

# 基于 SCE-UA 支持向量机的短期电力负荷预测模型研究

李刚\*, 程春田, 曾筠, 林剑艺

(大连理工大学 水电与水信息研究所, 辽宁 大连 116024)

**摘要:** 支持向量机(support vector machine, SVM)作为一种新颖的机器学习方法已成功应用于短期电力负荷预测,然而应用研究发现 SVM 算法性能参数的设置将直接影响负荷预测的精度.为此在对 SVM 参数性能分析的基础上,提出了 SCE-UA(shuffled complex evolution-University of Arizona)支持向量机短期电力负荷预测模型建模的思路及关键参数的选取,在建模过程中引入了径向基核函数,简化了非线性问题的求解过程,并应用 SCE-UA 算法辨识 SVM 的参数.贵州电网日 96 点负荷曲线预测的实际算例表明,所提 SCE-UA 支持向量机模型不仅克服了 SVM 参数选择的盲目性,而且能提高预测准确率,是一种行之有效的短期电力负荷预测模型.

**关键词:** 负荷预测;支持向量机;SCE-UA;相似日

**中图分类号:** TM74 **文献标志码:** A

## 0 引言

短期负荷预测对电力系统的安全、经济运行起着重要的作用,无论是电网的经济调度、水火协调、发电计划还是系统安全评估等,都需要有可靠的负荷预测数据作为前提.长期以来,国内外专家学者对短期电力负荷预测的理论和方法都做了大量研究工作,提出了许多先进方法.其中,人工神经网络具有良好的非线性逼近能力和自适应自学习功能,在短期负荷预测上得到了广泛的应用,是近十几年来研究和使用较多的一种方法,但神经网络模型在提高模型的泛化能力上缺乏理论上的突破<sup>[1]</sup>.近几年来一种新的学习机——支持向量机(support vector machine, SVM)模型在小样本、非线性、高维问题及泛化能力方面表现突出,得到了广泛关注,与神经网络学习方法不同,它以最小化结构风险代替了传统的经验风险,且实现了统计学中的 VC 维理论,对有限样本有较好的泛化性能<sup>[2,3]</sup>.

支持向量机包括用于解决分类问题的分类支持向量机和用于函数拟合的回归支持向量机<sup>[4]</sup>.

近来已经有一些将 SVR 用于短期电力负荷预测方面的研究<sup>[5~9]</sup>,且取得了较为理想的结果.然而,支持向量机在应用上仍存在一些困难,即如何设置影响算法性能的一些关键参数,如支持向量机自身的参数、核函数选取及相应的参数等.该问题也是将支持向量机应用于实际工程所碰到的热点问题,许多学者对此做了大量研究工作<sup>[1,10~13]</sup>,并应用于短期电力负荷预测,取得了较好的结果.本文提出 SCE-UA (shuffled complex evolution-University of Arizona) 算法优化参数的 SVM 短期电力负荷预测模型,对某实际电网进行日 96 点负荷预测.

## 1 SCE-UA 算法优化回归支持向量机模型

### 1.1 回归支持向量机<sup>[14,15]</sup>

给定一个训练集

$$\{(x_1, y_1), \dots, (x_i, y_i), \dots, (x_l, y_l)\} \subset \mathbf{R}^n \times \mathbf{R} \quad (1)$$

式中: $x_i$  为输入向量; $y_i$  是  $x_i$  对应的输出值; $l$  为样

本个数. 支持向量回归的基本思想是通过非线性映射  $\phi$  将输入空间数据映射到一个高维特征空间中, 然后在空间中进行线性回归, 即

$$f(\mathbf{x}) = \langle \mathbf{w}, \phi(\mathbf{x}) \rangle + b; \mathbf{w}, \mathbf{x} \in \mathbf{R}^n, b \in \mathbf{R} \quad (2)$$

可以把上述问题定义为对一个损失函数进行风险最小化的问题, 回归函数实际上就是在在一定约束条件下的最小化规则化风险泛函:

$$\frac{1}{2} \|\mathbf{w}\|^2 + C \cdot \frac{1}{l} \sum_{i=1}^l L_\epsilon(y_i, f(\mathbf{x}_i))$$

式中: 第 1 项称为规则化项, 可以使函数较为平坦, 提高泛化能力; 第 2 项称为经验风险泛函, 通过不同的损失函数确定; 常数  $C > 0$ , 它控制对样本超出误差  $\epsilon$  的惩罚程度.  $\epsilon$ -不敏感损失函数为

$$L_\epsilon(y_i, f(\mathbf{x}_i)) = \begin{cases} |y - f(\mathbf{x}_i)| - \epsilon; & |y - f(\mathbf{x}_i)| \geq \epsilon \\ 0; & |y - f(\mathbf{x}_i)| < \epsilon \end{cases} \quad (3)$$

