

基于蚁群支持向量机的短期负荷预测

魏俊¹, 周步祥¹, 林楠², 邢义¹

(1.四川大学电气信息学院, 四川 成都 610065; 2.四川电力职业技术学院, 四川 成都 610072)

摘要: 支持向量机(SVM)是一种在统计学习理论基础之上发展起来的针对小样本数据且具有优良推广性能的机器学习方法。阐述了SVM的基本原理及特性,并采用一种新的适用于连续问题的蚁群优化算法(MG-CACO)对SVM核函数的参数进行了优化。同时介绍了基于MG-CACO算法的支持向量机技术的设计思想和特点。并对一实际电网的短期负荷预测进行了实例研究,其结果验证了基于MG-CACO算法的支持向量机预测方法提高了预测精度,此方法在短期负荷预测中的可行性和有效性。

关键词: 支持向量机; 连续蚁群算法; 参数优化; 短期负荷预测

Short-term load forecasting based on MG-CACO and SVM method

WEI Jun¹, ZHOU Bu-xiang¹, LIN Nan², XING Yi¹

(1.Sichuan University, Chengdu 610065, China; 2.Sichuan Electric Power College, Chengdu 610072, China)

Abstract: Support Vector Machine (SVM) as a machine learning method is based on solid theory foundation of Statistical Learning Theory and focuses on small samples. It has good generalization and has received good applications. The theory and characteristics of SVM are expatiated, a new continuous ant colony optimization algorithm called MG-CACO is used to optimize the kernel parameter of SVM in this paper. The design idea and characteristics of MG-CACO and SVM are introduced. Then the application of SVM to a short-term load prediction of an actual power system is proposed. Forecasting results of calculation examples show the feasibility and effectiveness in the short-term forecasting, and that the MG-CACO and SVM method improves the accuracy of prediction.

Key words: support vector machine; continuous ant colony optimization algorithm; parameter optimization; short-term load forecasting

中图分类号: TM715 文献标识码: A 文章编号: 1674-3415(2009)04-0036-05

0 引言

对电力系统负荷预测的研究和应用开发工作已有六十多年的历史,至今还有许多负荷预测方法。但是由于影响负荷的因素非线性程度极高,所以精确的电力负荷预测仍然存在相当的困难。实践证明没有一个方法是绝对准确的,也没有一个方法可以适用于一切电力系统。某一个特殊的方法只适用于某一个特定场合。

传统的负荷预测方法有外推法、回归分析法、相关分析法、时间序列法,卡尔曼(Kalman)滤波法和灰色系统法等等^[1,2]。80年代中后期,专家系统在负荷预测中取得一定的成果。进入90年代以后,人工神经网络 ANN (Artificial NeuNetwork)等人工智能方法在电力系统负荷预测的研究中日益增多^[3,4]。在统计学理论上,由 Vapnik 等人于 20

世纪 90 年代提出了一种新的通用学习方法——支持向量机 SVM (Support Vector Machine),它表现出了许多优于已有方法的性能,正广泛应用于函数回归、模式识别等问题中^[5-7]。和传统的神经网络方法不同,SMV 方法应用的是结构风险最小化(SRM)原则,寻求的是一般误差上界的最小化,而不是单纯训练误差的最小化。SVM 具有模型的自由选择(参数、基函数的位置等)、可将训练转化为二次规划问题(有全局最优解)以及对于小样本良好的学习能力等特性,正成为继神经网络之后的新的研究热点。但是, SVM 在具体应用中存在一个突出问题,即如何设置影响算法性能的一些关键参数,如平衡参数 C 、不敏感性参数 ϵ 及核函数参数等,才能获得较好的预测效果。本文提出了蚁群支持向量机方法,用于电力系统短期负荷预测。利用蚁群算法对其进行参数优化,以减少参数选择的盲目性,提高 SVM 的

