

曾 燕,王 晓,成新文,等. 基于 ABC-LSSVM 的芹菜总黄酮提取量预测[J]. 江苏农业科学,2015,43(1):298-300.
doi:10.15889/j.issn.1002-1302.2015.01.099

基于 ABC-LSSVM 的芹菜总黄酮提取量预测

曾 燕¹,王 晓¹,成新文¹,陈欲云²

(1. 四川理工学院计算机学院,四川自贡 643000; 2. 四川理工学院化学与制药工程学院,四川自贡 643000)

摘要:针对传统的芹菜总黄酮含量测定过程复杂、时间长的问題,提出一种基于人工蜂群算法(ABC)优化最小二乘支持向量机(LSSVM)的芹菜总黄酮提取量预测方法。首先对标准人工蜂群算法进行改进,然后利用改进的人工蜂群算法对最小二乘支持向量机的核宽度和正规化参数进行寻优,最后对芹菜总黄酮的提取量进行预测。结果表明,该方法具有预测精度高、性能稳定的特点,有利于实现芹菜总黄酮提取量的网络在线预估和优化控制。

关键词:人工蜂群算法;最小二乘支持向量机;总黄酮;预测;芹菜

中图分类号: S126 **文献标志码:** A **文章编号:** 1002-1302(2015)01-0298-03

芹菜是 1 年生或 2 年生的伞型花科植物,在 15 世纪被作为药食两用栽培植物引入中国。现代医学研究发现,芹菜茎、叶中含有丰富的黄酮类物质,具有清除自由基、抗氧化、抑菌、抗病毒、抗癌、抗肿瘤作用,对心血管疾病和肝病也有一定疗效^[1-2]。从芹菜中提取黄酮类物质,生产具有抗氧化作用的天然保健产品,已成为农产品资源高效利用的研究热点^[3-4]。

目前测定芹菜总黄酮提取量大多采用分光光度法、光谱法等方法,测定过程复杂、时间长、成本较高^[5]。采用软件方法实现芹菜总黄酮提取量预测,是解决此类生物量参数难以实时在线测量和控制的有效方法^[6]。近年来神经网络、支持向量机等预测技术越来越受到重视^[7-8],其中神经网络技术对样本数据量要求高,应用受限;支持向量机预测技术虽然适用于非线性小样本数据,但其泛化能力有限,对异常样本敏感,在实际应用中容易出现大的预测偏差。最小二乘支持向量机技术提高了泛化能力,具有精度高、性能稳定的优点,可用于农业现代化应用中^[9]。影响最小二乘支持向量机预测精度的重要参数指标是正规化参数和核宽度系数,传统的交叉验证确定法可保证较高的预测精度,但预测速度慢,限制了其实际应用。本研究采用改进的人工蜂群算法对最小二乘支持向量机的参数指标进行优化选择^[10-11],避免了标准人工蜂群算法过早收敛问题,实现了芹菜总黄酮提取量的快速预测,以期对蔬菜黄酮类物质提取量的在线测量提供参考。

1 算法基础

1.1 最小二乘支持向量机

最小二乘支持向量机(LSSVM)对支持向量机(SVM)进行了改进,很好地解决了小样本、非线性、高维数的问题,提高了问题的求解速度和泛化能力。最小二乘支持向量机通过构建如下回归函数,将非线性问题转化为高维特征空间的线性

估计问题。

$$f(x) = \omega^T \varphi(x) + b. \quad (1)$$

式中: $\varphi(x)$ 是核空间映射函数; ω 为权值向量; b 为偏置量。

最小二乘支持向量回归算法就是求解如下约束优化问题。

$$\min J(\omega, \xi) = \frac{1}{2} \omega^T \omega + \frac{1}{2} \gamma \sum_{i=1}^n \xi_i^2; \quad (2)$$

$$\text{s. t. } y_i = \omega^T \varphi(x_i) + b + \xi_i, i = 1, 2, \dots, n. \quad (3)$$

