

支持向量机法及其在中药研究中的应用

杨 岩^{1,3}, 肖佳妹^{1,3#}, 周 晋^{1,2,3}, 贺福元^{1,2,3}, 曾慧杰⁴, 杨岩涛^{1,2,3*}

1. 湖南中医药大学药学院, 湖南 长沙 410208

2. 中药成药性与制剂制备湖南省重点实验室, 湖南 长沙 410208

3. 湖南中医药大学 中医药超分子机理与数理特征化实验室, 湖南 长沙 410208

4. 湖南省林业科学研究院, 湖南 长沙 410004

摘要: 在中医药现代化进程中, 有效地结合现代科学技术手段对传统中医药进行更为合理的分析是一个重要研究方向。计算机科学技术与人工智能算法在中医药研究领域已取得长足发展, 尤其是分类算法所具有的快速鉴别和分析能力, 使其在复杂的中医药研究体系中发挥举足轻重的作用。对支持向量机分类算法及其结合其他算法在解决中医药相关领域问题方面及成果进行综述, 以相关智能算法为基础探讨中医药研究过程中的诸多难题, 为推动中医药现代化进程提供支撑。

关键词: 分类算法; 智能算法; 中医药; 支持向量机; 模式识别

中图分类号: R286 **文献标志码:** A **文章编号:** 0253-2670(2020)08-2258-09

DOI: 10.7501/j.issn.0253-2670.2020.08.037

Research on classification algorithm based on support vector machine in Chinese materia medica

YANG Yan^{1,3}, XIAO Jia-mei^{1,3}, ZHOU Jin^{1,2,3}, HE Fu-yuan^{1,2,3}, ZENG Hui-jie⁴, YANG Yan-tao^{1,2,3}

1. College of Pharmacy, Hunan University of Chinese Medicine, Changsha 410208, China

2. Hunan Key Laboratory of Druggability and Preparation of Chinese Medicine, Changsha 410208, China

3. Laboratory of Supramolecular Mechanism and Mathematic-Physics Characterization for Chinese Materia Medica, Hunan University of Chinese Medicine, Changsha 410208, China

4. Hunan Academy of Forestry, Changsha 410004, China

Abstract: In the course of the modernization of traditional Chinese medicine, the more reasonable analysis of the traditional Chinese medicine is an important research direction in combination with the modern scientific and technological means. The rapid development of the computer science and technology and the artificial intelligence algorithm in the field of traditional Chinese medicine research, especially the rapid identification and analysis ability of the classification algorithm, can play a very important role in the complex Chinese medicine research system. This paper reviews the support vector machine classification algorithm and its application and results in combination with other algorithms to solve the problems in the traditional Chinese medicine research, explore the application of the related intelligent algorithm in solving the problems in the research of traditional Chinese medicine, and provide support for promoting the modernization of Chinese medicine.

Key words: classification algorithm; intelligent algorithm; traditional Chinese medicine; support vector machine; pattern identification

中医药传承几千年, 是中华文明的重要组成部分, 其现代化研究虽已取得长足的进步, 但仍存在许多关键性问题亟待解决。中药本身化学成分复杂, 同时基原品种、产地来源、生长环境、采收季节、

收稿日期: 2019-12-13

基金项目: 湖南省重点研发计划项目(2018NK2041); 湖南省教育厅优秀青年基金项目(2015B172); 湖南省中医药科研基金资助项目(201494); 湖南省教育厅一般项目(2015C1039); 湖南省“十二五”省级药学重点学科开放基金项目(1007); 湖南中医药大学化学工程与技术一流学科建设

作者简介: 杨 岩(1995—), 男, 湖北武汉人, 在读硕士研究生, 主要从事中药药剂学、中药制药工程等研究。

Tel: (0731)88458223 E-mail: 1043624698@qq.com

***通信作者** 杨岩涛, 男, 博士, 副教授, 硕士研究生导师。Tel: (0731)88458223 E-mail: xdyt1@163.com

#并列第一作者 肖佳妹, 女, 硕士研究生, 主要从事中药学相关研究。E-mail: 1319286212@qq.com

加工炮制工艺等诸多因素^[1]形成了中药复杂多变的体系,外观相似的中药药理活性大相径庭,假冒伪劣的中药饮片常常被不法分子利用,严重影响中医药现代化的快速推进。传统鉴别手段需借助该行业相关专家以传统经验进行分析,但该方法存在着耗时长、效率低、准确率不能保证等一系列问题。人工智能算法的出现有效地解决了这一问题,同时进一步结合其他学科,发挥多学科联合应用的优势,取长补短,将大数据的分析模式与中医药整体观有机统一,可以从宏观上对中药进行综合鉴别判定分析,同时也可将之应用到符合传统中医药理论的中药配伍、中药药性药效、定性定量分析等领域,有力地推动了中医药现代化进程。