预测精度。

1 支持向量机的基本原理及特性分析

1.1 支持向量回归的基本原理

在解决函数回归问题时, SVM 方法的基本思想是: 通过事先定义的非线性映射 $\phi: R_n \rightarrow R_m(m \geq n)$, 把输入空间的数据 x 映射到一个高维特征空间, 然后在该空间中做线性回归 $\phi: R_n \rightarrow R_m(m \geq n)$ 。给定数据点集为 $\{(x_i, y_i), i=1, 2, \dots, N\}$, 其中: $x_i \in R_n$ 为输入向量; $y_i \in R_n$ 为与 x_i 相对应的输出向量; N 是数据点总数。SVM 通过下式进行函数回归估计:

$$f(x) = w \bullet \phi(x) + b \quad (1)$$

式中: w 、 $\phi(x)$ 为 m 维向量; “ \bullet ” 表示特征空间中的点积; $b \in R$ 为阈值。SVM 采用最小化结构 R_{str} 来确定参数 w 和 b , 即

$$\min R_{\text{str}} = \frac{1}{2} \|w\|^2 + CR_{\text{emp}} \quad (2)$$

其中

$$R_{\text{emp}} = \frac{1}{l} \sum_{i=1}^l L_{\varepsilon}[x_i, y_i - f(x_i)] \quad (3)$$

$$L_{\varepsilon}(x, y - f(x)) = \max\{0, |y - f(x)| - \varepsilon\} \quad (4)$$

式(2)等号右端第1项 $\frac{1}{2} \|w\|^2$ 为决策函数复杂性的表达能力项, 是正规化项; 第2项经验风险 R_{emp} 为训练误差项; 平衡因子 C ($C \geq 0$) 为权重系数。经验风险 R_{emp} 由惩罚函数(loss function)来度量, 通常采用 ε -不敏感损失函数 $L_{\varepsilon}[x, y - f(x)]$ 。

引入松弛变量 ξ_i 和 ξ_i^* ($i=1, 2, \dots, l$), 式(2)可改写为:

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 + c \sum_{i=1}^l (\xi_i + \xi_i^*) \\ \text{s.t.} \quad & [w \bullet \phi(x_i)] + b - y_i \leq \varepsilon + \xi_i \\ & y_i - [w \bullet \phi(x_i)] - b \leq \varepsilon + \xi_i^* \\ & \xi_i, \xi_i^* \geq 0 \quad (i=1, 2, \dots, l) \end{aligned} \quad (5)$$

称上式为 SVM 的原始问题, 可以看出原始问题是一个有线性约束的二次规划问题。根据强对偶定理, 引入 Lagrange 乘子 α_i 和 α_i^* ($i=1, 2, \dots, l$), 建立 Lagrange 函数, 并对 w 、 b 、 ξ_i 和 ξ_i^* 求偏导置零, 得到原始问题的对偶问题为:

$$\begin{aligned} \min \quad & \frac{1}{2} \sum_{i,j=1}^l (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) K(x_i, x_j) + \varepsilon \sum_{i=1}^l (\alpha_i^* + \alpha_i) - \sum_{i=1}^l y_i (\alpha_i^* - \alpha_i) \\ \text{s.t.} \quad & \sum_{i=1}^l (\alpha_i^* - \alpha_i) = 0 \\ & 0 \leq \alpha_i, \alpha_i^* \leq C \quad (i=1, 2, \dots, l) \end{aligned} \quad (6)$$

式中: $K(x_i, x_j) = [\phi(x_i) \bullet \phi(x_j)]$ 为核函数, 其作用是不必知道从低维输入空间到高维特征空间非线性映射 $\phi(x)$ 的具体形式, 通过引入核函数就可得到决策回归方程。最常用最优解的核函数是高斯函数

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right), \text{ 其中 } \sigma \text{ 是高斯函数的宽度}$$

