

改进支持向量机在棉铃虫人工饲料配方优化中的应用

李俊^{1,2}, 谭显胜^{1,2}, 谭泗桥¹, 袁哲明^{1,2,*}, 熊兴耀²

(1. 湖南农业大学生物安全科学技术学院, 长沙 410128; 2. 湖南省作物种质创新与资源利用重点实验室, 长沙 410128)

摘要: 发展新的实验设计与分析方法, 通过实施尽可能少的实验而获得满意配方对动植物营养、发酵工程等复杂多因素多水平寻优问题极为重要。本研究结合均匀设计(uniform design, UD)与支持向量回归(support vector regression, SVR), 提出了一种新的配方优化实验设计与分析方法 UD-SVR, 将其应用于棉铃虫 *Helicoverpa armigera* (Hübner) 幼虫人工饲料配方优化。结果表明: 在考虑 6 因素时仅通过 2 轮 22 个实验, 表征配方优劣的指标——平均蛹重即由初始的 0.2436 g 高效提升至 0.3044 g, 明显优于二次多项式偏最小二乘回归等经验风险最小参比模型。UD-SVR 预测精度高、指导性强、可解释性好、优化高效, 有望在多因素多水平配方优化中得到广泛应用。

关键词: 棉铃虫; 人工饲料; 均匀设计; 支持向量回归; 配方优化

中图分类号: Q965 文献标识码: A 文章编号: 0454-6296(2010)04-0420-07

Application of improved support vector machine in the optimization of artificial diet for the cotton bollworm, *Helicoverpa armigera* (Lepidoptera: Noctuidae)

LI Jun^{1,2}, TAN Xian-Sheng^{1,2}, TAN Si-Qiao¹, YUAN Zhe-Ming^{1,2,*}, XIONG Xing-Yao² (1. College of Bio-safety Science and Technology, Hunan Agricultural University, Changsha 410128, China; 2. Hunan Provincial Key Laboratory of Crop Germplasm Innovation and Utilization, Hunan Agricultural University, Changsha 410128, China)

Abstract: The development of new experimental design and analysis methods that can provide satisfactory formula through conducting as few experiments as possible is extremely important for complex multi-factor and multi-level optimization problem, such as animal and plant nutrition, and fermentation engineering. Based on uniform design and support vector regression, a novel experimental design and analysis method named as UD-SVR for the prescription optimization was proposed. It was applied to optimize the complicated artificial diet including six factors for the cotton bollworm, *Helicoverpa armigera* (Hübner). The optimization results showed that the mean pupal weight increased from 0.2436 g in the initial benchmark formulation to 0.3044 g in the optimal prescription after carrying out only 22 schemes. UD-SVR is more efficient than the reference models and has the potential to be widely used for experimental design and analysis of the prescription optimization.

Key words: *Helicoverpa armigera*; artificial diet; uniform design; support vector regression; prescription optimization

在动植物营养、发酵工程等诸多领域中, 发展新的实验设计与分析方法, 通过实施尽可能少的实验获得较优配方对复杂多因素多水平实验极为重要(方开泰, 1980; 莫美华和庞雄飞, 2007; 袁哲明等, 2009)。均匀设计(uniform design, UD)抛开正

交设计整齐可比的要求, 只考虑试验点的均匀分散性, 可使每因素每水平仅做一次试验, 因而试验次数可大幅度降低(方开泰, 1980; 唐启义和冯明光, 2002)。均匀设计后续回归分析常采用基于经验风险最小的逐步线性回归(stepwise linear

基金项目: 教育部新世纪优秀人才计划项目(NCET-06-0710); 国家自然科学基金项目(30570351); 高等学校博士点基金(200805370002); 湖南省教育厅青年基金(05B025)