1 支持向量机(SVM)的定义及其在中药相关领域的应用

SVM 作为近 20 年来高速发展的人工智能算法,属于二分类模型,如果数据服从某个(固定但未知的)分布,要使机器的实际输出与理想输出之间的偏差尽可能小,则机器应当遵循结构风险最小化(structural risk minimization, SRM)原则,而不是经验风险最小化原则^[2]。通过事先选择的非线性映射将输入向量映射到高维特征空间,寻找最优超平面,然后在此空间中求解凸优化问题^[3],SVM 正是这一理论的具体实现,其中内核函数作为 SVM 的精髓,将原来线性不可分的样本可以在新的核空间下转变为线性可分。常见的核函数主要为线性核函数、多项式核函数、径向基核函数和 sigmoid 核函数。SVM 的基本结构见图 1,图中 $x_1, x_2, \dots, x_n$ 表示的是输入变量,经过 SVM 内积核函数计算,不考虑变换拉伸过程中的具体映射关系,得到输出

变量 y 。SVM 包括主要应用于模式识别与分类及差别分析等的支持向量分类(support vector classification, SVC)和主要用于数据的拟合并产生回归模型,用于相关预测等的支持向量回归(support vector regression, SVR)^[4]。SVM 是泛化能力很强的优质分类器,准确率也很高,SVM 算法的分类作用目前主要应用于中药鉴定、中药药性、中药活性成分、中药安全性评价、药物分析、制剂工艺及中药配伍规律等研究领域。

1.1 产地识别鉴定研究

中药产地众多,质量参差不齐,且“一药多源”现象广泛存在,质量监控困难。而 SVM 算法为中药产地识别及中药鉴定研究提供了方法,已有研究以近红外光谱技术为基础,通过测定内部特征值及外部特征值,结合 SVM 算法,成功识别黄连饮片^[5],并能快速鉴别枸杞子产地^[6]。以此延伸至“内外结合”的中药质量控制新模式,实现传统“辨状论质”经验,得以现代“察外知内”的再现^[7]。结合近红外光谱或者太赫兹时域光谱技术,能较为精准地鉴别中药,且具有自动识别的特点,而这种方法难以对紫外数据及中药指纹图谱数据进行分类鉴别。此外,该方法存在着缺乏对个体预测可信度的测量及整体预测准确性较差的问题。因此,欲将其更好地应用于中药鉴定领域需在其基础上进一步融合其他化学计量学方法。

1.2 中药药性研究

中药四性是中药药性理论的主要内容之一,是中药作用于生物体后出现的寒、热、温、凉 4 种生物效应的表征,长期指导着中医临床用药。目前对中药四性的研究主要集中在阐明其物质基础上。因此多采用药效学、药物化学等手段从药物成分结构方面寻找四性之奥秘,其研究方向也主要集中在运用现代科学技术方法预测药效成分与寒热药性的相关性。而支持向量机算法能够从初生物质这一难点入手对中药药性进行预测分析,有研究^[8-9]指出通过支持向量机建立寒热药性的数学判别模型,发现游离脂成分以及多糖水解成分与中药寒热药性存在明显的相关性,可依据模型对药物药性进行判别。但要对中药药性展开进一步研究则需要将多种初生物质结合进一步分析,借助多学科的知识方可充分发掘其药性规律。

1.3 中药毒副作用研究

中药在临床应用常被认为是安全有效且毒副作

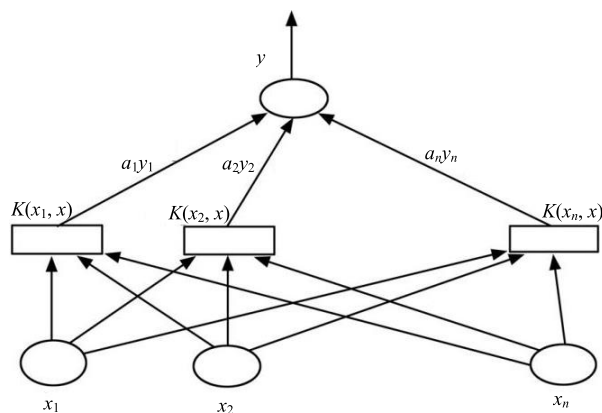


图 1 支持向量机原理

Fig. 1 Schematic of support vector machine

用小,但随着现代科学的不断发展及研究的不断深入,类似于何首乌具有肝毒性等中药的毒副作用不断被发现,给中药的合理应用提出了新的课题,即如何准确有效地识别中药中潜在的毒性成分,建立中药安全评价方法体系。SVM 凭借其强大的分类判别能力,结合外源数据库,构建合理的预测分析模型对具有潜在毒性的中药化学成分进行筛选,并在此基础上,结合系统药理学分析进一步阐明其毒性机制,在药物研发的早期阶段对化合物的毒性进行评价并将毒性较强的化合物尽早排除^[10],可有效提高药物开发的效率和成功率。如在中药肾毒性研究中,有研究^[11]借鉴化学药的定量-结构-活性关系(quantitative structure-activity relationship, QSAR)模型进行检测,结果显示 SVM 算法在自身预测准确率方面远高于 K-邻近算法。在其基础上,将肾毒化合物和无肾毒化合物转化为分子描述符,用 SVM 建立判别模型,用肾小管损伤模型^[12]进一步判别肾毒性模型筛查结果为阳性的中药成分,其准确性较高。后将代谢组学技术结合 SVM 研究肾毒性小分子代谢标志物^[13],设计出肾毒性判别模型,能够快速、高效地筛查中药中致毒性的成分,以及肾毒性生物标志物。此外,在山豆根^[14]神经毒性成分筛选中,采取了与肾毒性成分筛选相似的处理方式,且进行了验证。在肝毒性成分^[15]研究中,其模型较为复杂,以多种机器算法结合分子指纹构建的组合模型,整理出肝毒性位于前 10 的中药,为中药肝毒性的预测提供了合理的研究策略。在未来中药毒副作用研究过程中加强对致毒性成分的筛选判定,以达到祛除毒性成分,保留效应成分,更好地为中药临床合理用药提供技术支撑。