式中: ξ_i 为第 i 个样本真实值与预测值的误差; γ 为可以动态调整的正归化参数。

可采用拉格朗日函数求解这个优化问题。

$$L(\omega, \beta, \xi, \alpha) = \frac{1}{2} \omega^T \omega + \frac{1}{2} \gamma \sum_{i=1}^n \xi_i^2 - \sum_{i=1}^n \alpha_i [\omega \varphi(x_i) + b + \xi_i - y_i]. \quad (4)$$

式中: $\alpha_i \in R(i = 1, 2, \dots, N)$ 为 Lagrange 乘子。 a 和 b 的求解与核函数的选择有关,常选择如下形式的径向基核函数(RBF)。

$$K(x_i, x_j) = \exp\left(-\frac{|x_i - x_j|^2}{2\sigma^2}\right). \quad (5)$$

式中: σ 为核宽度。

由此可见,对于最小二乘支持向量机,核宽度 σ 和正规化参数 γ 是 2 个重要参数,它们的选取直接影响着算法的学习能力和泛化性能。

1.2 人工蜂群算法的改进

人工蜂群算法(ABC)模拟蜜蜂群智能搜索行为,具有控制参数少、算法简单等优点,受到国内外学者广泛关注。但是标准 ABC 算法局部搜索能力较差,容易产生早熟和停滞现象^[11]。

改进标准 ABC 算法的随机初始化种群生成方式,采用反向学习^[12]的初始化策略,确保初始种群的多样性,可以提高求解效率和改善解的质量。

首先,随机生成初始解集 $\{x_{ij} | x_{ij} \in [\min_j, \max_j]\}$ 、 (i, j) 分别为解的个数和维度, \min_j, \max_j 分别为第 j 维的下界、上界)。然后,对每个初始解求出对应的反向解,计算方法如下。

收稿日期:2014-09-25

基金项目:四川理工学院科研项目(编号:2013KY04);酿酒生物技术及应用四川省重点实验室开放基金(编号:NJ2011-09)。

作者简介:曾 燕(1979—),女,四川自贡人,硕士,讲师,研究方向为人工智能。E-mail:zy261365@suse.edu.cn。

$$x_{ij} = \min_j + \max_j - x_{ij}。$$
 (6)

最后,对随机种群和反向种群的合集进行排序,选择适应度较优的解作为初始种群。

在标准 ABC 算法中,跟随蜂根据雇佣蜂种群适应度大小,按照轮盘赌策略选择 1 个雇佣蜂跟随,易导致群体多样性下降,算法过早收敛。采用“信息素 - 灵敏度”模型代替轮盘赌选择策略^[11]。

信息素反映了解的质量,信息素的大小与目标函数值呈正比,每完成 1 次搜索过程,信息素进行动态更新。灵敏度确定搜索的区域方向。信息素和灵敏度相结合的方法如下。

$$p(i) = \begin{cases} \frac{f(i) - f_{\min}}{f_{\max} - f_{\min}} & f_{\max} \neq f_{\min} \\ 0 & p(k) \leq s(i) \end{cases}。$$
 (7)

式中: $f(i)$ 为个体的适应度值; f_{\max} , f_{\min} 为最大适应度值和最小适应度值; $p(k)$ 为不等于 i 的第 k 个食物源的信息素; $s(i)$ 为第 i 个跟随蜂的灵敏度; $s(i) \sim U(0,1)$ 。

“信息素 - 灵敏度”模型在一定程度上避免了算法陷入局部最优,保证了算法快速进化的方向。

2 预测建模

2.1 芹菜总黄酮成分提取和测量方法

对于芹菜总黄酮的提取,首先将芹菜原料进行低温烘干,以降低机械粉碎的强度。将其粉碎后,称取芹菜干粉 1 g 于 50 mL 锥形瓶中,加入一定量乙醇浸泡,从而使有效成分充分溶解。然后将浸泡好的溶液投入超声波提取器提取,之后抽滤、定容,作为测定总黄酮含量的待测液,最后测定总黄酮含量。芹菜总黄酮的提取工艺流程如图 1 所示。