作者简介: 李俊, 男, 1984 年生, 四川眉山人, 硕士研究生, 主要从事昆虫生理生化研究, E-mail: ljun163com@163.com

* 通讯作者 Corresponding author, E-mail: zhmyuan@sina.com

收稿日期 Received: 2009-10-24; 接受日期 Accepted: 2010-04-26

regression, SLR)、多元线性回归(multiple linear regression, MLR)、二次多项式回归(quadratic polynomial regression, QPR)、偏最小二乘回归(partial least-squares regression, PLR)、偏最小二乘二次多项式回归(PLS-QPR)和人工神经网络(artificial neural networks, ANN)等方法,其理论基础是大数定律,只有当训练样本数目趋近于无穷大时,规律才能被统计出来;而配方优化是典型的小样本,基于经验风险最小的传统方法在小样本条件下泛化能力不足,优化效果往往并不理想。同时,配方中各因子、因变量之间往往存在非线性关系,而 SLR, QPR 和 PLR 是线性或拟线性的,非线性解析能力不足(Rosipal and Trejo, 2001)。以响应面方法优化配方时,重构函数的选择是一个关键问题,通常采用的二次多项式对非线性函数的逼近能力有限(Gupta and Manohar, 2004)。ANN 虽能解决非线性问题,但存在模型结构难以确定、可解释性差、易于出现过度训练和训练不足、易于陷入局部极小与维数灾难等问题(Chakraborty *et al.*, 1992)。支持向量机(support vector machine, SVM)基于结构风险最小,较好地解决了小样本、非线性、过拟合、维数灾难和局部极小等问题,泛化推广能力优异(Vapnik, 1995; 邓乃扬和田英杰, 2004),但存在可解释性差、核函数选择依赖经验、对大训练样本计算复杂度高等缺陷。SVM 包括支持向量分类(support vector classification, SVC)和支持向量回归(support vector regression, SVR)。

棉铃虫 *Helicoverpa armigera* (Hübner) 为分布广泛的杂食性多寄主害虫,因其危害严重、抗药性发展快、抗逆能力强等原因成为昆虫生理、抗药性与防治技术等研究的重要实验昆虫。为实现室内种群的继代饲养,获得大量生长健壮、发育整齐的试虫,棉铃虫人工饲料配方优化非常重要(卓乐姝等, 1981; 吴坤君, 1985)。本研究在改进 SVR 缺陷的基础上,结合均匀设计与 SVR 提出了一种新的实验设计与分析方法 UD-SVR, 将其应用于棉铃虫人工饲料配方优化,结果表明新方法优化效果显著、指导性强、可解释性好。

1 材料与方法

1.1 虫源与饲养管理

棉铃虫卵块购自武汉科诺生物农药有限公司,并由本实验室在基准配方人工饲料上继代饲养 3 代

以上。初孵幼虫单虫饲于 24 孔 Corning 细胞培养板中(孔板孔直径 15.6 mm, 生长面积 1.9 cm²), 3 龄前后转移至 6 孔 Corning 细胞培养板单虫饲养(孔板直径 34.8 mm, 生长面积 9.5 cm²), 定期更换饲料、清理培养板,直至幼虫化蛹。培养板置于温度 27 ± 1℃, 光周期 L:D = 14:10, RH 60% ~ 80% 的光照培养箱中。

1.2 基准人工饲料配方

棉铃虫基准人工饲料配方为本室多年饲养的经验配方,室内继代饲养多代,种群无明显衰退。基准人工饲料含以下 3 组分: I 组为黄豆粉 100 g, 麦麸 100 g, 酵母浸膏 100 g, 蔗糖 4 g, 菜籽油 4 滴; II 组为琼脂 12 g; III 组为维生素 C 片(维福佳牌)50 粒, 21 金维他片 0.5 粒, 山梨酸 1.6 g, 冰乙酸 1 mL(维生素 C 片和 21 金维他片用时磨细)。配制时,将 I 组倒入 500 mL 温水中搅拌均匀, II 组琼脂于 500 mL 水中煮沸融化,混合 I 组和 II 组,高压蒸汽灭菌 30 min,待温度降至 50℃ 左右时加入 III 组成分,快速混匀后冷藏备用。

1.3 支持向量回归与 LIBSVM 2.8 软件包

SVM 起初是用于解决线性可分情况下两类样本的分类问题(SVC),其核心思想是找到一个最优分类超平面,使两类样本的分类间隔最大化。SVR 与 SVC 相似,但 SVR 所求超平面是使所有样本点到超平面的距离为最小。对于线性回归问题,给定样本集 (x_i, y_i) , 其中 $i = 1, \dots, n; x \in R^d; y \in R$, 问题变为寻求一个最优超平面,使得在给定精度 ε ($\varepsilon \geq 0$) 条件下可以无误差地拟合 y , 即所有样本点到最优超平面的距离都不大于 ε ; 考虑到允许误差的情况,可引入松弛变量 ξ 和 $\xi^* \geq 0$ 以及惩罚参数 $C > 0$, 其寻优问题转化为相应的带约束条件的二次规划问题;将该优化问题转化成对偶问题后即可解得最优回归函数。仅有少部分样本的 Lagrange 乘子 α 不为零,因此决定该超平面的样本只能是这些支持向量。对于非线性回归问题,可通过核函数变换将样本映射到一个高维特征空间中用线性回归来解决。通常,特征空间具很高甚至无穷维数,致使空间变换后计算量巨增而面临维数灾难等问题。幸运的是 SVM 中待解对偶问题只包含一个变换后特征空间内积运算,而这种运算能在原空间中通过核函数来实现。有关 SVM 和 SVR 的详细内容参见 Vapnik (1995), 张学工 (2000) 和边肇祺等 (2000)。

