

# 基于 SVM 的 RSM 模型拟合方法研究

崔庆安<sup>1,2</sup>, 何 桢<sup>2</sup>, 崔 楠<sup>3</sup>

(1. 郑州大学管理工程研究所, 郑州 450001; 2. 天津大学管理学院, 天津 300072; 3. 武汉大学经济与管理学院, 武汉 430072)

**摘要:** 对于多极值、存在高阶交互作用和约束的复杂过程, 参数 RSM 整体代表性差, 难以达到全局最优; 而非参数 RSM 在样本量有限时泛化性差, 模型难以优化. 将 RSM 模型拟合归结为一类有限制条件、可主动获取样本点的小样本学习问题; 提出一种基于 SVM 的复杂过程 RSM 模型拟合方法, 并提出了适用于 RSM 的实用性 SVM 核函数及参数选择方法. 算例研究表明, 所提的核函数及参数选择方法得到的泛化误差与其最小值的平均偏离率在 20% 以内; 基于 SVM 的 RSM 拟合模型对因子约束、误差分布无严格限制, 泛化性能、曲面重现能力均优于现有 RSM, 其平均泛化误差与样本量分别比非参数 RSM 降低约 20% 和 30%, 说明了所提方法的有效性 with 优越性.

**关键词:** 支持向量机; 响应曲面法; 复杂过程; 模型拟合

**中图分类号:** F253.3   **文献标识码:** A   **文章编号:** 1007-9807(2008)01-0031-11

## 0 引 言

提高产品质量是企业赢得竞争优势的根本途径. 工业中大多数过程的质量响应与影响因子的作用机理复杂, 无法建立工程或物理模型. 如何较真实地描述这种作用关系, 从而选择合适的因子的值, 使响应满足目标, 是质量改进和优化关键问题. 目前常用方法是 Box 和 Wilson<sup>[1]</sup> 提出的响应曲面法 (Response Surface Methodology, RSM). RSM 首先采用实验设计方法获取数据, 然后利用回归估计, 拟合二阶模型近似表示因子与响应之间的关系; 再采用最速上升法优化响应得到合适的因子组合. RSM 包括实验设计、模型拟合、过程优化三方面的内容<sup>[2]</sup>, 经过多年的发展, 其理论体系涵盖了参数方法: 如双响应曲面法<sup>[3]</sup>、多响应优化<sup>[4, 5]</sup>、广义线性模型 (general linear model, GLM)<sup>[6, 7]</sup> 等; 非参数方法: 如核函数回归 (kernel function regression, KFR)<sup>[8]</sup>、人工神经网络 (artificial neural networks, ANN)<sup>[9, 10]</sup> 等; 应用上则涉及了离散制造业<sup>[11, 12]</sup>、食品与生物工程<sup>[13-15]</sup> 等诸

多领域的质量优化.

随着生产过程的日益复杂和质量要求不断提高, 现有 RSM 存在如下问题有待解决: 首先, 参数 RSM 只能在很小区域内近似响应过程, 无法对存在高阶交互作用和约束条件的区域进行拟合, 整体代表性较差; 其次, 非参数 RSM 是一种大样本渐进性理论, 当样本量有限时泛化性能较差. KFR 的窗宽参数、ANN 网络结构的确定多依赖经验, 极容易出现“过学习”, 使拟合模型过于复杂, 拟合性好而泛化性差. 而且非参数 RSM 拟合模型的解析形式复杂 (如 KFR) 或根本不具有解析形式 (如 ANN), 难以进行响应寻优.

针对上述不足, 本文将机器学习领域的支持向量机技术引入到了 RSM 的模型拟合之中, 提出了一种能较好的反应曲面形状与极值分布, 泛化能力强, 克服现有 RSM 代表性差、泛化性差、样本量大等缺点的 RSM 拟合方法. 由于现有 RSM 可以较好对简单的响应过程的进行拟合与优化, 因

收稿日期: 2005-07-04; 修订日期: 2007-11-18;  
基金项目: 国家自然科学基金资助项目 (70572044).  
作者简介: 崔庆安 (1974—) 男, 山西襄垣人, 博士, 讲师. Email: cuiqia@zsu.edu.cn

此本文研究对象定位为存在多个极值或鞍点、因子间存在高阶交互作用或约束条件等难以有效应用现有 RSM 的复杂的响应过程 (简称复杂过程)。

### 1 RSM 的机器学习描述

#### 1.1 RSM 本质分析

首先, RSM 是对未知系统的认识与学习过程, 学习性能好的 RSM 应该尽可能真实地反映过程的响应曲面形状, 具有良好的拟合与泛化性能; 其次, 拟合模型是影响 RSM 学习能力的重要因素, 现有 RSM 对复杂过程的代表性差是由于模型不合适而引起的. 因此, 要想提高学习能力, 需寻找合适的拟合模型. 由于最优点是由模型优化, 而非实验得到, 因此模型泛化性能的比拟合性能更为重要; 第三, 拟合模型决定了实验设计方法和过程优化方法. 例如一阶、二阶多项式导致了经典 RSM 相对简单的优化方法, 但是为估计模型系数需严格限制实验点位置; 非参数 RSM 虽不限制模型形式, 对设计要求不严格 (如空间网格设计), 但是过程优化较为困难.

#### 1.2 RSM 机器学习描述

RSM 的模型拟合阶段的主要方法是回归估计. 而回归估计作为一类机器学习问题可以描述如下: 设系统  $S$  的输出  $y \in R$  与输入  $x \in R^n$  之间存在一定的未知依赖关系  $y = f(x)$  或联合概率  $F(x, y)$ , 由方法  $G$  产生了的  $l$  个独立同分布的观测样本

$$T = \{ (x_1, y_1), (x_2, y_2), \dots, (x_l, y_l) \},$$
$$x_i \in X, y_i \in Y, i = 1, 2, \dots, l \quad (1)$$

回归估计就是利用  $T$  中的数据, 在函数集中  $\{f(x)\}$  寻找一个最优的函数  $\{f_0(x)\}$ , 使预测的期望风险 (或泛化误差)

$$R(\hat{f}) = \int L(y, f(x)) dF(x, y) \quad (2)$$

最小. 其中  $\{f(x)\}$  由学习机  $LM$  来实现;  $L(y, f(x))$  称为损失函数, 用于度量预测值和实际值之间的差异.