1.4 中药组方配伍研究

中药系多成分、多靶点作用,因此七情和合等配伍规律是中药组方基本理论体系的重要组成,但中药组方物质基础不明严重阻碍了中药配伍规律的现代化阐述。借助现代药理学的研究成果,筛选在临床上发挥作用的中药活性成分,回归预测治疗相关疾病的最佳药物配比已被证明是一条行之有效的途径。在大黄芩元治疗脑缺血配伍研究中^[16],研究者采用了均匀设计-药效学验证-数学建模相结合的方法,对组分配伍剂量进行优化以阐明临床配伍量效关系,结果显示该模型能够模拟大黄各配伍组治疗脑缺血的药效预测。此外,在原有配伍剂量条件下,结合 SVM 算法对中药复方进行了二次开发^[17],

通过模型预测和实验验证,多个配伍组的药效优于原方,这项研究成果将推动中药经典名方的进一步研究开发及现有中药品种的二次开发。

1.5 中药复方制剂工艺优化和质量研究

中药制剂工艺参数的优化是决定药品质量及成本的关键因素,采用现代化科学智能算法对制剂工艺数据进行处理将会获得更精确的优化结果。在优化中药提取工艺所得到的数据集中,比较决策树和 SVM 算法的优劣性,结果表明决策树算法的结果可理解性较好,但 SVM 分类算法有更高的精度^[18]。通过构建相关分类器,在总黄酮提取工艺中得到实际应用^[19],其预测模型精确且泛化能力强,获得了较好的优化结果。在中药生产过程中,提取液浓度是一个非常重要但又难以测定的参数,常规检测方法主要是人工检测、离线检测,这些方法往往会造成每个批次之间具有主观差异^[20]。但在线检测或线边检测是一个良好的解决方案,基于 SVM 建立提取液浓度的在线自适应软测量模型^[21],实现了对提取液浓度的在线测量及控制,对中药生产加工的全局统筹及对物质基础动态变化规律的认识,可从根本上提高中药产品的质量^[22]。如能进一步基于 SVM 建立中药质量标志物(Q-marker)体系的在线检测及控制,必将有效解决中药生产过程中物质传递稳定性等影响到成药质量的重要问题。同时在中药不同剂型质量分析中引入 SVM 算法,将为其质量研究提供更为精确的数据。如针对苦碟子注射剂的研究中^[23],将受试者工作特征曲线(receiver operating characteristic curve, ROC)结合指纹图谱和 SVM 的指纹谱-ROC-SVM 策略充分发挥了 3 者之间优势,能准确地将合格品、不合格品、高温加工样品、高照度加工样品区分开,表明该分析策略具有较强的实用性和准确性,可以进一步应用于中药注射液的监测或预测,大大提高中药注射剂的临床安全性。

1.6 中药有效成分研究

有效成分是中药治疗疾病的物质基础,而中药复方多成分在炮制、提取、浓缩、干燥、制剂加工等生产过程中,成分-成分、成分-辅料之间有可能形成非共价键形式结合的超分子体系,这也给中药物质基础研究带来了复杂性,如能阐明中药复方成分在生产过程中的结合与变化规律,将极大地推动中药药效物质基础的研究。有研究^[24]建立了分类预测模型,对中药方剂在不同浓度下均形成聚集体这一现象进行了研究,精确地对中药有效成分三维结

构与性质数据库的分子进行了预测,为中药有效成分研究提供了新的思路。

2 SVM 与其他算法的结合及在中药研究中的应用

SVM 作为一种传统的二分类算法,不仅能解决线性分类,还可以使用核函数有效地进行非线性分类。但由于 SVM 是借助二次规划求解支持向量,对大规模样本难以实施,在解决多分类问题上有困难,且由于 SVM 主要参数的选择能够在很大程度上影响分类性能和效果,目前参数优化缺乏理论指导^[25]。因此当前针对复杂中医药体系研究往往都是与常见化学计量法相结合,以此达到分类识别预测等效果。化学计量学类似于生物计量学及经济计量学,主要使用统计学和数学工具从多元化的化学数据中提炼有效的相关信息,在探索多种中药物质相关联的数据、药材质量控制、多成分变化规律等中医药研究领域中发挥重要的作用。其本质是将复杂数据的变化删繁就简,从中提取更为有效且简洁的信息。针对中药复杂体系中具备多个特征数据的研究,多种算法交叉结合应用,主要遵循先降维,再回归,最后建模预测分类分析的原则以解决问题。