LIBSVM 2.8 软件包的 SVR 功能需用到 4 个常

用程序: Svm scale 用于对原始数据规格化, Svmtrain 用于训练, Svmpredict 用于预测, Gridregression. py 用于自动搜索核函数最优参数 c , g , p ($c \in [-1, 6]$, $g \in [-8, 0]$, $p \in [-8, -1]$, 步长均为 1)。各程序用法及其参数设置参见文献(Chang and Lin, 2001)。

2 UD-SVR 优化过程与分析

先给出 UD-SVR 配方优化流程框图(图 1), 再结合实验进程分步详细介绍。

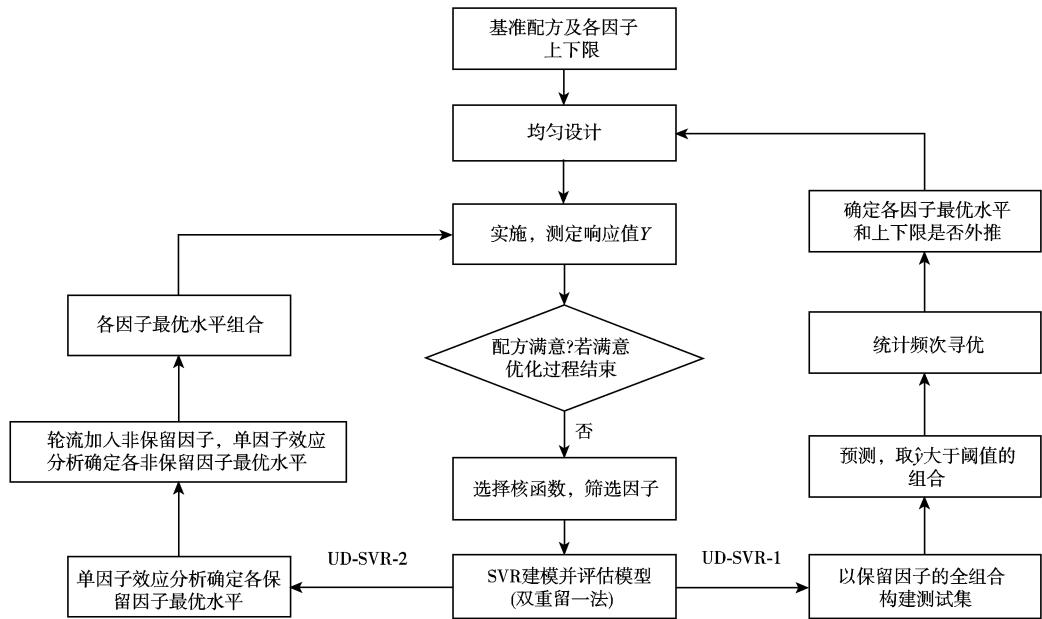


图 1 UD-SVR 配方优化流程
Fig. 1 Flow chart of prescription optimization methods based on UD-SVR

2.1 基准配方及各因素上下限

配方优化考虑 6 因素, 各因素基准、上下限见表 1。其他组分同 1.2 基准人工饲料配方。

表 1 基准方案中各因素的上下限
Table 1 Lower limit and upper limit of components in benchmark formulation

因子 Factor	基准 Benchmark	上限 Upper limit	下限 Lower limit
黄豆粉 Soybean powder, x_1 (g/L)	100	128	40
麦麸 Wheat bran, x_2 (g/L)	100	110	10
酵母浸膏 Yeast extract, x_3 (g/L)	100	150	50
蔗糖 Sucrose, x_4 (g/L)	4	20	2
菜籽油(滴/L) Rapeseed oil, x_5 (drop/L)	4	20	2
维生素 C(粒/L) Vitamin C, x_6 (tablet/L)	50	98	10