根据机器学习理论, 在 RSM 中, 系统  $S$  是所研究的响应过程;  $G$  代表实验设计方法, 如因子设计等等;  $LM$  是拟合模型, 如二阶多项式, ANN 等等; 损失函数是最小二乘估计 (如经典 RSM)、最小化经验风险 (如 ANN) 等等. 回归问题的样本来源于历史数据且样本量比较大, 但对于 RSM, 一

方面可以通过精心地设计实验来获取有利于模型拟合的样本, 另一方面由于客观条件的制约, 实验要遵循一定准则, 且样本量一般较小. 因此说, RSM 模型拟合是一类有限制条件的、可以主动获取样本的小样本机器学习问题. 这样就可以在机器学习的框架之下, 将有关机器学习的最新理论引入 RSM, 结合质量改进与优化的特点, 发展一种全新的 RSM 模型拟合方法.

### 2 统计学习理论及 SVM

由 Vpanik<sup>[16]</sup> 提出的统计学习理论及支持向量机 (support vector machines, SVM) 技术, 被认为是目前关于小样本统计学习和预测的最佳机器学习理论. 它侧重于研究有限样本情况下的学习性能, 可以根据结构风险最小化原则自动确定模型结构, 并且得到全局最优点. 国内外学者在 SVM 的模型性质、训练算法、参数选择等方面做了大量的研究, 其理论体系已基本成熟<sup>[17~19]</sup>; 并且广泛应用于模式识别<sup>[20, 21]</sup>、建模与预测<sup>[22, 23]</sup>、异常检测<sup>[24]</sup>、信号处理<sup>[25]</sup> 等领域, 但是在 RSM 中的应用还很少见报道.

SVM 中,  $L(y, f(x))$  一般定义为  $\rho$ -不敏感区函数: 给定  $\rho > 0$

$$L(y, f(x)) = \begin{cases} 0 & |y - f(x)| \leq \rho \\ |y - f(x)| - \rho & |y - f(x)| > \rho \end{cases} \quad (3)$$

即当  $x$  的观测值  $y$  与估计值  $f(x)$  之差不超过  $\rho$  时, 则认为无损失.

统计学习理论认为, 拟合模型的泛化误差  $R$  是经验风险 (即拟合误差)  $R_{emp}$  与置信范围  $\sigma$  之和, 为使  $R$  最小, 应同时最小化  $R_{emp}$  与  $\sigma$ . 这就是结构风险最小化 (structural risk minimization, SRM) 原则. 传统的最小二乘法、KFR、ANN 等只注重最小化  $R_{emp}$ , 因而得不到最小的  $R$ .

作为 SRM 的实现工具, SVM 以拟合误差作为约束条件, 求解如下的最优化问题

$$\min_{w, b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*)$$
$$s.t. \quad (w \cdot T(x_i) + b) - y_i \leq \rho - \xi_i$$
$$y_i - (w \cdot T(x_i) + b) \leq \rho + \xi_i^*$$
$$\xi_i, \xi_i^* \geq 0, i = 1, 2, \dots, l \quad (4)$$



式中,  $\frac{1}{2} w^2$  为置信区间,  $T(x)$  为非线性变换函数.  $\epsilon_i, \epsilon_i^*$  为松弛变量.  $C$  为惩罚参数, 用于对超出  $\epsilon$ -不敏感区的样本进行惩罚. 模型的泛化误差由  $\min_{w, b} \left\{ \frac{1}{2} w^2 + C \sum_{i=1}^l (\epsilon_i + \epsilon_i^*) \right\}$  来控制.

式 (4) 的对偶问题为

$$\begin{aligned} \min_{a_i^*} & \frac{1}{2} \sum_{i,j=1}^l (a_i^* - a_i)(a_j^* - a_j) k(x_i, x_j) + \\ & \sum_{i=1}^l (a_i^* + a_i) - \sum_{i=1}^l y_i (a_i^* - a_i) \\ \text{s t} & (a_i - a_i^*) = 0 \\ & 0 \leq a_i, a_i^* \leq C, \quad i = 1, 2, \dots, l \end{aligned} \quad (5)$$

式中  $k(x_i, x_j) = T(x_i) \cdot T(x_j)$  称为核函数, 用来简化非线性映射后的高维计算. 常用的核函数有高斯径向基核、多项式核、傅立叶核, 等等.

式 (5) 的最优解为

$$(\bar{a}, \bar{a}^*) = (\bar{a}_1, \bar{a}_1^*, \bar{a}_2, \bar{a}_2^*, \dots, \bar{a}_l, \bar{a}_l^*) \quad (6)$$

计算  $\bar{w} = \sum_{i=1}^l (\bar{a}_i - \bar{a}_i^*) k(x_i, x)$ , 选择位于区间  $(0, C)$  中的  $\bar{a}_i$  或  $\bar{a}_i^*$ , 根据相应下标计算

$$\bar{b} = y_j - \sum_{i=1}^l (\bar{a}_i - \bar{a}_i^*) k(x_i, x) \quad (7)$$

拟合函数  $f(x)$  变为

$$\begin{aligned} f(x) &= \bar{w} \cdot x + \bar{b} \\ &= \sum_{i=1}^l (\bar{a}_i - \bar{a}_i^*) k(x_i, x) + \bar{b} \end{aligned} \quad (8)$$

计算出的  $\bar{a}_i = 0$  或  $\bar{a}_i^* = 0$  所对应的  $x_i$  称为支持向量. 支持向量具有稀疏性, 也就是说, 并不是所有的样本点都是支持向量, 将非支持向量样本删除后, 重新学习所得回归函数不变. 此外, 根据 Vapnik 的管道模型<sup>[16]</sup>, 支持向量的数目受  $\epsilon$ -不敏感参数控制,  $\epsilon$  越小, 支持向量的数目越少;  $\epsilon$  越大, 支持向量的数目则越多.

### 3 基于 SVM 的 RSM 拟合方法

#### 3.1 适用于 RSM 的 SVM 核函数及参数的选取方法

利用 SVM 进行 RSM 拟合, 首先要解决 SVM 核函数和参数的选择问题. 目前 SVM 参数的选择主要有交叉验证法<sup>[26]</sup>、理论分析法<sup>[27]</sup>、启发式算法<sup>[28]</sup> 等等. 这些方法或者计算复杂, 需要较大的

样本量; 或者需要额外样本作为测试样本, 因此不适合于复杂过程的 RSM 模型拟合. 下面结合复杂过程的特点, 提出 SVM 核函数及参数选择方法:

##### 1) 核函数的选择

研究表明, 通过选取适当的核参数  $\gamma$ , Gauss 核可以拟合任意的函数<sup>[27]</sup>, 并且在应用中表现优良<sup>[28-31]</sup>, 因此这里将其选作为 SVM 的核函数, 基本形式如下

$$K(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right) \quad (9)$$

式中,  $\sigma > 0$  是需要确定的核参数,  $\|x - x'\|$  代表向量空间中点  $x$  与点  $x'$  的距离. Gauss 核由如下性质: 1) 有界性:  $0 < K(x, x') \leq 1$ ; 2) 可旋转性:  $K(x, x')$  只与  $\|x - x'\|$  有关, 而与  $x$  的具体位置无关; 3) 单调性:  $\|x - x'\|$  不变时,  $K(x, x')$  随着  $\sigma$  的增大而增大.