引入松弛变量  $\xi_i$  和  $\xi_i^*$ , 式(3)可以重新描述为下面的优化问题:

$$\begin{aligned} \min \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ \text{s. t.} \quad & \begin{cases} y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - b \leq \epsilon + \xi_i \\ \langle \mathbf{w}, \mathbf{x}_i \rangle + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned} \quad (4)$$

这是一个二次优化问题, 可以通过其对偶问题来求解, 其对偶问题为

$$\begin{aligned} \max \quad & W(\alpha_i, \alpha_i^*) = -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l [(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle] - \epsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + \sum_{i=1}^l y_i (\alpha_i - \alpha_i^*) \\ \text{s. t.} \quad & \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0; \alpha_i, \alpha_i^* \in [0, C] \end{aligned} \quad (5)$$

求解后得到  $\alpha_i$  和  $\alpha_i^*$  的值,  $\mathbf{w}$  的表达式为

$$\mathbf{w} = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \phi(\mathbf{x}_i) \quad (6)$$

$$f(\mathbf{x}) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}) \rangle + b \quad (7)$$

对于非线性回归问题, 是通过非线性变换将输入向量  $\mathbf{x}_i$  映射到高维特征空间  $\phi(\mathbf{x}_i)$ , 转化为类似的线性回归问题加以解决. 其中  $K(\mathbf{x}_i, \mathbf{x}) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}) \rangle$  称为核函数, 描述了高维特征空间的内积, 在满足 Mercer 条件的情况下选取, 可以

有效避免空间中的“维数灾难问题”. 表 1 列出了较为常用的核函数.

表 1 核函数

Tab. 1 The kernel functions

核	表达式	参数
线性核函数	$\langle \mathbf{x}, \mathbf{x}_i \rangle$	
多项式核函数	$(\langle \mathbf{x}, \mathbf{x}_i \rangle + 1)^d$	$d$
径向基核函数	$\exp(-\ \mathbf{x} - \mathbf{x}_i\ ^2 / 2\sigma^2)$	$\sigma$
Sigmoid 核函数	$\tanh(b\langle \mathbf{x}, \mathbf{x}_i \rangle + c)$	$b, c$

## 1.2 SCE-UA 算法<sup>[16,17]</sup>

SCE-UA 算法是一种十分有效的进化算法, 其结合了单纯形法、随机搜索和生物竞争进化等方法的优点, 可以快速地搜索到连续型流域水文模型参数优化的全局最优解<sup>[18]</sup>. Sorooshian 等<sup>[19]</sup>和 Kuczrea<sup>[20]</sup>分别采用 SCE-UA 算法解决水文模型参数优选问题, 结果表明, SCE-UA 算法更有效, 鲁棒性更强, 收敛速度更快, 且收敛效果更佳. 所以, 本文采用 SCE-UA 算法优化 SVM 相关参数, 详细的步骤过程见文献<sup>[16]</sup>.

SCE-UA 算法中的参数可分为不确定性参数和确定性参数两个部分, 参数的设置会影响到优化的效果, 只有选择好算法中的参数, 才能使算法达到全局最优. 研究表明, SCE-UA 算法建议取  $m = 2n + 1, q = n + 1, \alpha = 1, \beta = 2n + 1$ . 其中  $m$  为复合形中点的个数,  $q$  为子复合形中点的个数,  $\alpha$  为复合形生成的连续子辈数目,  $\beta$  为复合形进化的代数,  $n$  为维数.

## 1.3 支持向量机的 SCE-UA 参数辨识

SVM 模型选择包括以下两个步骤: (1) 选取核函数; (2) 支持向量机参数 ( $C, \epsilon$ ) 及核函数对应参数的辨识.