2.2 第 1 轮均匀设计

依表 1 按 6 因素 12 水平以 DPS 进行均匀设计(唐启义和冯明光, 2002), 共 12 个处理($N_1 \sim N_{12}$), 每处理供试幼虫 48 头, 单虫饲养(即每处理 48 重复, 下同)。各处理因子水平组合及实测平均蛹重见表 2。可知, 第 1 轮均匀设计蛹重最大的 4 个处理 N_1, N_2, N_7, N_8 与基准人工饲料 CK_1 均无显著差异, 需进一步优化。

2.3 最优核函数选择与非线性变量筛选

SVR 常用核函数有 4 种: 线性核函数($t = 0$), 多项式核函数($t = 1$), 径向基核函数($t = 2$)以及 sigmoid 核函数($t = 3$)。针对 SVR 核函数的选择依赖经验、缺乏先验理论指导的缺陷, 袁哲明等(2007)、张永生等(2007)发展了以 4 种常用核函数依次按留一法进行交叉测试、取均方误差(mean squared error, MSE)最小者为最优核函数的核函数自动寻优算法与程序。不同于逐步线性回归等线性自变量筛选方法, 袁哲明等(2007)、谭显胜等(2008)还基于 SVR 给出了一种非线性的变量筛选

表 2 第 1 轮均匀设计方案及实测蛹重
Table 2 The 1st-round uniform design and mean pupal weight

实验组 Scheme	x_1	x_2	x_3	x_4	x_5	x_6	蛹重 y (平均值 $\pm$ 标准误) (g) Pupal weight (mean $\pm$ SE)
N ₁	120	40	140	5	10	74	0.2563 $\pm$ 0.0052 aA
N ₂	72	50	150	11	2	34	0.2445 $\pm$ 0.0063 abAB
N ₃	104	90	120	8	6	10	0.2291 $\pm$ 0.0061 bcBCDE
N ₄	48	70	80	2	9	26	0.2039 $\pm$ 0.0050 dF
N ₅	88	120	100	3.5	4	82	0.2048 $\pm$ 0.0063 dF
N ₆	128	110	50	12.5	8	42	0.2330 $\pm$ 0.0054 bcABCDE
N ₇	112	60	70	18.5	3	66	0.2464 $\pm$ 0.0043 abAB
N ₈	96	20	90	15.5	12	18	0.2439 $\pm$ 0.0061 abABC
N ₉	40	30	110	14	7	90	0.2100 $\pm$ 0.0072 dEF
N ₁₀	80	80	60	9.5	13	98	0.2354 $\pm$ 0.0055 bcABCD
N ₁₁	64	10	40	6.5	5	50	0.2198 $\pm$ 0.0052 cdCDEF
N ₁₂	56	100	130	17	11	58	0.2294 $\pm$ 0.0055 bcBCDE
CK ₁	100	100	100	4	4	50	0.2436 $\pm$ 0.0070 abABC

N₁ ~ N₁₂ 为均匀设计处理, CK₁ 为基准人工饲料对照。蛹重数据后有不同字母表示差异显著, 小写和大写字母分别代表显著水平为 $P < 0.05$ 和 $P < 0.01$ (Duncan 氏新复极差法检验); 下同。N₁ - N₁₂ are schemes based on uniform design; NK is the benchmark formulation for the control. Different small and capital letters mean significant difference by Duncan's new multiple range test at the 5% level and 1% level, respectively. The same for the following tables.

方法：以多轮末尾淘汰法从包含全部输入变量的 SVR 模型中以留一法依 MSE 最小原则非线性逐次剔除对提高预测精度不利的变量, 筛选后的变量即为保留因子。在此基础上为使 SVR 具备更好的解释性, 进一步发展了多轮末尾强制淘汰法对保留因子进行强制筛选, 该方法可给出各因子相对重要性顺序, 使模型具备一定解释能力(袁哲明等, 2007; 张永生等, 2007; 谭泗桥等, 2007; 谭显胜等, 2008)。最优核函数选择和非线性变量筛选应同步进行。

