013, 39(6): 745-758.
- [46] 李鹏飞, 闫亚娣, 张凯兵, 等. 基于AdaBoost局部二值模式特征的色织物纹理分类[J]. 西安工程大学学报, 2018, 32(6): 670-677.
- [47] 佟旭. 基于复杂网络理论的糖尿病肾病辨证建模研究[D]. 北京: 北京中医药大学, 2016.
- [48] 胡申宁. 中医舌诊中舌色、苔色自动分类的研究[D]. 杭州: 浙江理工大学, 2010.
- [49] 吴墩华. 面向中医面诊诊断信息提取的若干关键技术研究[D]. 厦门: 厦门大学, 2008.
- [50] 佟旭. 基于复杂网络理论的糖尿病肾病辨证建模研究[D]. 北京: 北京中医药大学, 2016.
- [51] 夏淑洁, 李灿东. “因发知受”的思维原理剖析[J/OL]. 中国中医基础医学杂志, 2020: 1-8 [2020-07-15]. <http://kns.cnki.net/kcms/detail/11.3554.r.20200414.1328.002.html>.
- [52] 李灿东. 中医诊断学[M]. 北京: 中国中医药出版社, 2012: 4-5.
- [53] 夏淑洁, 周智慧, 李佐飞, 等. 四诊现代化研究原理与应用[J]. 天津中医药, 2020, 37(3): 259-265.
- [54] 张晓航, 石清磊, 王斌, 等. 机器学习算法在中医诊疗中的研究综述[J]. 计算机科学, 2018, 45(S2): 32-36.

【责任编辑: 贺小英】