2.1 SVM 与主成分分析 (principal component analysis, PCA) 的结合

PCA 在化学计量法中属于经典算法,主要是对特征光谱数据进行降维处理,可以在数据可视化的同时降低分类的时间成本。利用近红外漫反射光谱 (near infrared diffuse reflectance spectra, NIRS) 法,结合 PCA 和 SVM 算法,在中药鉴定、中药构效关系及中药药性识别研究中得到了广泛的应用,也取得了良好的效果。PCA-SVM 联用不仅能满足指标成分的筛选要求,同时能充分发挥在解决小样本数据及非线性等实际问题中的优势。

中药 NIR 光谱 PCA-SVM 鉴别模型^[26]被建立并应用在 6 种树脂及其他类中药的鉴别过程中,实现了对中药的无损、快速鉴别。以 PCA-SVM 作为分类鉴别算法,结合比例-积分-微分 (proportion integral differential, PID) 算法研制了一套智能化自动检测系统,实现了对气室温度的智能控制,能够对红参、白参、西洋参、太子参、三七进行较高精度的鉴别^[27],这些研究都推动了 SVM 与 PCA 结合在中药鉴定学科中的应用。阐明中药产生药效的物质基础是实现中药现代化研究的关键,也是广大中药研究者长期以来努力的方向,通过计算机算法模型对活性物质进行筛选预测已被证明是一条行之有效的途径。有

研究表明在活血化瘀类中药中^[28-29],选取显著代表化合物空间结构的描述参数,建立了相关的定量预测模型,此方法实现了对未知中药相应性质的分类和预测,对于药物的筛选或新药的合成都具有重要的参考价值。在探讨中药红外光谱与药性的相关性研究中^[30],采用 PCA 对中药光谱数据降维之后,选择主成分作为 SVM 输入变量,构建得到平性药和非平性药的识别模型,可较好地地区分平性药与非平性药,为药性物质基础研究提供了一种新的研究思路。如能适当增加不同提取部位的红外数据以增加 SVM 计算所需要的特征变量,有可能提高中药药性的识别率。

2.2 SVM 与层次分析法 (analytic hierarchy process, AHP) 的结合

AHP 是一种系统评价方法,常用于权重决策分析^[31],属于定性与定量的结合,其计算方法有 4 种,即几何平均法、算术平均法、特征向量法、最小二乘法^[32]。该分析方法在优化权重赋值时常可以摆脱人为主观影响,故应用广泛,已与 SVM 结合应用于中药方剂的优选。其具体结合思路首先利用 AHP 计算出指标权重,其次依权重来处理数据;最后用 SVM 对数据进行分类^[33],两者结合能够更加全面、科学、客观地反映指标层对实验结果的影响。中药组合用药是新药研究的一个热点,根据方剂配伍理论,在中医传承辅助平台构建中药成方制剂的结构化数据库,经过 AHP 映射处理,在 SVM 模型下对方剂-疾病与方剂-证候数据集进行有效预测^[34],将中医证候与中药方剂有力地关联起来,为进一步用现代科学语言阐明中医证候概念提供了支撑。

2.3 SVM 与最小二乘法的结合

最小二乘支持向量机 (least squares support vector machine, LS-SVM) 回归模型是经典 SVM 的改进,其算法是一种人工智能技术结合,将线性思想引入向量机,用以处理非线性问题,通过核函数降低了计算复杂性,加快了求解速度,主要应用在光谱数据的建模分析。有研究^[35]对银杏叶提取物指纹图谱及其清除氧化自由基能力的训练集和测试集建立 LS-SVM 预测模型,能够对银杏叶抗氧化活性进行测定。并且在根据化学计量学理论建立相关预测模型时,发现中药的生物活性是复杂体系的整体作用。该模型的结合预测效果优于目前普遍单独使用的误差反向传播神经网络和偏最小二乘回归。不仅如此,凭借对指纹图谱的多维色谱数据区分和预

报能力, LS-SVM 得到更好的分类效果^[36]。在羌活的指纹图谱研究中, 并未将共有峰作为输入变量, 而是将全部色谱数据进行分析处理, 其预测效果优于单独采用共有峰的预测, 更为重要的是这种方法的研究思路与中医药的整体观念不谋而合, 是一种研究中药活性物质及确定中药 Q-marker 的重要方法。

2.4 SVM 与遗传算法 (genetic algorithm, GA) 的结合

在近红外定量分析常见的建模方法中, LS-SVM 有建模速度快、优化参数少、泛化能力强等优点而备受关注^[37]。而针对 LS-SVM 参数优化中的启发式算法则更多的被发掘应用, 如蚁群算法、GA、粒子群算

法等。通常情况下, 因光谱数据无效数据较多, 干扰信息庞杂, 故如何对样本进行有效筛选, 前处理是降低 LS-SVM 处理成本的重要条件。而基于 GA 的特征波长筛选可以在有效消除紫外 (UV) 光谱数据冗余信息的同时, 保留光谱有效信息, 从而降低模型的复杂度。在秦皮提取液快速定量分析中^[38], 径向基核函数 (radial basis function, RBF) 凭借其拟合精度与预测性能均较好而非常适合作为建模函数, 在其基础上进一步结合 GA 算法, 成功地对秦皮甲素、秦皮乙素、秦皮素等进行含量测定。具体结合使用方法见图 2^[39]。同样在不同威灵仙^[40]的代谢物分析中, GA-SVM 模型显现出优秀的预测性能。