非线性因子筛选结果表明, 最优核函数为径向基核($t=2$), 保留因子为 x_4, x_2, x_1 , 对蛹重 y 预

测精度影响的相对重要性为 $x_1 > x_2 > x_4$; 在表 1 给定的上下限范围内, 因子 x_6, x_3, x_5 对蛹重影响较小(表 3)。

2.4 模型评估

由此得到一个含 3 个输入因子的 SVR 模型, 现以双重留一法评估该模型的泛化推广能力(袁哲明等, 2007)。从 $t=2$ 及其 3 个保留因子出发, 从 13 个样本中每次拿出 1 个作为独立测试样本, 其余 12 个样本再次以留一法 Gridregression.py 搜索寻找最优 SVR 参数, 以最优参数 c, g, p 对该 12 个样本 Svmtrain 后预测 (Svmpredict) 独立测试样本。双重留一法预测结果 MSE 为 0.0018, 表明所建模

表 3 非线性因子筛选与保留因子强制筛选及其 MSE 值(径向基核)
Table 3 Screening variables and compulsorily screening retained variables and their MSE ($\times 10^4$)

筛选历程 Stage	轮次 Round	筛选前的 MSE 值 MSE before screening	x_1	x_2	x_3	x_4	x_5	x_6	筛选掉的因子 Variables screened
因子筛选 Screening	1	2.20	2.85	1.90	1.81	1.89	2.14	1.77	x_6
variables	2	1.77	2.85	1.73	1.69	1.72	1.72	-	x_3
强制筛选 Compulsorily screening	3	1.69	2.85	1.93	-	1.76	1.42	-	x_5
	4	1.42	2.85	1.88	-	1.54	-	-	x_4
	5	1.54	2.85	1.76	-	-	-	-	x_2

型有较好的泛化推广能力;实测值最大的4组有3组与预测值接近,仅 N_2 稍差(数据未显示)。此处配方优化更多关注的是预测高值的报准程度,通过预测高值找到实际最好的处理。可见, $t=2$ 且保留因子为 x_4, x_2, x_1 的SVR模型可以信赖。

2.5 第2轮实验设计

从表2第1轮均匀设计结果出发,可基于频次统计寻优或基于单因子效应分析两种方案进行第2轮实验设计与优化;为显示UD-SVR的有效性,同时设计若干参比模型进行比较。

2.5.1 基于频次统计寻优的配方优化(UD-SVR-1):以 $t=2$ 为最优核函数,以 x_4, x_2, x_1 为保留因子,对表2的13个样本训练建模,对 x_1 (12水平)、 x_2 (12水平)、 x_4 (12水平)共1728个全组合待测样本进行预测,取预测平均蛹重 $\hat{y}$ (0.25的213个处理统计各因子频次如表4)。

表4 预测平均蛹重大于0.25的213个样本的频次统计寻优结果

Table 4 Factor's frequency statistics for 213 samples with the predicted mean pupal weight >0.25 g

x_1	频次 Frequency	x_2	频次 Frequency	x_4	频次 Frequency
40	0	10	42	2	8
48	0	20	40	3.5	11
56	0	30	39	5	12
64	0	40	35	6.5	15
72	0	50	27	8	16
80	0	60	20	9.5	19
88	0	70	10	11	21
96	0	80	0	12.5	21
104	19	90	0	14	24
112	53	100	0	15.5	24
120	67	110	0	17	22
128	74	120	0	18.5	20

预测最优组合为 $x_1=128, x_2=10, x_4=15.5$,对应最大预测值 $\hat{y}=0.2594$ 。结合预测最优组合和表4, x_1 (黄豆粉)应超过128,上限外推至180,下限调整为120;基准配方中 x_2 (麦麸)含量明显偏高,下限外推至0,上限调整为21.6;基准配方中 x_4 (蔗糖)含量明显偏低,同时饲养经验表明较高的蔗糖含量可改善饲料口感、刺激昆虫取食,故下限调整为16,上限外推至22。以3因子4水平进行第2轮均匀设计,共8处理($N_{13} \sim N_{20}$)。表3已指出,在表1给定的上下限范围内 x_6, x_3, x_5 对蛹

重影响较小,故考虑成本可取下限值,稳妥起见可取均值或基准值;此处取 $x_3=100, x_5=8, x_6=48$ 。第2轮均匀设计及各处理平均蛹重预测值、实测值见表5,每处理供试幼虫48头。

由表2和表5可知,第1轮均匀设计的12个处理实测最高蛹重为0.2563 g;经SVR指导的第2轮均匀设计,8个处理($N_{13} \sim N_{20}$)中有4个蛹重超过0.2563 g,特别是 N_{15} 处理平均蛹重达0.2889 g,显著高于其他7个处理及基准对照 CK_2 。表明基于频次统计寻优的UD-SVR整体优化性能较好。