函数。设对偶问题的解为 $\bar{\alpha} = (\bar{\alpha}_1, \bar{\alpha}_1^*, \dots, \bar{\alpha}_l, \bar{\alpha}_l^*)^T$, 根据 SVM 的稀疏性, 一对 $\bar{\alpha}_i$ 和 $\bar{\alpha}_i^*$ 最多只有一个不为零, 即 $\bar{\alpha}_i \bar{\alpha}_i^* = 0, \forall i \in (1, 2, \dots, l)$ 且只有少数 $\bar{\alpha}_i$ 、 $\bar{\alpha}_i^*$ 可不为零, 这些不为零的参数所对应的输入向量 x_i 称为支持向量, 则决策方程只由支持向量决定, 与非支持向量无关。决策回归方程式(1)可改写为:

$$f(x) = \sum_{i=1}^l (\bar{\alpha}_i^* - \bar{\alpha}_i) K(x_i, x) + \bar{b} \quad (7)$$

其中

$$\bar{b} = \begin{cases} y_j - \sum_{i=1}^l (\bar{\alpha}_i^* - \bar{\alpha}_i) K(x_i, x_j) + \varepsilon & \text{若 } \bar{\alpha}_j \in (0, c) \\ y_j - \sum_{i=1}^l (\bar{\alpha}_i^* - \bar{\alpha}_i) K(x_i, x_j) + \varepsilon & \text{若 } \bar{\alpha}_j^* \in (0, c) \end{cases} \quad (8)$$

1.2 支持向量机方法的参数特性分析

SVM 的推广性能好坏, 相当程度上取决于参数的选择。参数 c 、 ε 和核函数参数 (高斯核为 σ^2) 的选择, 对于支持向量机回归估计的精度影响很大。误差惩罚参数 c 的作用是在确定的数据子空间中调节学习机器置信范围和经验风险的比例以使学习机器的推广能力最好, 不同数据子空间中最优的 c 不同。在确定的数据子空间中, c 的取值小表示对经验误差的惩罚小, 学习机器的复杂度小而经验风险值较大; 反之亦然。前者称为“欠学习”现象, 而后者则为“过学习”。 ε 不敏感损失函数的大小决定了支持向量的个数。 ε 越大, 支持向量的个数越少, 函数估计的精度越低; ε 越小, 则支持向量的个数越多, 函数估计的精度越高。但是 ε 也不是越小越好, 因为虽然精度提高了, 但是算法所需要的时间也变长了, 所以, 必须选择适当的 ε , 以保证 SVM 的速度和精度。核函数的参数反映了训练数据的特性, 对于系统的泛化能力影响较大。因此, 只有正确选择参数, 才可以使支持向量机回归估计得到很好的拟合效果^[8~10]。

如何合理选择支持向量机的参数, 目前尚未有有效的方法。一般通过交叉验证试算或梯度下降法求解, 这些方法或者人为因素太多, 或者要求函数连续可导, 易陷入局部极小。为此本文选用蚁群算法

对支持向量机的优化，它是一种全局搜索寻优的方法，经过实例检验，本方法具有很好的效果。

2 蚁群优化算法

蚁群算法最早是由意大利学者 Dorigo 提出的，它是一种集体智能算法^[11,12]。该算法虽然提出的时间较晚，但是发展很快，目前在求解 TSP、车间作业调度、网络路由(Qos)、背包问题等方面都得到了广泛应用，且仿真结果显示，蚁群算法具有较好的效果。本文拟从生物学行为出发，运用蚁群成群募集和海量募集两种机制，并结合厌食现象，构建一种适用于连续问题的蚁群优化系统(MG-CACO)，它无需计算梯度等信息，且有较高的全局寻优效率^[13-15]。

2.1 目标函数的选取

一般 MG-CACO 法将优化问题表述为

$$\begin{cases} \min f(x) = f(x_1, x_2, \dots, x_m) \\ \text{s.t. } x_i^l \leq x_i \leq x_i^u; i = 1, 2, \dots, m \end{cases} \quad (9)$$