选取合适的核函数可以大大简化具体问题的求解难度, 因此很多学者在核函数比较分析及如何选择上展开相关研究<sup>[15,20]</sup>. 研究表明, 径向基核函数相对而言是较为通用的核函数, 并且核函数本身参数太多影响模型参数的优选. 鉴于以上原因, 本文选择径向基核函数作为模型的核函数, 对应的 SVM 模型可以描述为

$$y = f(\mathbf{x} | C, \epsilon, \sigma) \quad (8)$$

上式表示对于给定的参数 ( $C, \epsilon, \sigma$ ), 输入向量  $\mathbf{x}$  就可以得到输出值  $y$ . 对于式(8)抽象的 SVM 模型,

为了准确模拟给定的训练集  $\{(x_1, y_1), \dots, (x_i, y_i), \dots, (x_l, y_l)\}$ , 模型中参数  $(C, \epsilon, \sigma)$  的选择非常重要, 这也就是 SVM 模型的参数辨识优选问题. 本文采用 SCE-UA 算法进行参数寻优, 其优化计算流程图如图 1 所示. 在 SCE-UA 算法搜索过程中, 以日负荷预测准确率最大为目标函数, 其计算公式如下:

$$E_d = \left( 1 - \sqrt{\frac{1}{n} \sum_{i=1}^n \left( \frac{y_i - y_i^*}{y_i^*} \right)^2} \right) \times 100\% \quad (9)$$

其中  $E_d$  为日负荷预测准确率,  $y_i$  为  $i$  时刻负荷预测值,  $y_i^*$  为  $i$  时刻负荷实际值,  $n$  为考核点数.

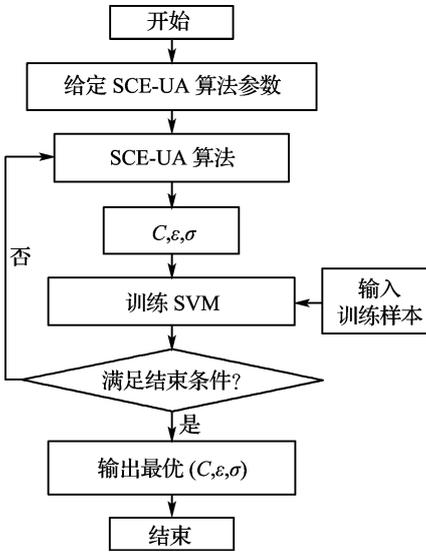


图 1 SCE-UA 算法优化 SVM 参数流程图

Fig. 1 Flow chart of optimizing SVM parameters by SCE-UA algorithm

## 2 基于 SCE-UA 支持向量机的短期负荷预测

### 2.1 预测步骤

短期电力负荷预测需要考虑的影响因素很多, 主要因素有日分类(正常日、节假日等)、星期类型(周一~周日)、日天气类型(晴、阴等)、日温度(最高、平均、最低)等, 而相似日的选取直接考虑各种相关因素. 本文通过负荷预测时间间隔差异、日类型差异和日温度差异综合度量选取多个相似日, 以相似日负荷数据作为样本对 SVM 进行训练, 最后用训练好的 SVM 对预测日 96 点负荷进行预测.

基于 SCE-UA 支持向量机的短期电力负荷预测步骤如下:

(1) 负荷数据预处理. 电力系统短期负荷预测模型在建模时需要大量的历史实际数据, 而实际数据大多来源于 SCADA 数据库, 由于采集过程中的各种原因会有多种错误产生, 历史负荷数据中往往包含非真实数据, 通常称为“不良数据”或“坏数据”. 数据预处理就是在利用这些数据之前, 先对其进行分析加工, 去除不规则数据和填补缺失数据, 消除不良数据或坏数据的影响. 具体处理方法参见文献[21].

(2) 基于相似日方法选择样本. 考虑了“近大远小”的原则, 取预测日前 4 个星期内的已知日作为待选的相似日, 选择星期类型、天气类型、日最高气温、日最低气温作为影响负荷变化的主要因素, 根据文献[22]提供的思想和方法将相应的影响因素映射到特征量映射表中. 根据映射以后的标准特征值确定日特征量的差异度、相似度, 选取相似日. 差异度和相似度计算公式如下:

$$d_{ij} = \sqrt{\sum_{k=1}^m |x_{ik} - y_{jk}|^2} \quad (10)$$

$$r_{ij} = \frac{\sum_{k=1}^m x_{ik} y_{jk}}{\sqrt{\sum_{k=1}^m x_{ik}^2} \sqrt{\sum_{k=1}^m y_{jk}^2}} \quad (11)$$

式中:  $d_{ij}$ 、 $r_{ij}$  分别为  $i$ 、 $j$  两天特征量的差异度和相似度;  $x_{ik}$ 、 $y_{jk}$  分别为  $i$ 、 $j$  两天第  $k$  个特征量;  $m$  为特征量总数.