2.5.2 基于单因子效应分析的配方优化(UD-SVR-2):以 x_4, x_2, x_1 为保留因子,对表2的13个样本训练建立SVR预测模型。为分析 x_j 的单因子效应,可将除 x_j 外的各因子固定为其均值,令 x_j 在合适区间内按一定步长取值,产生待测样本集,代入SVR预测模型得预测值 $\hat{y}_i$,可知因变量 $\hat{y}_i$ 随 x_j 的变化规律,取因变量为极大时 x_j 的值为该因子最优水平,依次获得3个保留因子的最优水平。

在基于频次统计寻优的配方优化中,非保留因子的取值具一定随意性;本方案基于单因子效应分析,对非保留因子同样给出最优水平:基于表2的13个样本、3个保留因子,轮流加入一个非保留因子 x_j 训练建模;3个保留因子取其均值,令 x_j 在合适区间内按一定步长取值,产生待测样本集,代入模型得预测值 $\hat{y}_i$,依次获得3个非保留因子的最优水平。

基于单因子效应分析模型(UD-SVR-2)推测的6因子最优水平组合 $x_1 \sim x_6$ 分别为172, 14.4, 68, 21.2, 2, 40。实测平均蛹重达0.3044 g,在所有处理中最高(表5),显示基于单因子效应分析的UD-SVR优化效果十分突出。

2.5.3 参比模型:参比经验风险最小模型中,由表2获得SLR模型为 $\hat{y}=0.18074+0.00046x_1-0.000162x_2+0.00016x_3+0.00071x_4$ (调整相关系数 $R_a=0.7420$);PLS-QPR考虑交互项和平方项模型、PLS-QPR仅考虑交互项模型、PLS-QPR仅考虑平方项模型均相当冗长,且可依表2数据由DPS软件快捷获得(唐启义和冯明光,2002),故未列出。4个经验风险最小模型推测的最优组合及其验证结果见表5。可见,参比模型实测最高的为PLS-QPR-2,但也仅为0.2647 g,与基准人工饲料(CK_2)差异不显著;参比模型推测的最优组合预测值均高于实测值,尤以PLS-QPR-2偏离最大;表明参比模型优化效果较差。与此相反,基于频次统

表 5 第 2 轮实验设计及实测平均蛹重
Table 5 The 2nd-round experimental design and mean pupal weight

实验组 Scheme	x_1	x_2	x_3	x_4	x_5	x_6	蛹重 y (平均值 $\pm$ 标准误)(g)	
							Pupal weight (mean $\pm$ SE)	
							预测值 Predicted value	实测值 Observed value
N ₁₃	180	14.4	100	18	8	48	0.2844	0.2492 $\pm$ 0.0057 bC
N ₁₄	120	7.2	100	20	8	48	0.2644	0.2650 $\pm$ 0.0059 bBC
N ₁₅	160	7.2	100	16	8	48	0.2782	0.2889 $\pm$ 0.0073 aAB
N ₁₆	140	14.4	100	22	8	48	0.2716	0.2644 $\pm$ 0.0096 bBC
N ₁₇	120	21.6	100	16	8	48	0.2605	0.2448 $\pm$ 0.0081 bC
N ₁₈	160	21.6	100	20	8	48	0.2771	0.2641 $\pm$ 0.0068 bBC
N ₁₉	180	0	100	22	8	48	0.2876	0.2499 $\pm$ 0.0077 bC
N ₂₀	140	0	100	18	8	48	0.2727	0.2481 $\pm$ 0.006 bC
UD-SVR-2	172	14.4	68	21.2	2	40	0.2796	0.3044 $\pm$ 0.0077 aA
SLR	128	10	150	20	8	48	0.2756	0.2574 $\pm$ 0.006 bC
PLS-QPR-1	124	39.6	148	6.4	13	68	0.2737	0.2613 $\pm$ 0.006 bC
PLS-QPR-2	128	10	150	2	2	98	0.3292	0.2647 $\pm$ 0.0042 bBC
PLS-QPR-3	120	49.2	99.6	12	13	66	0.2793	0.2515 $\pm$ 0.007 bC
CK ₂	100	100	100	4	4	50	0.2265	0.2428 $\pm$ 0.0053 bC

N₁₃ ~ N₂₀: UD-SVR-1 给出的均匀设计处理; PLS-QPR-1: PLS-QPR 考虑交互项和平方项模型; PLS-QPR-2: PLS-QPR 仅考虑交互项模型; PLS-QPR-3: PLS-QPR 仅考虑平方项模型; CK₂: 基准人工饲料对照。N₁₃ - N₂₀ were schemes based on UD-SVR-1; PLS-QPR-1, PLS-QPR-2 and PLS-QPR-3 are the optimal schemes predicted by the model of PLS-QPR with reciprocation and square, PLS-QPR with reciprocation only and PLS-QPR with square only, respectively; CK₂ is the benchmark formulation for the control.