分量都满足式(9)约束条件的 $x = [x_1, x_2, \dots, x_m]^T$ ，称为可行解或个体， $f(x)$ 越小，表示 x 处的食物源越优；反之亦然。第 i 只蚂蚁在时间 t 时的位置对应于一个可行解，记为 $x_i(t)$ 。一组个体构成种群，其个体数为种群规模 N 。这里我们取 $x_i = [c, \varepsilon, \sigma^2]$ ，目标函数为：

$$\min F = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (10)$$

$$\text{约束条件为} \begin{cases} c_{\text{low}} \leq c \leq c_{\text{up}} \\ \varepsilon_{\text{low}} \leq \varepsilon \leq \varepsilon_{\text{up}} \\ \sigma_{\text{low}}^2 \leq \sigma^2 \leq \sigma_{\text{up}}^2 \end{cases} \quad (11)$$

也就是要让训练样本集的总误差要最小。这里 y_i 为实际值， \hat{y}_i 为通过 SVM 计算出的预测值。寻参过程就是最小化 F 。

2.2 蚁群搜索步骤

Step1 根据目标函数的优劣，找出当前最优位置的蚂蚁，即当前最优解，称这只蚂蚁为头蚁。

Step2 除头蚁外，选择最差的一半蚂蚁进行成群募集操作，剩余的一半执行海量募集操作。其中成群募集操作的作用是让头蚁引领较差的个体向当前最优的位置移动，而海量募集操作则是蚂蚁根据各自信息素大小选择其移动的位置。这两部分操作的具体描述可以参见文献[11]。

Step3 头蚁进行局部搜索，搜索策略如下：

$$\begin{aligned} x_{\text{temp}}(t) &= x_{\text{leader}}(t) + a * \text{step} \quad p < 0.5 \\ x_{\text{temp}}(t) &= x_{\text{leader}}(t) - a * \text{step} \quad p \leq 0.5 \end{aligned} \quad (12)$$

其中： p 是(0,1)上的随机数， a 随着迭代次数的增加

而减少。如果 $x_{\text{temp}}(t)$ 优于 $x_{\text{leader}}(t)$ ，则用此解替换，同时头蚁的信息素也要进行相应的更新，更新的式子可以表示成：

$$\tau(x_{\text{leader}}(t)) = \tau(x_{\text{leader}}(t)) + k[f(x_{\text{leader}}(t)) - f(x_{\text{temp}}(t))] \quad (13)$$

这一步与 MG-CACO 方法不同。MG-CACO 方法对每只蚂蚁都进行局部搜索，这样做会导致多次训练 SVM，降低寻优速度，因此文中在局部搜索中只对头蚁进行局部搜索。

Step4 所有蚂蚁完成一轮搜索后要执行信息素挥发操作，即

$$\tau(x, t) = \rho * \tau(x, t) \quad (14)$$

ρ 为挥发因子。其算法流程如下图 1 所示。

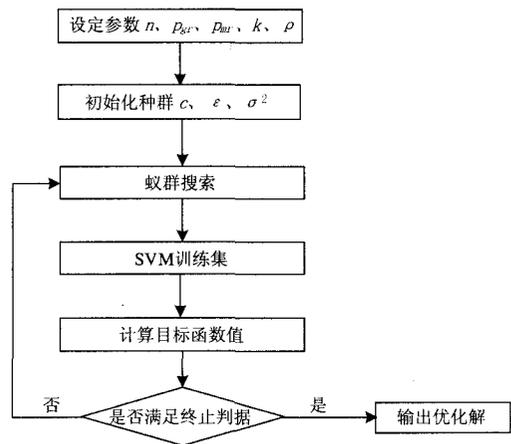


图 1 基于蚁群算法的 SVM 参数优化的一般过程

Fig.1 Process of optimizing SVM parameters based on ant colony algorithm

3 基于 MG-CACO 的 SVM 短期负荷预测

3.1 基于 MG-CACO 的 SVM 回归估计模型

为降低问题求解规模，对一天 24 点的每一点分别建立其负荷预测模型，分别预测对应的整点负荷。蚁群支持向量机的输入为：预测日之前 60 天根据相似日法选出的与预测时刻相似点的负荷数据；相似点的日最高气温 T_{max} 、最低气温 T_{min} ；相似点的周属性 $W=(1,2,\dots,7)$ ，对应于星期一到星期日；预测日的节日属性 $F=(1, 0)$ ，表征预测日是否为重大节假日，1 表示该预测日为重大节假日。输出为预测日第 i 小时负荷。具体的负荷预测步骤如下：