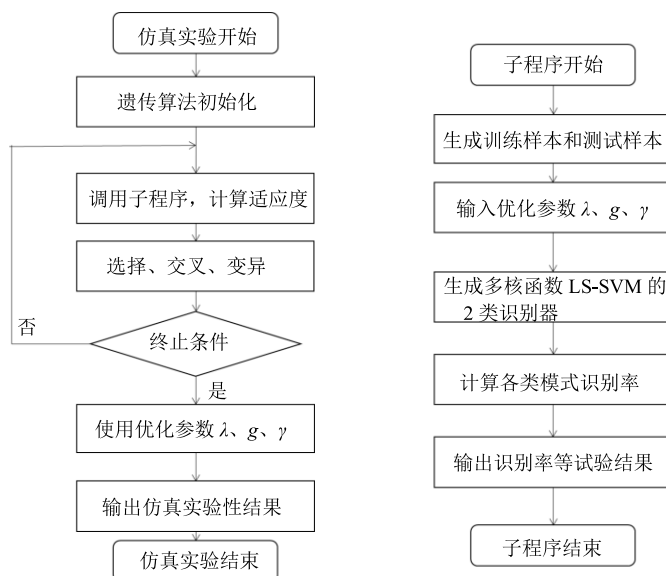


图 2 仿真实验及其子程序

Fig. 2 Simulation experiment and subprogram

2.5 SVM 与粒子群算法的结合

粒子群算法 (particle swarm optimization, PSO) 模拟鸟群捕食行为, 其算法概念简单, 控制参数少, 易于实现, 同时兼有进化计算和群智能优化的特点, 通过个体间的协作与竞争, 可以实现对复杂空间最优解的搜索^[41], 该算法不仅具有很强的全局搜索能力, 而且是解决整数非线性优化、非线性连续优化和组合优化等问题的有效工具。利用 LS-SVM 多输入、多输出、非线性的特点, 可进一步结合粒子群算法对复杂生产过程进行全局预测优化。在中药糖析出建模相关研究中, 以 SVM 建立相关回归模型, 结合多种粒子群算法对其参数进行寻优, 解决模型参数较多或范围较大的问题。对山茱萸药材中多个质控指标 (水分、浸出物、马钱苷和莫诺苷) 的定

量分析模型中^[42], 通过比较基于偏最小二乘回归法 (partial least squares regression, PLSR) 和人工神经网络 (artificial neural networks, ANN) 建立的 NIR 模型, 得到 PSO-LS-SVM 模型能对山茱萸药材质量进行更为精确的控制。除在中药质量控制中的应用外, 这种模型还用于建立红花^[43]提取过程关键质控指标的定量分析模型。该优势在于避免使计算陷入局部最优, 且容易实现, 调整参数较少, 因此 SVM 与粒子群算法相结合有可能用于解决中药从药材到制剂的动态迁移过程中 Q-marker 体系的研究, 进一步充实中药 Q-marker 体系的建立与研究。本文以邵永^[44]研究中的多种群粒子群算法 (multipopulation particle swarm optimization, MPSO) 对 SVR 模型参数进行优化流程举例, 见图 3。

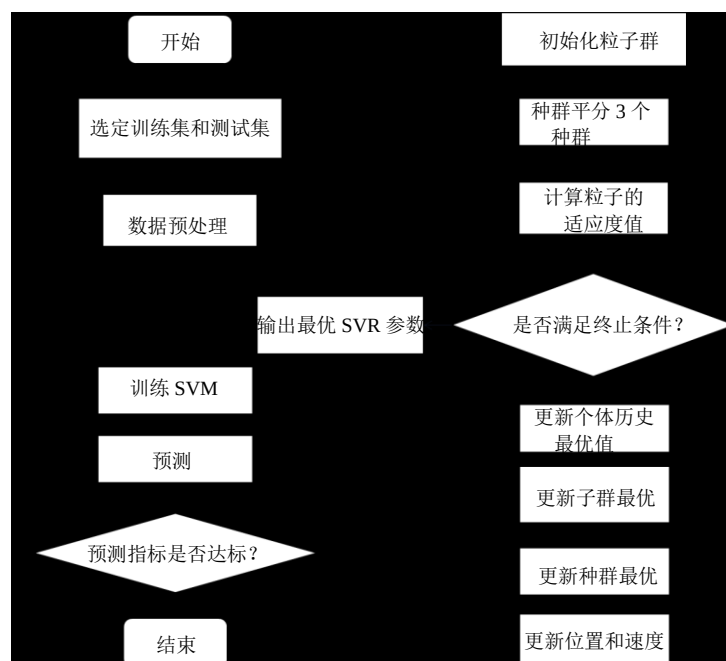


图 3 基于 MPSO-SVR 参数选择流程

Fig. 3 Flow chart of selection based on MPSO-SVR parameters

2.6 SVM 与人工神经网络的结合

在中医药各相关研究领域, SVM 和人工神经网络 (artificial neural network, ANN) 作为数据挖掘算法均有非常广泛的应用。ANN 是一种以生物学中神经网络为基本原理, 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的处理机制的一种数学模型。其实际上是一个有大量简单节点相互连接传递的复杂网络, 能够进行复杂的逻辑运算的系统, 具有高度的非线性、非局限性、非定性和非凸性 4 个基本特征^[45]。有研究^[46]利用基于 SVM (径向基函数与多项式函数 2 种算法) 和 ANN [一对多网络 (ACON) 与一对一网络 (OCON) 2 种结构的网络] 对心血管疾病的中医临床信息和证候类别之间的关系进行分析, 建立模型, 观察其证候预测的准确性, 反映一定的客观规律, 实现中医诊断规范化, 这也给 ANN 与 SVM 结合应用在中医药研究领域提供了参考。