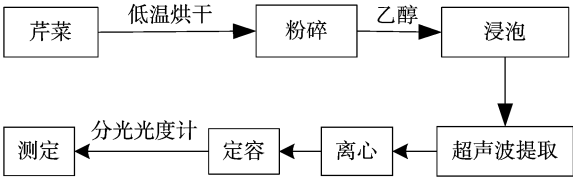


图1 芹菜总黄酮提取流程

采用 $\text{Al}(\text{NO}_3)_3 - \text{NaNO}_2$ 分光光度方法^[4]测定总黄酮含量。取上述样品液 2 mL,用 30% 乙醇溶液定容至 5 mL,加入 15% NaNO_2 溶液 0.3 mL,摇匀静置 6 min,再加入 10% $\text{Al}(\text{NO}_3)_3$ 溶液 0.3 mL,摇匀静置 6 min 后加入 1 mol/L NaOH 溶液 4 mL,再用蒸馏水定容,摇匀静置 15 min 后测定吸光度,根据吸光度计算样品液中总黄酮含量。由此可见,芹菜总黄酮提取测定过程复杂,时间长。

2.2 基于 ABC - LSSVM 的芹菜总黄酮提取量预测模型

采用吸光度方法测定芹菜总黄酮提取量,由于测定过程复杂,时间长,制约了其生产应用。因此,构建芹菜总黄酮提取含量的预测模型,减少试验测量次数,实现在线实时预测尤为必要。基于 ABC - LSSVM 的芹菜总黄酮预测步骤如下: (1)对芹菜总黄酮提取试验中的相关数据进行归一化处理; (2)参数初始化,设置最大迭代次数,按反向学习策略生成初始解,计算适应度函数值,按适应度函数值优劣进行排序,排在前 50% 的为雇佣蜂,排在后 50% 的为跟随蜂; (3)对于第 n 步的雇佣蜂,记录最优值并展开邻域搜索,产生 1 个新位置;

(4)雇佣蜂按贪婪选择方法,当搜索解优于记忆中的最优解时,替换记忆解,反之保持不变; (5)全部雇佣蜂完成邻域搜索后,跳摆尾舞与跟随蜂共享食物源信息;跟随蜂按公式(7)的选择机制选择雇佣蜂; (6)同(3)(4),跟随蜂记下种群最终更新后达到的最优适应度值以及相应的参数; (7)当跟随蜂的邻域搜索次数到达阈值而仍未找到更优位置时,侦察蜂重新初始化食物源位置; (8)如果满足停止准则,则停止计算,输出最优适应度值及相应参数,否则转向(3); (9)根据最优解得到的 LSSVM 参数对测试数据集进行仿真,用建立的芹菜总黄酮预测模型对总黄酮提取量进行预测。

3 结果与分析

根据芹菜总黄酮提取试验,发现总黄酮提取量主要与提取时间、乙醇浓度、料液比等 3 个非线性因素的关系最为密切,三者之间的交互作用较小,且容易测量。因此,在 5 种料液比(1 : 10、1 : 20、1 : 30、1 : 40、1 : 50, g : mL)、5 种乙醇浓度水平(50%、60%、70%、80%、90%)、5 种提取时间(10、20、30、40、50 min)条件下进行芹菜总黄酮提取的单因素试验。

在单因素试验基础上,选用 $L_9(3^3)$ 正交试验表,考察料液比、提取时间、乙醇浓度对芹菜总黄酮提取的影响。因素水平见表 1。

表 1 芹菜总黄酮提取条件正交试验因素与水平

水平	因素		
	提取时间(min)	乙醇浓度(%)	料液比(g : mL)
1	30	70	1 : 20
2	40	80	1 : 30
3	50	90	1 : 40