(1) 对历史负荷数据进行平滑处理和基于相似日法选择样本，再将这些样本归一化处理，温度数据量化处理，形成训练样本集和验证样本集；

(2) 用训练样本集建立如式(10)的目标函数；

(3) 利用蚁群算法求解式(10), 得到优化参数 c 、 ε 、 σ^2 ;

(4) 将 c 、 ε 和 σ^2 代入式(6), 并求解 α_i 、 α_i^* ;

(5) 将 α_i 、 α_i^* 代入式(7)、(8), 得到回归函数表达式, 对未来某一时刻的负荷进行预测。

3.2 算例分析

为了验证本文所提蚁群支持向量机方法的可行性和有效性, 对四川省某市进行了短期负荷预测的计算和分析, 采用 2007 年 5 月-6 月的负荷和气象数据、节假日信息构成预测模型, 预测 6 月 28 日的 24 点负荷值。

在本文算例中, c 的取值范围为(0.001, 200), 初值为 1; ε 的取值范围为 (0, 0.8), 初值为 0.1; σ^2 的取值范围为(0.1, 200), 初值为 60; 蚂蚁种群大小取 $N=50$, 最大迭代为 500 次, 成群募集 $p_{GR}=0.5$, 海量募集 $p_{MR}=0.5$, 信息素更新中比例因子 $k=0.01$, 挥发因子 $\rho=0.8$ 。用 Matlab 和 libsvm 编制仿真程序。为了校验本方法的有效性, 又选用使用交叉验证试算的传统 SVM 法和基于梯度下降法的传统 BP 神经网络法做短期负荷预测。

文中取日平均相对误差 (MAPE) 和日均方根误差 (RMS) 作为各种方法预测效果判断的依据, 即:

$$MAPE(t) = \frac{1}{24} \sum_{i=1}^{24} \left| \frac{Y'(t) - Y(t)}{Y(t)} \right| \times 100\% \quad (15)$$

$$RMS(t) = \sqrt{\frac{1}{24} \sum_{i=1}^{24} \left(\frac{Y'(t) - Y(t)}{Y(t)} \right)^2} \quad (16)$$

式中: $Y'(t)$ 和 $Y(t)$ 分别指 t 时刻预测负荷值和实际值。

图 2 和图 3 为 2007 年 6 月 28 日的预测结果对比图, 蚁群支持向量机法的预测曲线明显比其它两种方法更接近于实际负荷曲线, 同时它的相对误差曲线也更小。