2.7 国外 SVM 与其他算法结合的研究

目前国外针对 SVM 的研究很多, 根据数据处理难度衍生出很多相关算法的结合, 但针对中药方面的先进算法却是少数, 抛开各专业局限性, 根据分类预测 2 大核心功能, 如能够综合借鉴其优点, 将更好地运用在复杂多变的中药领域。关于 SVM 有几种常见的结合: (1) 多实例 SVM (MI-SVM)^[47]

广泛应用于各种实际任务例如药物和图像分析。(2) GA-共轭梯度 (CG)-SVM^[48]目前用于药物动力学预测及相关毒性预测^[49]。(3) 特征变换和 SVM 的联合学习 (F-SVM)^[50]属于扩展为联合学习非线性变换的分类器, 能大幅度提高图像分类性能。(4) PmSVM^[51]是一种用于一般非线性核的线性回归 SVM 框架, 实现了最高的学习速度和最高的准确性。前两者在药理学等专业有所应用, 后 2 种方法则是 SVM 自身做出的改进, 相比于 SVM, 都实现了更为精确的演算。若将其正确的引用至复杂的中医药行业, 不离开分类预测的大框架下, 必将加快中医药现代化进程。

3 基于 SVM 的中医药研究新思路

中药质量标准决定着中药的安全性与有效性, 完善的质量标准将会促进中医药产业走向现代化^[52]。常规使用的中药质量表征方法有薄层色谱法、高效液相色谱法、气相色谱法、近红外光谱法等, 但这些分析方法都存在数据集庞大且不明确、特征数据有效关联程度不够等一系列问题, 尤其是近红外光谱技术的局限性, 对不同类成分的研究缺乏有效分类技术手段, SVM 的应用及其他分类算法的多学科交叉融合对中药质量标准研究具有重大指导意义。

中药质量应全面遵循中医药基础理论, 应全面

包括中药药性、药效、复方配伍及制剂加工等相关内容,才能够支撑中药现代化、规范化研究。鉴于此,结合 SVM 在中药相关研究领域中的应用,提出以下 2 种研究思路:(1)基于中药 Q-marker 概念^[53],凭借 SVM 分类判别算法泛化能力强、分类预测能力强、客观真实且不需要大样本数据支撑等特点,将其融合到中药 Q-marker 的研究中来,有望获得更精准的 Q-marker 及相关质量标准,更好地推动 Q-marker 研究。在元胡止痛片^[54]Q-marker 鉴别中,以 LS-SVM 建立成分和功效之间的相关性模型,最终选择了 3 种成分(四氢巴马汀、羟乙基脯氨酸和欧前胡素)作为最小组合,既可以作为区分伪品的指标成分,又可以作为指示生物活性强度的 Q-marker,成功改善该药片的质量控制。为 Q-marker 结合机器算法这一理念提供了有力证据,也证明基于 SVM 算法对 Q-marker 研究是一条行之有效的中药质量控制体系策略。(2)总量统计矩法具有整体、宏观和模糊分析等特点,可以对指纹图谱进行定性定量描述,是一种综合性、可量化的鉴定方法^[55]。可在获得总量零、一、二阶矩等有相应意义的总量统计矩参数基础上,进一步运用 SVM 算法对其进行优化,有望为中药复方制剂加工过程中多成分质量迁移动态规律的研究提供支撑。在总量统计矩现有的参数体系,充分结合现代科学技术,发挥机器算法优势,进一步拓展总量统计矩的应用。

如何建立符合中医药特色的中药质量控制体系一直是中医药走向现代化的关键。SVM 作为辅助中药研究的有效手段,其潜力不断被挖掘以提供更精确的数据信息。SVM 算法属于有监督模式的化学模式识别方法,为新兴的人工智能技术,理论还在不断完善和发展中,具有巨大的研究潜力。SVM 能够对色谱图信息进行数据化处理,使其数据整合后被计算机识别,现有研究表明其预测分类效果优于其他分类算法,为中药现代化研究提供了一个明确的方向。但目前 SVM 在中医药领域的应用尚存在结构不完善、结果不够精准等问题。如能将该算法与其他更多的专业算法相结合,对其进行优化,将对阐明中医药理论内涵提供逻辑更为精确的技术支持。

参考文献

[1] 孙立丽,王 萌,任晓亮.化学模式识别方法在中药质

量控制研究中的应用进展 [J]. 中草药, 2017, 48(20): 4339-4345.