控制着拟合模型的线性程度,  $\sigma$  越大, 模型越接近线性<sup>[32]</sup>. 而对于复杂过程, 线性模型的一般均存在拟合不足, 因此  $\sigma$  应限制在一定范围内. Cristianini 和 Campbell 等<sup>[31]</sup> 证明了对于基于 Gauss 核的 SVM, 其泛化误差  $R$  的上界是  $\sigma$  的光滑函数, 但是给出的  $R$  的上界极难计算.

考察 SVM 回归方程式 (8), 对于某一确定的  $x$ , 所有的  $l$  个样本点  $x_i$ , 通过  $K(x_i, x)$ , 与常数  $\bar{a}_i$  和  $\bar{a}_i^*$  共同形成了  $f(x)$ . 根据 Gauss 核的单调性, 当  $\|x_i - x\|$  固定时,  $\sigma$  越小,  $K(x_i, x)$  越小,  $x_i$  对  $f(x)$  的影响也越小. 由于复杂过程的响应输出变化一般比较大, 为产生较准确的预测输出  $f(x)$ , 距  $x$  较远的  $x_i$  对  $f(x)$  的影响应尽可能的小.

此外, 由于  $0 < K(x_i, x) \leq 1$ , 因此也可以看作  $x_i$  对  $f(x)$  的影响权值  $w_i$ , 即

$$w_i = K(x_i, x) \quad (10)$$

结合式 (9), 有

$$= \left[ \frac{\|x_i - x\|^2}{2 \ln K(x_i, x)} \right]^{\frac{1}{2}} = \left[ \frac{\|x_i - x\|^2}{2 \ln w_i} \right]^{\frac{1}{2}} \quad (11)$$

对于式 (11) 中的  $\|x_i - x\|^2$ , 根据 RSM 一般做法, 可首先将各因子线性变换至  $[-1, 1]$  内, 有  $x = [x_1, x_2, \dots, x_k]^T \in [-1, 1]^k$ , 可行域变成长度为 2 的  $k$  维超立方体, 距中心点  $x_0 = [0, 0, \dots, 0]^T$  最远的距离  $d_{\text{radi}}$  是超立方体对角线长度的一半, 即

$$d_{\text{radi}} = \sqrt{k} \tag{12}$$

在考察距  $x$  较远的  $x_i$  对  $f(x)$  的影响时,若以  $d_{\text{radi}}$  作为基准距离,则有

$$x_i - x^2 = d_{\text{radi}}^2 = k \tag{13}$$

综上所述,这里给出确定核参数 范围的方法如下:首先给定距离  $x$  为  $d_{\text{radi}}$  的远端点  $x_i$  对  $f(x)$  的影响权值  $w_i$  的上界  $w_{\text{max}}$  和下界  $w_{\text{min}}$ ,有

$$0 < w_{\text{min}} < w_i < w_{\text{max}} < 1 \tag{14}$$

然后由式 (11)、(13)、(14) 可得 的上限和 下限:

$$\left( \frac{-k}{2 \ln w_{\text{min}}} \right)^{\frac{1}{2}} < < \left( \frac{-k}{2 \ln w_{\text{max}}} \right)^{\frac{1}{2}} \tag{15}$$

图 1 给出了不同  $k$  值时,  $w$  值与 的对照曲线,具体应用时也可查图得到 的范围.

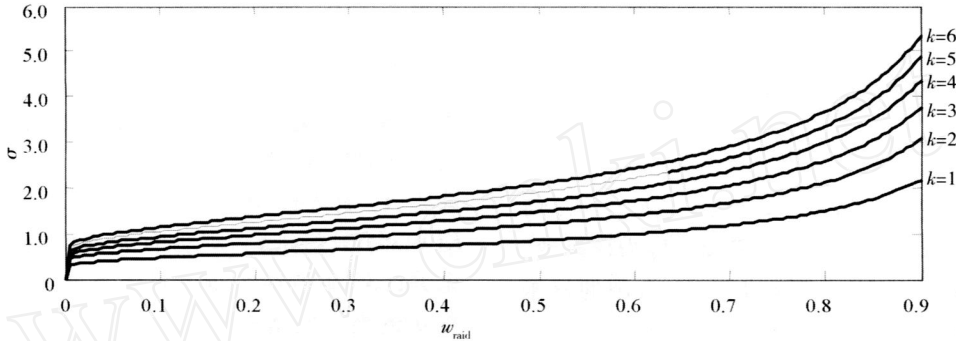


图 1  $w_{\text{radi}}$  与 的对照曲线

Fig. 1 Diagram between  $w_{\text{radi}}$  and

$w_{\text{max}}$  应根据样本量  $l$  的大小来选取,并且与  $l$  成反比,  $w_{\text{max}}$  的选择更多地影响的是搜索最优的效率,而不会错过最优的 .

式 (15) 虽然大幅度地缩小了 值范围,但是不足以确定准确的 值. 考察最优化问题式 (4), 式中  $\frac{1}{2} w^2$  意味着置信区间,  $\sum_{i=1}^l (y_i - \hat{y}_i)^2$  意味着拟合误差之和,  $\frac{1}{2} w^2 + C \sum_{i=1}^l (y_i - \hat{y}_i)^2$  则意味着 SVM 泛化误差  $R(\cdot)$  的上界. 在样本数据集不变、参数 与  $C$  确定的情况下,泛化误差的上界是关于核参数 的光滑函数,即

$$P(\cdot) = \frac{1}{2} w^2 + C \sum_{i=1}^l (y_i - \hat{y}_i)^2 \tag{16}$$

因此选择使式 (16) 最小的 值,就可以在其他参数固定的条件下实现泛化误差的上界最小. 由于 值的理论范围是  $> 0$ ,若采用穷举法搜索最优的 势必效率低下,这时可首先采用式 (15) 确定的大概范围,再利用穷举法搜索使  $P(\cdot)$  最小的 .