计寻优的 UD-SVR-1 所得最优处理 N₁₅ 和基于单因子效应分析的 UD-SVR-2 实测蛹重分别为 0.2889 和 0.3044 g, 均显著或极显著高于参比模型与基准对照, 充分表明在小样本前提下 UD-SVR 优化高效。

3 讨论

基于第 1 轮均匀设计数据(表 2), 非线性因子筛选指出 x_1 , x_2 , x_4 为保留因子。基于频次统计寻优的 UD-SVR-1 认为 x_1 (黄豆粉) 上限应超过 128; 基准配方中 x_2 (麦麸) 为 100, 明显偏高, 应调整为 10 或 10 以下; 基准配方中 x_4 (蔗糖) 为 4, 明显偏低, 应调整为 15.5 左右。基于单因子效应分析的 UD-SVR-2 认为 x_1 , x_2 , x_4 应分别为 172, 14.4, 21.2。UD-SVR-1 和 UD-SVR-2 对第 2 轮实验设计的指导意见较为一致。表 5 实际验证结果, UD-SVR-1 指导的 8 个均匀设计处理实测最优的为 N₁₅, 其 x_1 , x_2 , x_4 分别为 160, 7.2, 16。UD-SVR-2 与 N₁₅ 处理蛹重均极显著高于基准配方。这说明 UD-SVR 具较好的解释性与指导性。

梁革梅等(1999)曾报道一种棉铃虫人工饲料配方, 其平均蛹重达 0.33037 g, 高于本文报道的 0.3044 g。然而, 由于饲料配方组成的差异、棉铃虫种群不一、室内继代饲养代数不一等, 两者并不具可比性。再者, 本研究的目的更多的是提供一种多因素多水平配方优化与实验设计新方法。

对一个 6 因子的棉铃虫人工饲料复杂配方优化问题, UD-SVR 仅通过 2 轮 22 个实验, 即获得一个满意配方: 黄豆粉 172 g/L, 麦麸 14.4 g/L, 酵母浸膏 68 g/L, 蔗糖 21.2 g/L, 菜籽油 2 滴/L, 维生素 C 片 40 粒/L, 21 金维他片、琼脂、山梨酸、冰乙酸用量同基准配方, 实测平均蛹重 0.3044 g。UD-SVR 预测精度高、指导性强、可解释性好、优化高效, 为解决多因素多水平的配方优化试验设计与分析提供了新的有效途径, 但其普适性仍需更多研究实例的支持。

参考文献 (References)

Bian ZQ, Zhang XG, 2000. Pattern Recognition. 2nd ed. Tsinghua University Press, Beijing. 284 - 303. [边肇祺, 张学工, 2000. 模式识别. 第 2 版. 北京: 清华大学出版社. 284 - 303]