[2] 李晓宇,张新峰,沈兰荪.支持向量机(SVM)的研究进展 [J]. 测控技术, 2006, 25(5): 7-12.

[3] 程 跃,程文明,郑 严.支持向量机在中药浓缩浓度软测量中的应用 [J]. 计算机工程与应用, 2010, 46(5): 240-242.

[4] 龙 伟,刘培勋,曾 平,等.支持向量机与中药研究 [J]. 中草药, 2008, 39(10): 1441-1445.

[5] 陶梦琳,顾文涛,汪子青,等.基于支持向量机的黄连饮片产地识别研究 [J]. 中草药, 2015, 46(21): 3173-3178.

[6] 杜 敏,巩 颖,林兆洲,等.样品表面近红外光谱结合多类支持向量机快速鉴别枸杞子产地 [J]. 光谱学与光谱分析, 2013, 33(5): 1211-1214.

[7] 陶梦琳,顾文涛,侯珂惠,等.基于支持向量机的“内外结合”中药质量控制新模式探索 [J]. 中国药房, 2016, 27(1): 118-121.

[8] 王 鹏,王振国,薛付忠,等.基于支持向量机法的中药性状与药性相关性研究 [J]. 江西中医药, 2012, 43(7): 65-68.

[9] 陈永新,李 峰,孙稚颖,等.基于游离脂成分的 20 种中药寒热药性支持向量机分析 [J]. 辽宁中医杂志, 2011, 38(1): 127-129.

[10] 雷太龙.基于机器学习方法的药物毒性的理论预测研究 [D]. 杭州:浙江大学, 2017.

[11] 王新洲,朱永亮,金若敏,等.基于数学模型的药物成分肾毒性预测 [J]. 中国新药杂志, 2014, 23(13): 1565-1568.

[12] 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.

[13] 郭雪君,谷彩云,许妍妍,等.肾毒性小分子代谢标志物在 5 种中药毒性评价的适用性研究 [J]. 药物评价研究, 2017, 40(4): 472-478.

[14] 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.

[15] 吴淇慧.中草药肝毒性预测模型的构建及其应用研究 [D]. 广州:广州中医药大学, 2018.

[16] 吴纯伟,宿爱山,陈 超,等.基于支持向量机的 5 种大黄苷元治疗脑缺血的配伍研究 [J]. 中草药, 2015, 46(22): 3377-3381.

[17] 李淑贤,陈 超,王淑美,等.支持向量机回归算法用于筛选脑得生抗血栓有效部位 [J]. 中药材, 2011, 34(7): 1078-1080.

[18] 朱恒民,刘文杰,王宁生.数据挖掘技术在优化中药提取工艺中的应用 [J]. 计算机与应用化学, 2006, 23(3):