(3) 归一化处理. 为避免出现计算饱和现象, 要对训练样本集中的负荷数据进行归一化处理, 使输入的负荷数据在  $[0, 1]$ ,  $t$  时刻负荷数据采用如下归一化公式:

$$\hat{L}_t = \frac{L_t - L_{\min}}{1.5L_{\max} - L_{\min}}; t = 1, 2, 3, \dots, 96 \quad (12)$$

其中  $L_t$ 、 $\hat{L}_t$  分别为  $t$  时刻原始负荷数据和归一化数据,  $L_{\max}$ 、 $L_{\min}$  分别为训练样本集中负荷的最大值和最小值.

(4) 将选取的相似日负荷数据作为 SCE-UA 支持向量机预测模型的训练样本集, 找出最优参数向量  $(C, \epsilon, \sigma)$ , 赋给 SVM, 再用样本数据对 SVM 进行训练.

(5) 用训练好的 SVM 对预测日的 96 点负荷

曲线进行预测.

### 2.2 算例分析

运用本文所提模型,对贵州电网实际负荷数据进行日 96 点负荷预测研究.取 2007-07-01 至 2007-08-20 的数据作为样本数据集,预测 2006-08-21 至 2006-08-31 的日 96 点负荷值.利用样本集中各日的气候数据(包括日平均温度、日最高温度、日最低温度、天气类型)和日 96 点负荷数据建立预测模型.

用 matlab 编制程序,设定 SCE-UA 算法优

化解空间为 3 维,分别对应 SVM 的 3 个参数( $C, \epsilon, \sigma$ ),复合形的个数为 10,每个复合形的顶点数为 7,子复合形中点的个数为 4,复合形生成的连续子辈数目为 1,复合形进化的代数数为 7. SVM 参数  $C$  取值范围为 $[0.1, 150]$ , $\epsilon$  取值范围为 $[0.01, 1]$ , $\sigma$  取值范围为 $[0.1, 10]$ .图 2 和 3 是该电网 2006-08-31 1 d 96 点负荷  $P$  及其偏差率  $e$  的预测结果,并和常规 SVM 方法预测的结果进行了比较.参数优化结果为  $C = 41.261, \epsilon = 0.010537, \sigma = 1.2031$ .

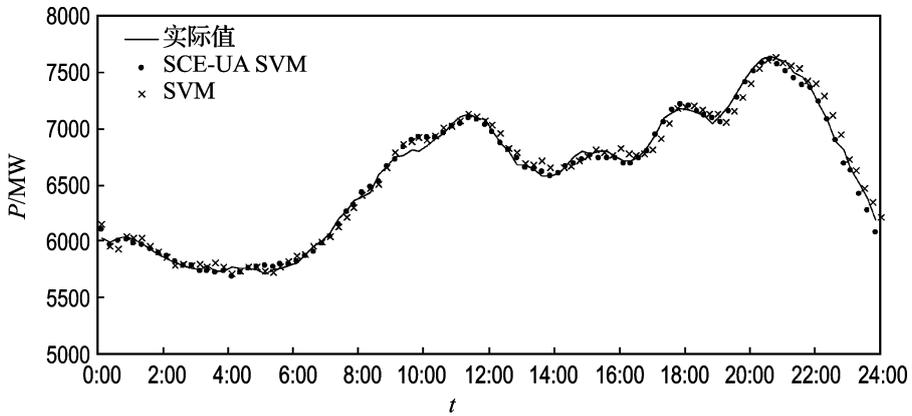


图 2 两种模型负荷预测曲线比较

Fig. 2 Comparison of load forecasting results of two models

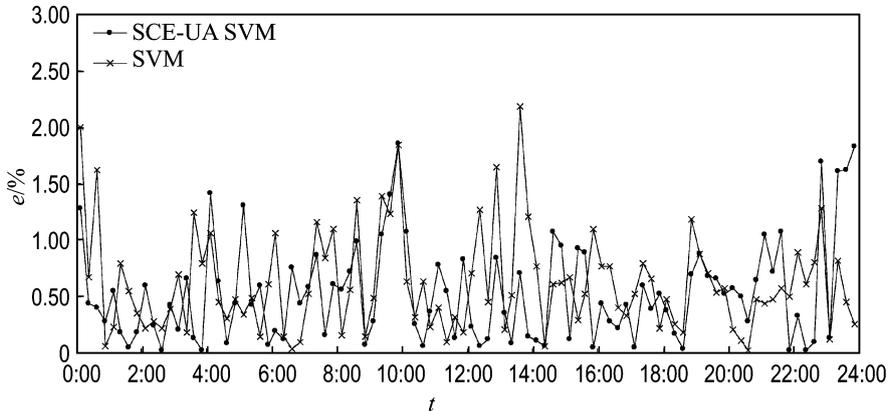


图 3 两种模型日 96 点负荷预测偏差率比较

Fig. 3 Comparison of load forecasting relative error at daily 96 points of two models