将试验所得的 24 例数据作为样本数据,在数据归一化之后,选取前 16 例试验样本数据作为训练样本数据,构建基于 ABC - LSSVM 的预测模型,再用后 11 例样本数据作为测试数据,对该模型进行验证。作为训练样本的 16 例原始数据如表 2 所示。

表 2 芹菜总黄酮预测模型训练样本数据

样品编号	提取时间(min)	乙醇浓度(%)	料液比(g : mL)	总黄酮含量(mg/g)
1	40	70	1 : 10	6.472
2	40	70	1 : 20	6.938
3	40	70	1 : 30	7.296
4	40	70	1 : 40	6.902
5	40	70	1 : 50	6.830
6	10	70	1 : 30	6.186
7	20	70	1 : 30	6.400
8	30	70	1 : 30	6.651
9	40	70	1 : 30	6.973
10	50	70	1 : 30	6.938
11	40	50	1 : 30	4.287
12	40	60	1 : 30	5.290
13	40	70	1 : 30	6.544
14	40	80	1 : 30	7.475
15	40	90	1 : 30	7.153
16	40	70	1 : 20	6.207

在预测仿真试验中,计算机配置为联想双核 E5800@3.2 GHz、2 G 内存、Windows XP 操作系统,测试软件采用 Matlab 7.0。

采用改进的人工蜂群算法优化最小二乘支持向量机得到的最优参数值为:核宽度 $\sigma = 262.22$,正规化参数 $\gamma = 1\,969.6$ 。将基于 ABC-LSSVM 预测模型仿真得到的数据与实际测量数据进行比较,结果见表 3。

表 3 基于 ABC-LSSVM 的芹菜总黄酮预测结果						
样品编号	提取时间 (min)	乙醇浓度 (%)	料液比 (g : mL)	实测含量 (mg/g)	预测含量 (mg/g)	相对误差 (%)
1	40	80	1 : 30	7.48	6.95	-7.08
2	40	90	1 : 30	7.15	7.18	0.36
3	30	70	1 : 20	6.28	6.56	4.50
4	40	80	1 : 20	6.49	6.82	5.09
5	50	90	1 : 20	6.60	7.08	7.23
6	30	80	1 : 30	6.96	6.92	-0.55
7	40	90	1 : 30	7.21	7.18	-0.43
8	50	70	1 : 30	6.46	6.74	4.28
9	30	90	1 : 40	7.07	7.21	2.03
10	40	70	1 : 40	6.67	6.77	1.52
11	50	80	1 : 40	7.35	7.03	-4.37

由表 3 可知,采用 ABC-LSSVM 的芹菜总黄酮预测模型得到的仿真数据与实测数据的最大相对误差为 7.23%,最小相对误差为 0.36%,平均相对误差为 1.14%,精度较高。实测数据与仿真数据的对比如图 2 所示。

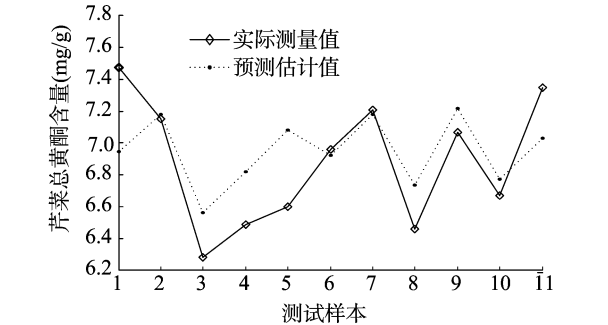


图2 芹菜总黄酮含量的实测数据与仿真数据对比

图 3 更直观地描述了模型估计值与实测值的相对误差情况。基于 ABC-LSSVM 的芹菜总黄酮提取量预测模型的进化曲线如图 4 所示,由图 4 可见,该预测模型收敛速度快、精度高、性能稳定。