2) 与  $C$  的选择

参数 越大,支持向量的数目越少,模型也就越平坦; 越小,支持向量的数目越多,模型复杂性越高,也越容易出现“过学习”. 参数  $C$  起着在模型平坦性与经验误差之间的折衷作用. 若  $C$  较大,即对超出 - 不敏感区的样本实施较严厉的

惩罚,那么将增大模型的复杂性;反之,若  $C$  较小,则允许相对大的经验风险,而模型平坦度将增加.

对于等方差过程,Cherkassky和 Ma<sup>[19, 33]</sup> 提出采用经验公式

$$= 3 \sqrt{\frac{\ln l}{l}} \tag{17}$$

和

$$C = \max(|\mu_y + 3 \sigma_y|, |\mu_y - 3 \sigma_y|) \tag{18}$$

来确定 和  $C$  其中  $\sigma_p^2$  代表过程噪声方差,  $\mu_y$  代表过程输出的平均值,  $\sigma_y$  代表过程输出的标准差. 该方法既考虑了样本的分布,同时又避免了离群值的影响,确定的 和  $C$  接近其最优值,较适合于 RSM 的模型拟合. 但 Cherkassky和 Ma 给出的  $\sigma_p^2$  估计较为复杂,从样本数据去估计  $\sigma_p^2$  较困难,而且存在偏差. 事实上,在 RSM 中,对  $\sigma_p^2$  的估计可以采用实验方法来完成. 对于等方差过程,可以在可行域的任意一个因子组合处做若干次重复实验来估计  $\sigma_p^2$ . 设在因子组合  $x_c$  处进行了  $n$  次重复实验,得到  $y_{c1}, y_{c2}, \dots, y_{cn}$  个响应输出,则可利用样本方差  $S_p^2$  作为过程噪声的方差  $\sigma_p^2$  的估计

$$S_p^2 = \sigma_p^2 = \frac{1}{n-1} \sum_{i=1}^n (y_{ci} - \bar{y}_c)^2 \tag{19}$$

其中,  $\bar{y}_c = \frac{1}{n} \sum_{i=1}^n y_{ci}$  且  $S_p^2$  是  $\sigma_p^2$  的无偏估计量.

### 3.2 基于 SVM 的复杂过程 RSM 的模型拟合方法描述

假设已经通过某种设计方法 (如等间距空间网格取样法) 得到了实验样本集:

$$T = \{ (x_1, y_1), (x_2, y_2), \dots, (x_l, y_l) \} \quad (20)$$

则可以采用如下的方法来得到 SVM 拟合方程:

- 1) 采用式 (19) 估计过程误差  $\rho$ , 也可以根据工程经验确定  $\rho$ ;
- 2) 根据  $\rho$  和样本集  $T$ , 利用式 (17)、式 (18) 计算 SVM 的  $\gamma$ -不敏感参数和惩罚系数  $C$ ;
- 3) 选择 Gauss 径向基核作为核函数, 指定  $x$  对于  $f(x)$  影响权值的上界和下界, 根据图 1 或式 (15) 确定核函数的参数  $\sigma$  的大致范围;
- 4) 在  $\sigma$  的范围内进行搜索, 使式 (16) 中的泛化误差的上界  $P(\sigma)$  最小, 得到最优的  $\sigma$  值;
- 5) 利用样本集  $T$  和参数  $\gamma$ 、 $C$  进行 SVM 拟合, 得到拟合方程, 计算拟合误差、泛化误差, 识别出样本集中的支持向量的集合, 为后续 RSM 过程优化做准备.

## 4 算例研究

### 4.1 评价指标的选择

为评价方法的泛化性能, 这里选择与样本输

出值的度量无关的标准泛化误差 (standard error of prediction, SEP) 作为评价指标

$$SEP(\%) = \frac{1}{y} \sqrt{\frac{\sum_{i=1}^l (\hat{y}_i - y_i)^2}{l}} \times 100\% \quad (21)$$

式中,  $l$  代表测试集的样本量,  $\hat{y}_i$  代表方法在测试点上的估计值,  $y_i$  代表测试点的实际值,  $\bar{y}$  代表所有测试点实际值的均值.

### 4.2 核函数及参数选择方法有效性的验证

算例 1 设过程  $S_1$  的响应  $y$  受单个因子  $x$  的影响, 且  $x \in [-10, 10]$ .  $y$  与  $x$  之间存在如下关系

$$y = 10 \frac{\sin x}{x} \quad (22)$$

设过程噪声  $\epsilon$  为等方差分布, 且  $\epsilon \sim N(0, \rho)$ . 式 (22) 是一个常用的测试函数, Vapnik<sup>[16]</sup>, Cherkassky 和 Ma<sup>[19]</sup>, 邓乃扬和田英杰<sup>[34]</sup> 等学者均采用过该系列函数对各类 SVM 核函数及参数选择方法进行过对比研究.

表 1 给出了不同样本量时与不同  $\rho$  时各方法的 SEP 值. 表中 MinSEP 是通过穷举法搜索得到的最小的 SEP 值; ProSEP 是采用本文所提的核函数及参数选择方法而得到的 SEP 值; RndSEP 是随机选择核参数  $\sigma$  而得到的 SEP 值.

表 1 不同  $l$  或  $\rho$  时各种 SEP (%) 的对比

Table 1 Comparison of SEPs (%) with different  $l$  or  $\rho$

$l$	MinSEP	ProSEP	RndSEP	$\rho$	MinSEP	ProSEP	RndSEP
30	9.69	11.61	27.6	1.1	45.81	50.16	60.60
40	10.75	11.11	17.28	0.9	21.33	28.29	47.17
50	8.68	10.30	114.73	0.7	26.96	30.82	40.74
60	8.89	9.67	83.10	0.5	21.75	23.76	33.13
70	6.44	8.88	17.30	0.3	9.69	11.61	27.60
80	6.82	8.40	31.78	0.1	4.69	5.15	172.17
平均	8.55	10.00	48.63	平均	21.71	24.97	63.57

### 4.3 经典 RSM, 基于 SVM、KFR、ANN 的 RSM 模型拟合的比较

为了能够全面地研究各类 RSM 方法的性能, 这里给出了一个复杂的例子:

算例 2 设要考察的工业过程  $S_2$  的响应  $y$  受两个因子  $x_1, x_2$  的影响,  $x_1, x_2$  均为可控因子, 且  $x = [x_1, x_2]^T \in [-3, 3] \times [-3, 3]$ .  $y$  与  $x_1, x_2$  之间存在如下关系

$$y = f(x_1, x_2) + \epsilon$$

$$= 3(1 - x_1)^2 \exp(-x_1^2 - (x_2 + 1)^2) - 10\left(\frac{x_1}{5} - x_1^3 - x_2^5\right) \exp(-x_1^2 - x_2^2) - \frac{1}{3} \exp(-x_1 + 1)^2 - x_2^2 + \epsilon \quad (23)$$

过程的噪声  $\epsilon$  为等方差分布, 且  $\epsilon \sim N(0, \rho)$ . 响应的图形如图 2 所示, 其分布较为复杂, 包

括三个极大值、三个极小值和三个鞍点,符合关于复杂过程的界定. 设过程初始点位于  $x_0 = [-1.1, 0.4]^T$ , 此时  $\mu_y = -2.5266$

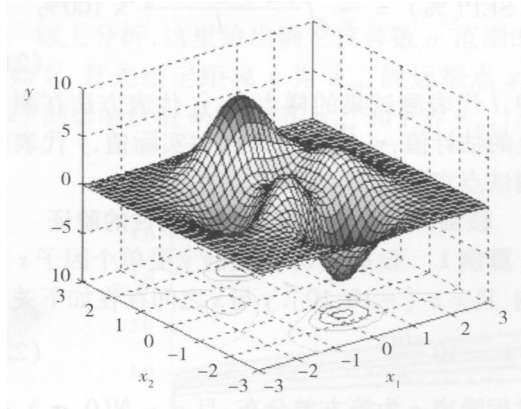


图 2  $S_2$  的响应曲面

Fig 2 Response surface of  $S_2$

1) 经典 RSM 的结果

为说明经典 RSM 的特点, 围绕初始点, 选择了两个均具有最速上升方向的、因子位级组合进行拟合, 此时过程误差  $\sigma_p = 0.05$  保持不变. 表 2 给出了最终寻优结果, 图 3 在过程等高线图上给出了的两次实验路线.

表 2 经典 RSM 优化结果

Table 2 Results of traditional RSM

序号	1	2
因子位级	$x_1: -1.0 \sim -1.2$ $x_2: 0.4 \sim 0.2$	$x_1: -1.0 \sim -1.2$ $x_2: 0.6 \sim 0.4$
实验次数	44	29
优化结果	$y^* = 3.8166$	$y^* = 8.1442$
最优点类型	局部最优点	全局最优点

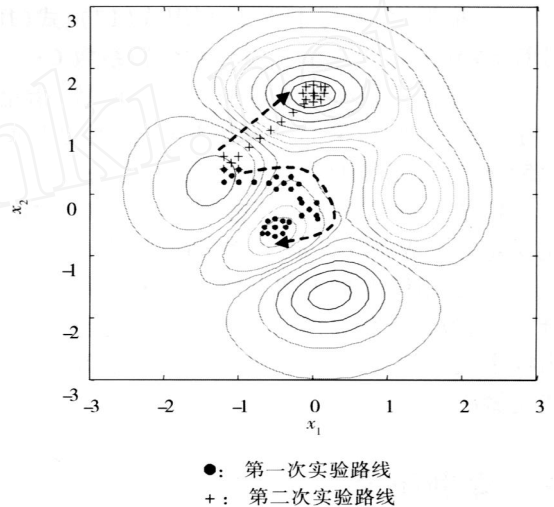
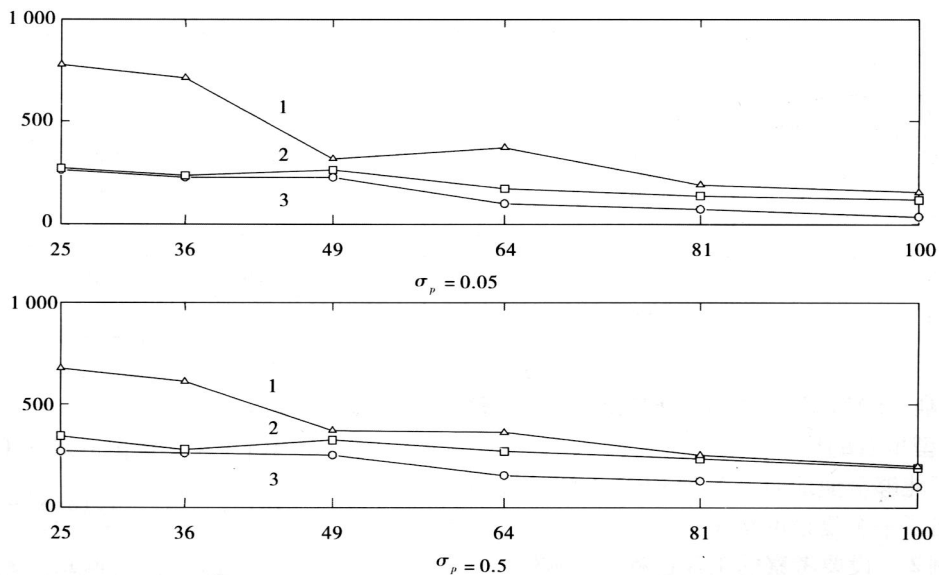


图 3  $S_2$  等高线图及经典 RSM 优化结果

Fig 3 Contour lines of  $S_2$  and optimal results of traditional RSM



1: ANN 的最小 SEP(%) 值 2: KFR 的最小 SEP(%) 值 3: SVM 的 SEP(%) 值  
图 4 不同  $l$  与  $\sigma_p$  时 SVM、KFR、ANN 的泛化误差对比

Fig 4 Comparison of generalized errors of SVM, KFR, ANN with different  $l$  or  $\sigma_p$

2) 基于 SVM, KFR, ANN 的结果

采用等间距的空间网格取样法进行实验设计, 实验点包括过程  $S_1$  可行域的边缘. 样本量  $l$  选为 3 个水平:  $l = 36, 64, 100$  过程误差的标准差  $\sigma_p$

选为 2 个水平  $\sigma_p = 0.05, 0.5$  选择等间距的  $60 \times 60 = 3600$  个空间网格点作为测试集计算 SEP. 图 4 给出了各种方法的 SEP 值, 图 5 给出了各种方法预测出的响应曲面.