从图 2 和 3 可以看出, SCE-UA SVM 模型虽然在个别点的偏差率大于常规 SVM 模型,但是总体预测准确率要明显高于常规 SVM 模型.该日 96 点负荷预测偏差率基本都在 2% 以内,最大偏差率为 1.85%,平均偏差率为 0.53%,日负荷

预测准确率达到 99.32%,与支持向量机方法(最大偏差率为 2.18%,平均偏差率为 0.62%,日负荷预测准确率达到 99.13%)相比,具有更高的预测准确率.

表 2 列出了分别用本文所提出的 SCE-UA

SVM 模型和 SVM 模型对贵州电网 2006-08-21 至 2006-08-31 日负荷进行预测的结果. 从表中仿真结果可以看出, 两个模型在某些负荷点的预测偏差都较大, 而且各日最大偏差不是很稳定, 这主要是因为预测日的个别特征指标(例如最高温度、最低温度)与其相似日有较大的出入, 造成这些点

负荷较难预测准确. SCE-UA SVM 模型预测各日的最大偏差率平均值为 2.09%, 较 SVM 模型的 2.37% 降低了一些, 而且各日预测的平均偏差率变化幅度比 SVM 模型小, 同时可以获得较高的日负荷预测准确率(都达到了 99.25% 以上).

表 2 SCE-UA SVM 模型和 SVM 模型的负荷预测结果

Tab. 2 Load forecasting results of the SCE-UA SVM model and SVM model

日期	SCE-UA SVM			SVM		
	准确率/%	最大偏差率/%	平均偏差率/%	准确率/%	最大偏差率/%	平均偏差率/%
2006-08-21	99.25	3.38	0.52	99.20	4.05	0.57
2006-08-22	99.28	2.40	0.55	99.34	2.42	0.49
2006-08-23	99.29	2.00	0.56	99.26	2.87	0.57
2006-08-24	99.27	2.33	0.58	99.32	2.49	0.51
2006-08-25	99.25	2.20	0.59	99.25	1.86	0.60
2006-08-26	99.29	2.15	0.51	99.23	2.40	0.61
2006-08-27	99.41	1.43	0.46	99.37	1.48	0.49
2006-08-28	99.33	1.88	0.51	99.28	1.80	0.58
2006-08-29	99.40	1.70	0.47	99.19	2.81	0.64
2006-08-30	99.35	1.70	0.52	99.32	1.68	0.52
2006-08-31	99.32	1.85	0.53	99.13	2.18	0.62
均值	99.31	2.09	0.53	99.26	2.37	0.56

### 3 结 语

支持向量机(SVM)是结构风险最小化准则实现的一种近似方法, 有效避免了过学习问题, 而且将函数估计最终转化为二次规划问题, 所以其解在理论上是全局最优的. 在对 SVM 的参数性能进行分析的基础上, 本文提出了在国内外水文模型参数优选中应用十分广泛的 SCE-UA 算法来优化 SVM 参数, 并将其用于短期电力负荷预测. 仿真结果表明, 与目前普遍应用的通过交叉验证试算确定 SVM 参数的方法相比, 本文所提基于 SCE-UA 支持向量机的短期电力负荷预测模型, 不仅克服了 SVM 参数选择的盲目性, 而且具有较高的预测准确率, 在短期电力负荷预测中有广泛的应用前景.

### 参考文献:

[1] 谢宏, 魏江平, 刘鹤立. 短期负荷预测中支持向量机模型的参数选取和优化方法[J]. 中国电机工程学报, 2006, 26(22):17-22

[2] 李元成, 方廷健, 于尔铿. 短期负荷预测的支持向量机方法研究[J]. 中国电机工程学报, 2003, 23(6):