- Chakraborty K, Mehrotra K, Mohan CK, Ranka S, 1992. Forecasting the behavior of multivariate time series using neural networks. *Neural Networks*, 5: 961–970.
- Chang CC, Lin CJ, 2001. LIBSVM: a library for support vector machines. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- Deng NY, Tian YJ, 2004. Support Vector Machine – A New Method in Data Mining. Science Press, Beijing. 77–162, 224–272. [邓乃扬, 田英杰, 2004. 数据挖掘中的新方法——支持向量机. 北京: 科学出版社. 77–162, 224–272]
- Fang KT, 1980. The application of uniform design – number-theoretic method in experimental design. *Acta Math. Appl. Sin.*, 3(4): 363–372. [方开泰, 1980. 均匀设计——数论方法在试验设计的应用. 应用数学学报, 3(4): 363–372]
- Gupta S, Manohar CS, 2004. An improved response surface method for the determination of failure probability and importance measures. *Structural Safety*, 26: 123–139.
- Liang GM, Tan WJ, Guo YY, 1999. The improvement of artificial breeding techniques in cotton bollworm. *Plant Protection*, 25(2): 15–17. [梁革梅, 谭维嘉, 郭予元, 1999. 人工饲养棉铃虫技术的改进. 植物保护, 25(2): 15–17]
- Mo MH, Pang H, Pang XF, 2007. Optimization of the semi-synthetic diet of diamondback moth. *Acta Sci. Nat. Univ. Sunyatseni*, 46(6): 63–68. [莫美华, 庞虹, 庞雄飞, 2007. 小菜蛾半合成人工饲料配方的优化. 中山大学学报(自然科学版), 46(6): 63–68]
- Rosipal R, Trejo LJ, 2001. Kernel partial least squares regression in reproducing Kernel Hilbert space. *J. Mach. Learn. Res.*, (2): 97–123.
- Tan SQ, Yuan ZM, Bai LY, Xiong JY, 2007. The QSAR modeling of herbicide using a combinatorial model based on support vector regression and K-nearest neighbor. *Chin. J. Pestic. Sci.*, 9(4): 324–329. [谭泗桥, 袁哲明, 柏连阳, 熊洁仪, 2007. 基于支持向量机回归与 K-最近邻法的组合预测用于除草剂 QSAR 建模. 农药学报, 9(4): 324–329]
- Tan XS, Yuan ZM, Zhou TJ, Wang CJ, Xiong JY, 2008. Multi-KNN-SVR combinatorial forecast and its application to QSAR of fluorine-containing compounds. *Chemical Journal of Chinese Universities*, 29(1): 95–99. [谭显胜, 袁哲明, 周铁军, 王春娟, 熊洁仪, 2008. Multi-KNN-SVR 组合预测在含氟化合物 QSAR 研究中的应用. 高等学校化学学报, 29(1): 95–99]
- Tang QY, Feng MG, 2002. DPS Data Processing System – Experimental Design, Statistical Analysis and Data Mining. Science Press, Beijing. 147–170. [唐启义, 冯明光, 2002. 实用统计分析及其 DPS 数据处理系统. 北京: 科学出版社. 147–170]
- Vapnik VN, 1995. The Nature of Statistical Learning Theory. Springer Verlag Press, New York. 87–189.
- Wu KJ, 1985. A lucerne-wheat germ diet for rearing the cotton bollworm, *Heliothis armigera* (Hübner). *Acta Entomol. Sin.*, 28(1): 22–29. [吴坤君, 1985. 棉铃虫的紫云英-麦胚人工饲料. 昆虫学报, 28(1): 22–29]
- Yuan ZM, Xiong JY, Zhang YS, 2007. A novel combinatorial forecast method based on support vector machine regression and k-near neighbor group and its application in QSAR. *J. Mol. Sci.*, 23(3): 163–169. [袁哲明, 熊洁仪, 张永生, 2007. 基于 SVR 和 k-近邻群的组合预测在 QSAR 中的应用. 分子科学学报, 23(3): 163–169]
- Yuan ZM, Zuo B, Tan SQ, Tan XS, Xiong XY, 2009. Experimental design and analysis of the optimal fermentation medium based on uniform design and support vector regression. *Chin. J. Process Eng.*, 9(1): 148–152. [袁哲明, 左斌, 谭泗桥, 谭显胜, 熊兴耀, 2009. 基于均匀设计与支持向量回归的发酵配方优化. 过程工程学报, 9(1): 148–152]
- Zhang XG, 2000. Introduction to statistical learning theory and support vector machines. *Acta Autom. Sin.*, 26(1): 32–42. [张学工, 2000. 关于统计学习理论与支持向量机. 自动化学报, 26(1): 32–42]
- Zhang YS, Yuan ZM, Xiong JY, Zhou TJ, 2007. Multidimensional time series analysis based on support vector regression and controlled autoregressive and its application in ecology. *Acta Ecol. Sin.*, 27(6): 2 419–2 424. [张永生, 袁哲明, 熊洁仪, 周铁军, 2007. 基于 SVR 和 CAR 的多维时间序列分析及其在生态学中的应用. 生态学报, 27(6): 2 419–2 424]
- Zhuo YS, Huang YL, Yang JR, 1981. Studies on the artificial diets of the cotton bollworm *Heliothis armigera* (Hubner). *Acta Entomol. Sin.*, 24(1): 108–110. [卓乐似, 黄月兰, 杨家荣, 1981. 棉铃虫人工饲料的研究. 昆虫学报, 24(1): 108–110]

(责任编辑: 袁德成)