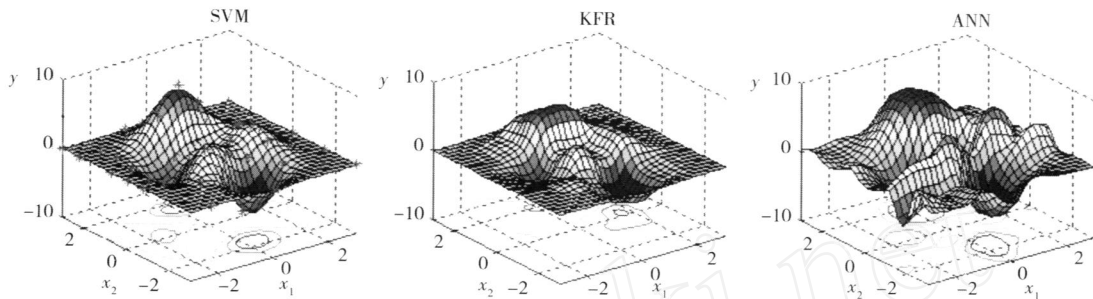


图 5 SVM、KFR、ANN ( $l = 64, \sigma_p = 0.05$ ) 的预测响应曲面

Fig. 5 Predicted response surfaces of SVM, KFR, ANN ( $l = 64, \sigma_p = 0.05$ )

4.4 基于 SVM 的 RSM 对过程误差为非正态的适用性

为进一步考察基于 SVM 的 RSM 模型拟合方法, 将过程的误差形式改为以 0 为中心的均匀分布, 表 3 给出了  $l = 100, \sigma_p = 0.05, 0.5, 1$  时的方法的 SEP 值, 表中 Dev 表示 ProSEP 与 MinSEP 的偏离率. 图 6 给出了  $\sigma_p = 0.05$  时的 SVM 拟合曲面及支持向量分布.

表 3 均匀分布过程的 SVM 的 SEP 值

Table 3 SEPs of SVM for uniform distributed process

$l$	$\sigma_p$	ProSEP (%)	MinSEP (%)	Dev (%)
100	0.05	27.28	27.28	0
	0.5	108.14	107.27	0.82
	1	179.76	177.84	1.08
平均		105.06	104.13	0.63

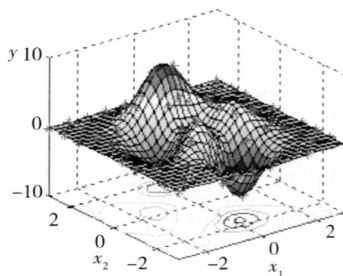


图 6 均匀分布过程的 SVM 的预测曲面

Fig. 6 Predicted surface of SVM for uniform distributed process

4.5 基于 SVM 的 RSM 对因子间带有约束的过程的适用性

为考察基于 SVM 的 RSM 模型拟合对于因子间带有约束条件的过程的适用性, 假设过程  $S_2$  的因子间存在如下约束条件:

$$s.t. \begin{cases} -\frac{1}{2} \exp(x_1) + x_2 + \frac{3}{2} & 0 \\ x_1 + x_2 + 1 & 0 \\ x_1^2 + x_2 - 3 & 0 \end{cases} \quad (23)$$

过程等高线图如图 7 所示, 其中被虚线围起来的区域即为因子的可行域. 表 4 给出了表 3 给出了  $l = 52, \sigma_p = 0.05, 0.5, 1$  时的方法的 SEP 值. 作为对比, 图 8 给出了 SVM 预测曲面的等高线图及实验点、支持向量分布.

表 4 因子间带有约束的过程的 SVM 的 SEP 值

Table 4 SEPs of SVM for factors with constraints process

$l$	$\sigma_p$	P_SEP (%)	Min_SEP (%)	Dev (%)
52	0.05	22.94	22.85	0.39
	0.5	21.94	20.36	7.76
	1	31.52	29.03	8.59
平均		25.47	24.08	5.58

5 结果讨论

5.1 核函数及参数选择方法分析

(1) 分析表 1 可以发现: 在  $\sigma_p$  不同取值条

件下,采用本文方法得到的 SEP值与最小的 SEP值相接近:  $p$  不变时, ProSEP与 MinSEP的偏离率平均为 13.62%;而  $l$  不变时, ProSEP与 MinSEP的偏离率平均为 15.87%,远小于 RndSEP值与 MinSEP值的平均偏离率(分别为 472.75%和 668.31%);

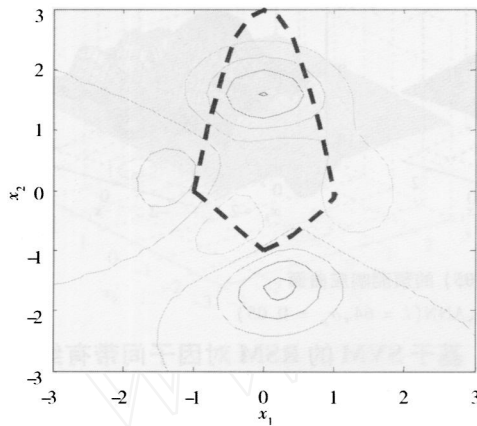


图 7 存在约束的  $S_2$  可行域

Fig 7 feasible zones of  $S_2$  for factors-with-constraints process

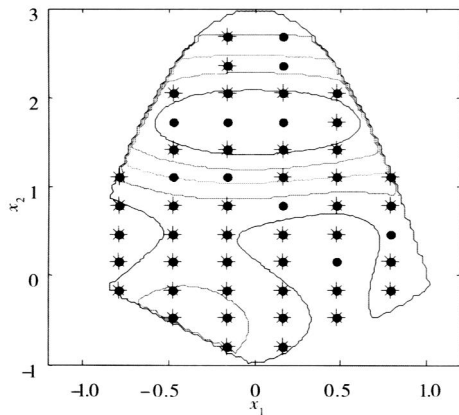


图 8 存在约束的过程 SVM 预测的等高线图

Fig 8 Predicted surface of SVM for factors-with-constraints process

2) 采用本文方法得到的 SEP值,反映了最小 SEP值随  $l$ 或  $p$ 变化的趋势,二者存在较强的线性相关关系.在不同  $l$ 和  $p$ 时, MinSEP和 ProSEP的 Pearson 相关系数分别为 0.900(P-Value = 0.014)和 0.993(P-Value < 0.001),在 95%置信水平线性相关;

(3) 由于存在一定的偏离率,本文方法并不能得到最小的泛化误差.其原因部分是过程误差,但更主要的是因为在式(15)中,  $P(\cdot)$ 代表的是泛化误差的上界,而使泛化误差的上界最小的

优化方法,并不一定能够得到最小的泛化误差.因此本文方法是一种经验性方法.另外,由于 Gauss核函数是关于轴线  $x = x$ 的对称可旋转函数,因此当  $x$ 为一维时的性质同样适用于多维  $x$ 时的情形.