- 233-236.
- [19] 曾 燕, 成新文, 陈欲云, 等. 最小二乘支持向量机优化芹菜总黄酮提取工艺 [J]. 食品研究与开发, 2015, 36(8): 49-52.
- [20] 梁 涛, 李长朋, 张军伟, 等. 支持向量机的中药提取浓度软测量 [J]. 计算机与应用化学, 2013, 30(11): 1371-1374.
- [21] 于晓明. 支持向量机及其在制浆过程重要参数软测量中的应用研究 [D]. 西安: 陕西科技大学, 2012.
- [22] 王启超. 基于组合核函数支持向量机的软测量技术及其应用研究 [D]. 赣州: 江西理工大学, 2011.
- [23] Yang B, Wang Y, Shan L, *et al.* A Novel and practical chromatographic “fingerprint-ROC-SVM” strategy applied to quality analysis of traditional Chinese medicine injections: Using KuDieZi injection as a case study [J]. *Molecules*, 2017, doi: 10.3390/molecules22071237.
- [24] 黄 钦, 庄 艳, 乔学斌, 等. 用支持向量机建立中药有效成分聚集体的预测模型 [J]. 物理化学学报, 2007, 23(8): 1141-1144.
- [25] 贺心皓, 罗 旭. 基于粒子群优化算法的支持向量机参数选择 [J]. 计算机系统应用, 2019, 28(8): 241-245.
- [26] 魏从师, 雷福汉, 艾伟霞, 等. 基于 NIRS 技术和 PCA-SVM 算法 6 种树脂及其他类中药的快速鉴别 [J]. 中国实验方剂学杂志, 2017, 23(9): 17-23.
- [27] 王 萌, 赵培陆, 田丽媛, 等. 用于中药人参品类鉴别的算法研究及硬件实现 [J]. 吉林大学学报: 信息科学版, 2017, 35(3): 235-242.
- [28] 徐明玲, 陈 光, 喻长远. 支持向量机法对活血化癥类药物的分类研究 [J]. 北京化工大学学报: 自然科学版, 2009, 36(6): 82-85.
- [29] 邓家刚, 刘 进, 覃洁萍, 等. 20 种活血化癥中药提取物的红外光谱分析及其与药性相关性的初步探讨 [J]. 光谱实验室, 2010, 27(2): 741-744.
- [30] 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.
- [31] 贾成友, 李 微, 张传辉, 等. 基于多指标权重分析和正交设计法优选白黄泄热止痢片复方提取工艺 [J]. 中草药, 2016, 47(6): 917-922.
- [32] 邓 雪, 李家铭, 曾浩健, 等. 层次分析法权重计算方法分析及其应用研究 [J]. 数学的实践与认识, 2012, 42(7): 93-100.
- [33] 许艳秋, 潘美芹. 层次分析法和支持向量机在个人信用评估中的应用 [J]. 中国管理科学, 2016, 24(S1): 106-112.
- [34] 金滋力, 胡建星, 金宏威, 等. 基于支持向量机与层次分析法的中药方剂配伍分析 [J]. 中国中药杂志, 2018, 43(13): 2817-2823.
- [35] 姚卫峰, 胡育筑, 牟玲丽, 等. 基于最小二乘支持向量机的色谱指纹图谱预测银杏叶总抗氧化活性 [J]. 分析化学, 2009, 37(3): 383-388.
- [36] 林 薇. 指纹图谱技术在某些中药和食品质量控制中的应用 [D]. 南昌: 南昌大学, 2012.
- [37] 徐 冰, 王 星, Tom D, 等. 基于遗传算法的多目标最小二乘支持向量机在近红外多组分定量分析中的应用 [J]. 光谱学与光谱分析, 2014, 34(3): 638-642.
- [38] 杨 铭, 陈佳蕾, 余敏英, 等. 基于遗传算法结合最小二乘支持向量机的秦皮提取液快速定量分析方法 [J]. 中国医院药学杂志, 2012, 32(2): 113-116.
- [39] 蒋 兆, 马义中. 基于多核函数最小二乘支持向量机的控制图模式识别方法 [J/OL]. 工业工程与管理, [2020-03-29]. <http://kns.cnki.net/kcms/detail/31.1738.T.20191104.1520.006.html>.
- [40] Chen H Y, Chen J Q, Li J Y, *et al.* Deep learning and random forest approach for finding the optimal traditional Chinese medicine formula for treatment of Alzheimer’s disease [J]. *J Chem Inf Model*, 2019, 59(4): 1605-1623.
- [41] 张 敏. 复杂生产过程质量控制的智能方法研究 [D]. 西安: 西南交通大学, 2013.
- [42] 刘雪松, 孙芬芳, 金 叶, 等. 近红外光谱技术结合粒子群最小二乘支持向量机算法在山茱萸药材质量控制中的应用研究 [J]. 药学学报, 2015, 50(12): 1645-1651.
- [43] 金 叶, 杨 凯, 吴永江, 等. 基于粒子群算法的最小二乘支持向量机在红花提取液近红外定量分析中的应用 [J]. 分析化学, 2012, 40(6): 925-931.
- [44] 邵 永. 中药溶液糖析出建模与参数优化方法研究 [D]. 沈阳: 东北大学, 2017.
- [45] 杨 岩, 肖佳妹, 王 韧, 等. 人工神经网络在中药相关研究领域的应用 [J]. 中草药, 2019, 50(13): 3230-3236.
- [46] 许朝霞, 王忆勤, 颜建军, 等. 基于支持向量机和人工神经网络的心血管疾病中医证候分类识别研究 [J]. 北京中医药大学学报, 2011, 34(8): 539-543.
- [47] Lin Z, Jia S, Luo G, *et al.* Dealing with heterogeneous classification problem in the framework of multi-instance learning [J]. *Talanta*, 2015, doi: 10.1016/j.talanta.2014.09.007.
- [48] Yang S Y, Huang Q, Li L L, *et al.* An integrated scheme for feature selection and parameter setting in the support vector machine modeling and its application to the prediction of pharmacokinetic properties of drugs [J]. *Artif Intell Med*, 2009, 46(2): 155-163.

- [49] Zhang H, Chen Q Y, Xiang M L, *et al.* In silico prediction of mitochondrial toxicity by using GA-CG-SVM approach [J]. *Toxicol In Vitro*, 2009, 23(1): 134-140.
- [50] Wu X, Zuo W, Lin L, *et al.* F-SVM: Combination of feature transformation and SVM learning via convex relaxation [J]. *IEEE Trans Neural Netw Learn Syst*, 2018, 29(11): 5185-5199.
- [51] Wu J, Yang H. Linear regression-based efficient SVM learning for large-scale classification [J]. *IEEE Trans Neural Netw Learn Syst*, 2015, 26(10): 2357-2369.
- [52] 余一鸣, 胡永慧, 韩立云, 等. 中药质量控制的研究进展 [J]. 中草药, 2017, 48(12): 2557-2563.
- [53] 刘昌孝, 陈士林, 肖小河, 等. 中药质量标志物 (Q-marker): 中药产品质量控制的新概念 [J]. 中草药, 2016, 47(9): 1443-1457.
- [54] Li K, Li J, Su J, *et al.* Identification of quality markers of Yuanhu Zhitong tablets based on integrative pharmacology and data mining [J]. *Phytomedicine*, 2018, doi: 10.1016/j.phymed.2018.03.002.
- [55] 贺福元, 周宏灏, 邓凯文, 等. 指纹图谱的一种定性定量研究新方法: 总量统计矩分析法 [J]. 药学学报, 2008, 43(2): 195-201.