4 结论

芹菜总黄酮提取受提取时间、乙醇浓度、料液比等 3 种非线性因素的影响,可利用改进的人工蜂群算法优化最小二乘支持向量机建立芹菜总黄酮预测模型。在标准人工蜂群算法的种群初始化阶段,引入反向学习策略,确保了个体分布的均匀性,并用“信息素-灵敏度”模型作为选择策略,提高了预

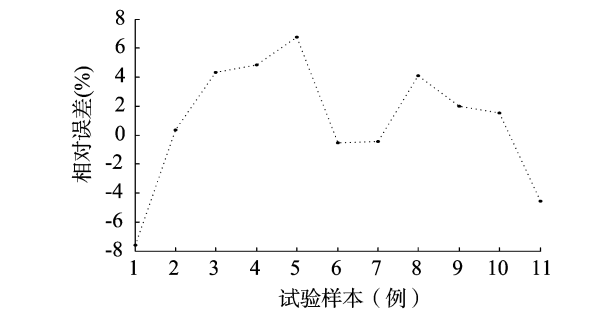


图3 ABC-SSVM 预测模型相对误差

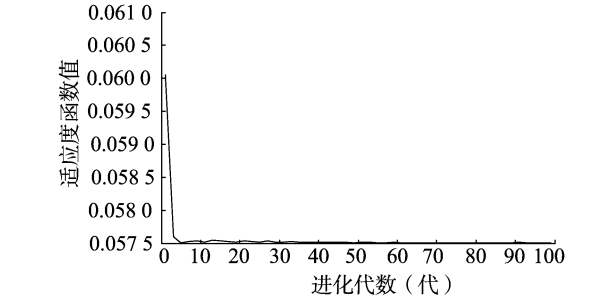


图4 ABC-LSSVM 预测模型进化曲线

测的精度和收敛速度,性能稳定,对于芹菜总黄酮提取的网络在线估计、降低测定成本具有重要意义。

参考文献:

[1]赵进,尤婷婷,林丹英,等. 超声波提取芹菜总黄酮及鉴别[J]. 时珍国医国药,2007,18(6):1454-1455.
[2]文开新,王成章,严学兵,等. 黄酮类化合物生物学活性研究进展[J]. 草业科学,2010,27(6):115-122.
[3]包丽芹,韩德权. 芹菜及芹菜黄酮的药用价值[J]. 黑龙江医药,2007,20(4):317-320.
[4]陈欲云,刘春丽,边名鸿. 芹菜黄酮抗氧化作用的研究[J]. 时珍国医国药,2012,23(7):1718-1719.
[5]何书美,乔兰侠,刘敬兰. 红外光谱法测定芹菜叶提取物中总黄酮的含量[J]. 分析科学学报,2008,24(2):201-204.
[6]俞金寿. 软测量技术及其应用[J]. 自动化仪表,2008,29(1):1-7.
[7]王红君,张梦,赵辉,等. 基于 BP 神经网络的温室黄瓜灌溉预测模型[J]. 江苏农业科学,2013,41(11):407-409.
[8]赵辰阳,徐明. 基于 FIG-SVM 的农产品价格趋势预测[J]. 江苏农业科学,2014,42(5):385-388.
[9]杨玮,孙红,郑立华,等. 基于土壤参数的冬小麦产量预测模型[J]. 农业工程学报,2013,29(23):118-123.
[10]朱林,陆春伟. 基于改进蚁群支持向量机的氢含量软测量建模[J]. 计算机仿真,2014,31(3):384-388.
[11]王慧颖,刘建军,王全洲. 改进的人工蜂群算法在函数优化问题中的应用[J]. 计算机工程与应用,2012,48(19):36-39.
[12]汪慎文,丁立新,谢大同,等. 应用反向学习策略的群搜索优化算法[J]. 计算机科学,2012,39(9):183-187.