上述分析说明,本文所提出的核函数及参数的选择和优化方法可以得到与最小泛化误差较为接近的泛化误差,而且计算方法简单,具有较强的实用性.

### 5.2 各类 RSM 模型拟合方法的对比

(1) 通常认为影响经典 RSM的主要有初始点位置、过程误差,但是表 2与图 3说明,在初始点、过程误差不变时,经典 RSM由于因子初始位级的略微不同而导致了截然不同的结果.说明初始点位级对实验过程和优化结果有显著影响.由于两种相邻位级均存在最速上升方向,且事先无法知晓哪个方向可以达到全局最优,也就无法确定最优的位级,因此说,这种影响是难以避免的;

(2) 图 4说明,对于各类 RSM的模型拟合方法,相同时,对于泛化误差, SVM 最小, KFR 次之, ANN 最大;泛化误差相接近时,对于所需的样本量, SVM 最小, KFR 次之, ANN 最大.  $p = 0.05$ 时, SVM 的 SEP值的均值比后两者分别降低 23.85%和 64.32%;对比  $l = 36$ 和  $l = 64$ ,  $l = 49$ 和  $l = 100$ ,  $l = 64$ 和  $l = 100$ 时 SVM、KFR、ANN的 SEP值发现, SVM的平均样本量分别比后两者降低了 32.89%和 39.02%;此外,随着样本量的增大, SVM的 SEP值减幅最大,  $p = 0.05$ 时,样本量从  $l = 25$ 增至  $l = 100$ 时, SVM、KFR、ANN的 SEP值分别下降了 89.37%、58.37%和 80.41%;在  $p$ 固定时,影响 SVM模型泛化能力的主要是  $l$ 的大小,而与样本点位置无关;

(3) 图 5说明,在  $l$ 和  $p$ 相同时,对于响应过程的重现能力, SVM最佳, KFR次之,而 ANN则最差.或者说, ANN模型的复杂程度最高, SVM模型的光滑程度最高. ANN复杂的响应曲面形状导致了拟合模型的“放大效应”(即预测的局部极值大量增加),而 SVM则在最小化经验风险和模型的复杂度之间取得了平衡,响应曲面较为光滑,在相邻两个样本点之间最多出现一个局部极值,不存在“放大效应”,可见,在实验点位置合适、样本量足够的情况下, SVM可以较为忠实地重现实际曲



面,有利于 RSM 寻优.

### 5.3 基于 SVM 的 RSM 对于非正态过程或因子间带有约束的过程的适用性

(1) 表 3 说明,在各  $\rho$  与  $l = 100$  的组合下,所提方法得到的 SEP 值、最小 SEP 值与过程误差为正态分布时对应的 SEP 值较为接近,同时图 6 的预测曲面形状也显示出良好的重现性能. 说明利用 SVM 拟合,不一定要要求过程误差必须服从正态分布,只要其分布形式对称即可,提出的 SVM 核函数及其参数选择方法也适用于非正态的对称分布过程;

(2) 表 4、图 7、图 8 说明,对于因子间存在约束的复杂过程,基于 SVM 的 RSM 拟合模型具有较低的泛化误差和良好的预测曲面的重现能力,同时,所提方法的 SEP 值与最小 SEP 值相接近. 说明了方法对于因子间存在约束条件的复杂过程的适用性.

## 6 结 论

本文将 SVM 引入到了复杂过程的 RSM 模型拟合. 算例分析表明,所提的 SVM 核函数及参数选择方法能够大幅度的提高拟合模型的泛化能力,其泛化误差接近于最小值,且不需要额外的测试样本; SVM 对拟合模型形式、因子是否约束、过程误差分布、实验点分布无严格要求,在有限样本量时泛化性能明显优于 KFR、ANN,因此应用于复杂过程具有显著优势,且由于 SVM 拟合方程解析形式简单,较 KFR 与 ANN 更有利于后续的过程优化,说明了基于 SVM 的 RSM 模型拟合方法的适用性与优势. 由于完整的 RSM 包括模型拟合、过程优化、实验设计等步骤,因此如何根据 SVM 拟合模型的特点,提出复杂过程的过程优化、实验设计方法将是本文进一步的研究方向.

### 参 考 文 献:

- [1] Box G E P, Wilson K B. On the experimental attainment of optimum conditions[J]. Journal of Royal Statistical Society, 1951, B (13): 1—45.
- [2] Myers R H. Response surface methodology—current status and future directions[J]. Journal of Quality Technology, 1999, 31 (1): 30—44.
- [3] Vining G G, Myers R H. Combining Taguchi and response surface philosophies: A dual response approach[J]. Journal of Quality Technology, 1990, 22 (1): 38—45.
- [4] Carlyle W M, Montgomery D C, Runger G C. Optimization problems and methods in quality control and improvement[J]. Journal of Quality Technology, 2000, 32 (1): 1—17.
- [5] Young-Hyun K, Kwang-Jae K, Chi-Hyuck J. A new loss function-based method for multiresponse optimization[J]. Journal of Quality Technology, 2005, 37 (1): 50—59.
- [6] Myers R H, Montgomery D C, Vining G G. Generalized Linear Models with Application in Engineering and the Sciences[M]. New York: John Wiley & Sons, 2002. 10—90.
- [7] Robinson T J, Wulff S S, Montgomery D C, Khuri A I. Robust parameter design using generalized linear mixed models[J]. Journal of Quality Technology, 2006, 38 (1): 65—75.
- [8] Vining G G, Bohn L L. Response surface for the mean and variance using a nonparametric approach[J]. Journal of Quality Technology, 1998, 30 (3): 282—291.
- [9] He Z, Xiao Y-X. Regression-based artificial neural network methodology in response surface methodology[J]. Transactions of Tianjin University, 2004, 10 (2): 153—157.
- [10] Anjum M F, Tasadduq I, Al-Sultan K. Response surface methodology: A neural network approach[J]. European Journal of Operational Research, 1997, 101: 63—65.
- [11] Al-Ahmari A M A. Predictive machinability models for a selected hard material in turning operations[J]. Journal of Materials Processing Technology, 2007, 190 (1): 305—311.
- [12] Palanikumar K. Modeling and analysis for surface roughness in machining glass fibre reinforced plastics using response sur-