55-59

[3] 林剑艺, 程春田. 支持向量机在中长期径流预报中的应用[J]. 水利学报, 2006, 37(6):681-686

[4] VAPNIK V. *Statistical Learning Theory* [M]. New York: Wiley, 1998

[5] 蔡国伟, 杜毅, 李春山, 等. 基于支持向量机的中长期日负荷曲线预测[J]. 电网技术, 2006, 30(23):56-60

[6] 胡国胜, 任震. 基于支持向量机混合模型的短期负荷预测方法[J]. 高电压技术, 2006, 32(4):101-103

[7] 潘峰, 程浩中, 杨镜非, 等. 基于支持向量机的电力系统短期负荷预测[J]. 电网技术, 2004, 28(21):39-42

[8] GUO Y C, NIU D X, CHEN Y X. Support vector machine model in electricity load forecasting [C] // *Proceedings of the 2006 International Conference on Machine Learning and Cybernetics*. Piscataway: IEEE, 2006

[9] ZHANG M G. Short-term load forecasting based on support vector machines regression [C] // *2005 International Conference on Machine Learning and Cybernetics*. Piscataway: IEEE, 2005

[10] 吴宏晓, 侯志俭. 基于免疫支持向量机方法的电力

- 系统短期负荷预测[J]. 电网技术, 2004, **28**(23): 47-51
- [11] 张红梅, 卫志农, 龚灯才, 等. 基于粒子群支持向量的短期电力负荷预测[J]. 继电器, 2006, **34**(3): 28-31
- [12] PAI P F, HONG W C. Support vector machines with simulated annealing algorithms in electricity load forecasting [J]. **Energy Conversion and Management**, 2005, **46**(17):2669-2688
- [13] SUN Chang-yin, GONG Deng-cai. Support vector machines with PSO algorithm for short-term load forecasting [C] // **Proceedings of the 2006 IEEE International Conference on Networking, Sensing and Control, ICNSC '06**. Piscataway: IEEE Computer Society, 2006:676-680
- [14] SMOLA A J, SCHÖLKOPF B. A tutorial on support vector regression [J]. **Statistics and Computing**, 2004, **14**(3):199-222
- [15] HSU Chih-wei, CHANG Chih-chung, LIN Chih-jen. A practical guide to support vector classification [J]. **Bioinformatics**, 2003, **1**(1):1-15
- [16] DUAN Q Y, SOROOSHIAN S, GUPTA V. Effective and efficient global optimization for conceptual rainfall-runoff models [J]. **Water Resources Research**, 1992, **28**(4):1015-1031
- [17] DUAN Q Y, GUPTA V, SOROOSHIAN S. Shuffled complex evolution approach for effective and efficient minimization [J]. **Journal of Optimization Theory and Applications**, 1993, **76**(3): 501-521
- [18] DUAN Q Y, SOROOSHIAN S, GUPTA V. Optimal use of the SCE-UA global optimization method for calibrating watershed models[J]. **Journal of Hydrology**, 1994, **158**(1):265-284
- [19] SOROOSHIAN S, DUAN Q Y, GUPTA V. Calibration of rainfall-runoff models: application of global optimization to the Sacramento soil moisture accounting mode [J]. **Water Resources Research**, 1993, **29**(4):1185-1194
- [20] KUCZREA G. Efficient subspace probabilistic parameter optimization for catchment models[J]. **Water Resources Research**, 1997, **33**(1):177-185
- [21] 杨奎河. 短期电力负荷的智能化预测方法研究[D]. 西安:西安电子科技大学, 2004
- [22] 康重庆, 程旭, 夏清, 等. 一种规范化的处理相关因素的短期负荷预测新策略[J]. 电力系统自动化, 1999, **23**(18):32-35

## Research on short-term electricity load forecasting model using support vector machine based on SCE-UA algorithm

LI Gang\*, CHENG Chun-tian, ZENG Yun, LIN Jian-yi

(Institute of Hydropower & Hydroinformatics, Dalian University of Technology, Dalian 116024, China)

**Abstract:** Support vector machine (SVM) is a novel type of learning machine, which has been successfully applied to short-term electricity load forecasting. However, its application indicates that how to confirm the parameters of SVM algorithm directly affects forecasting precision. On the basis of analyzing the parameter performance of SVM for regression estimation, a short-term electricity load forecasting model SCE-UA (shuffled complex evolution-University of Arizona) based on SVM is presented. In the process of establishing the model, the radial basis kernel function is introduced, which simplifies the course of solving non-linear problems, and the SCE-UA algorithm is applied to identifying the parameters of SVM. The model is applied to short-term electricity load forecasting using the actual 96 points daily data from Guizhou power grid. The results show that the proposed model not only overcomes the blindness of identifying SVM parameters, but also increases forecasting precision. It is a feasible and effective short-term electricity load forecasting model.

**Key words:** load forecasting; SVM; SCE-UA; similar day