- face methodology[J]. *Materials & Design*, 2007, 28(10): 2611—2618.
- [13] Liew Abdullah A G, Sulaiman N M, Arroua M K, *et al* Response surface optimization of conditions for clarification of carambola fruit juice using a commercial enzyme[J]. *Journal of Food Engineering*, 2007, 81(1): 65—71.
- [14] Katapodis P, Christakopoulou V, Kekos D, *et al* Optimization of xylanase production by *Chaetomium thermophilum* in wheat straw using response surface methodology[J]. *Biochemical Engineering Journal*, 2007, 35(1): 136—141.
- [15] Psomas S K, Liakopoulou-Kyriakides M, Kyriakidis D A. Optimization study of xanthan gum production using response surface methodology[J]. *Biochemical Engineering Journal*, 2007, 35(3): 273—280.
- [16] Vapnik V N. *The Nature of Statistical Learning Theory*[M] (2nd Edit). Berlin: Springer, 1999. 1—30.
- [17] 张学工. 关于统计学习理论与支持向量机[J]. *自动化学报*, 2000, 26(1): 32—42.  
Zhang Xuegong. Introduction to statistical learning theory and support vector machines[J]. *ACTA Automatica Sinica*, 2000, 26(1): 32—42. (in Chinese).
- [18] Katagiri SAbe S. Incremental training of support vector machines using hyperspheres[J]. *Pattern Recognition Letters*, 2006, 27(13): 1495—1507.
- [19] Cherkassky V, Ma Y. Practical selection of SVM parameters and noise estimation for SVM regression[J]. *Neural Networks*, 2004, 17(1): 113—126.
- [20] De Martino F, Gentile F, Esposito F, *et al* Classification of MRI independent components using IC-fingerprints and support vector machine classifiers[J]. *Neuro Image*, 2007, 34(1): 177—194.
- [21] 魏宏业, 吕永波, 何 葳等. 基于数据挖掘的企业合作伙伴的寻求方法研究[J]. *管理科学学报*, 2004, 7(1): 60—64.  
Wei Hongye, Lu Yongbo, He Wei, *et al* Analysis of method for enterprise to search partners base on data mining[J]. *Journal of Management Sciences in China*, 2004, 7(1): 60—64. (in Chinese).
- [22] Lorena A C, Andre C P. Protein cellular localization prediction with support vector machines and decision trees[J]. *Computers in Biology and Medicine*, 2007, 37(2): 115—125.
- [23] 李 栋, 王洪礼, 杜忠晓等. 城市生活用水量的支持向量回归预测[J]. *天津大学学报(社会科学版)*, 2006, 8(1): 64—67.  
Li Dong, Wang Hongli, Du Zhongxiao, *et al* SVM-Based prediction of city municipal and domestic water consumption[J]. *Journal of Tianjin University (Social Sciences)*, 2006, 8(1): 64—67. (in Chinese).
- [24] Cannas B, De Igo R S, Fanni A, *et al* Support vector machines for disruption prediction and novelty detection at JET[J]. *Fusion Engineering and Design*, 2007, 82(5): 1124—1130.
- [25] Stanislaw O, Konrad G. Forecasting of the daily meteorological pollution using wavelets and support vector machine[J]. *Engineering Applications of Artificial Intelligence*, 2007, 20(6): 745—755.
- [26] Muller K-R, Mika SRatsch G. An introduction to kernel-based learning algorithms[J]. *IEEE Transactions on Neural Networks*, 2001, 12(2): 181—201.
- [27] Cristianini N, Shawe-Taylor J. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*[M]. Cambridge: Cambridge University Press, 2000. 100—200.
- [28] Soares C, Brazdil P, B Kuba P. A meta-learning method to select the kernel width in support vector regression[J]. *Machine Learning*, 2004, 54(3): 195—209.
- [29] Keerthi S L, Lin C-J. Asymptotic behaviors of support vector machines with gaussian kernel[J]. *Neural Computation*, 2003, 15(7): 1667—1689.
- [30] Cherkassky V, Ma Y. Comparison of model selection for regression[J]. *Neural Computation*, 2003, 15(7): 1691—1714.
- [31] Cristianini N, Campbell C, Shawe-Taylor J. Dynamically Adapting Kernels in Support Vector Machines[M]. *Advances in Neural Information Processing Systems 11*. Cambridge, MA: MIT Press, 2004. 204—210, 1999. 20—50.
- [32] Burges C A. A tutorial on support vector machines for pattern recognition[J]. *Data Mining and Knowledge Discovery*, 1998, 2(2): 1001—1030.
- [33] Cherkassky V, Ma Y. Selection of Meta-parameters for Support Vector Regression[C]. in *Proceedings of the International Conference on Artificial Neural Networks (ICANN)*, Madrid, Spain, 2002. 687—693.

[34] 邓乃扬, 田英杰. 数据挖掘中的新方法: 支持向量机 [M]. 北京: 科学出版社, 2004. 268—270.

Deng Naiyang, Tian Yingjie. A New Approach in Data Mining: Support Vector Machine [M]. Beijing: Science Press, 2004. 268—270. (in Chinese).

## SVM-based RSM model fitting approach

*CUI Qing-an*<sup>1,2</sup>, *HE Zhen*<sup>2</sup>, *CUI Nan*<sup>3</sup>

1. Institute of Management Engineering, Zhengzhou University, Zhengzhou 450001, China;

2. School of Management, Tianjin University, Tianjin 300072, China;

3. School of Economics and Management, Wuhan University, Wuhan 430072, China

**Abstract:** When a complex process is featured with multi-extreme of quality responses as well as high order interactions and constraints among influential factors, parametric response surface method (RSM) fails to fit the real surface and is hard to achieve global optimization; While non-parametric RSM results in poor generalization performance when the sample size is finite and is hard to optimize the response as well. In this paper, the model fitting phase of RSM is described as a sort of restricted small-sample learning problem which is able to actively gain sample points. Then, a Support Vector Machine (SVM) based method is proposed for the model fitting phase of RSM. A practical method for selecting SVM kernel functions and parameters is put forward for RSM as well. The simulations show that, by using the proposed method to select kernel functions and parameters, the average deviation ratio of SVM generalized error from the exhaustively searched minimum is less than 20%. The SVM based RSM model fitting approach has no rigid restriction for the normality of the response and non-constraints among the factors. Furthermore, it outperforms the existing RSM approaches in generalization and surface reconstruction performance. Compared with non-parametric RSM, the average generalized error and the sample size of the proposed approach decrease by about 20% and 30% respectively. All these demonstrate the adaptability and superiority of the proposed approach.

**Key words:** support vector machine; response surface methodology; complex process; model fitting