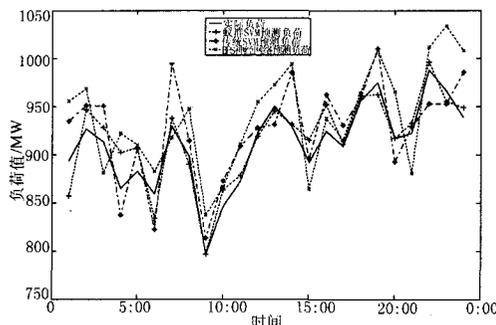


图 2 实际负荷与预测负荷曲线比较

Fig.2 Comparison between the actual load curve and the load forecasting curve

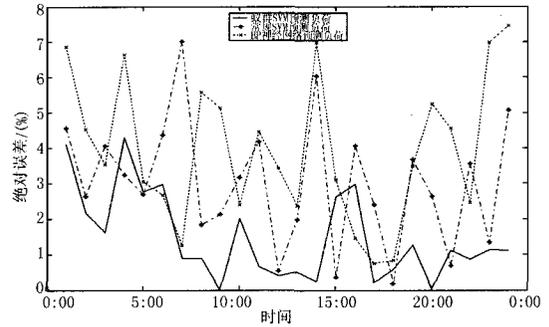


图 3 负荷预测的绝对误差比较

Fig.3 Comparison of the forecasting absolute errors

表 1 误差指标比较

Tab.1 comparison of error indexes

方法	蚁群 SVM	传统 SVM	传统 BP 神经网络
MAPE/(%)	1.480	3.180	3.967
RMS/(%)	1.915	3.475	4.449

表 1 为 2007 年 6 月 28 日误差指标比较结果, 从以上结果可以看出, 蚁群 SVM 预测模型比传统 SVM 和传统神经网络模型效果明显要好, 而传统 SVM 模型效果略好于传统 BP 神经网络。这因为 SVM 方法具有全局最优性, 不会陷入局部最小点, 避免了传统神经网络方法的缺陷, 提高了预测精度。而蚁群 SVM 方法由于对影响 SVM 方法计算精度的优化参数 c 、 ε 和 σ^2 基于蚁群算法的群智全局寻优, 避免传统 SVM 的缺陷, 从而明显提高预测精度。

4 结论

本文在对支持向量机方法应用于回归模型的短期负荷预测的基础上, 结合一种适应于连续问题的蚁群优化算法对其进行参数优化, 该方法通过募集机制并结合厌食现象采取全局寻优, 有效地解决了局部收敛问题, 提高了预测精度。通过一地区电网的实际算例, 并与传统 SVM 和传统 BP 神经网络预测法比较验证了该方法有较高的预测精度和较好的适应性。

参考文献

- [1] 康重庆, 夏清, 刘梅. 电力系统负荷预测[M]. 北京: 中国电力出版社, 2007.
KANG Chong-qing, XIA Qing, LIU Mei. Power System Load Forecasting[M]. Beijing: China Electric Power Press, 2007.
- [2] 康重庆, 夏清, 张伯明. 电力系统负荷预测研究综述与发展方向的探讨[J]. 电力系统自动化, 2004, 28(17): 1-9.
KANG Chong-qing, XIA Qing, ZHANG Bo-ming. Review of Power System Load Forecasting and Its

- Development[J]. Automation of Electric Power Systems, 2004, 28(17): 1-9.
- [3] 胡守仁, 余少波, 戴葵. 神经网络导论[M]. 长沙: 国防科技大学出版, 1993.
HU Shou-ren, YU Shao-bo, DAI Kui. Introduction to Neural Network[M]. Changsha: National Defense Scientific and Technical University Press, 1993.
- [4] 荣辉. 基于前馈神经网络的电力系统负荷预测研究[D]. 北京: 清华大学, 1998.
RONG Hui. A Load Forecasting System Based on BP Artificial Neural Network[D]. Beijing: Tsinghua University, 1998.
- [5] Vapnik V N. The Nature of Statistical Learning Theory, 2nd ed[M]. New York: Springer, 2000.
- [6] 瓦普尼克. 统计学习理论[M]. 张学工, 译. 北京: 电子工业出版社, 2004.
Vapnik V N. Statistical Learning Theory[M]. ZHANG Xue-gong, Trans. Beijing: Publishing House of Electronics Industry, 2004.
- [7] Vapnik V N. The Nature of Statistical Learning Theory[M]. New York: Springer, 1995.72-236.
- [8] 李元诚, 方廷建, 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.
LI Yuan-cheng, FANG Ting-jian, YU Er-keng. Study of Support Vector Machines for Short-Term Load Forecasting[J]. Proceedings of the CSEE, 2003, 23(6): 55-59.
- [9] 杨镜非, 程浩忠. SVM在电网短期负荷预测中应用研究[J]. 电力自动化设备, 2004, 24(2): 30-32.
YANG Jing-fei, CHENG Hao-zhong. Application of SVM to Power System Short-term Load Forecast[J]. Electric Power Automation Equipment, 2004, 24(2): 30-32.
- [10] 牛东晓, 刘达, 陈广娟, 等. 基于遗传优化的支持向量机小时负荷滚动预测[J]. 电工技术学报, 2007, 22(6): 148-153.
NIU Dong-xiao, LIU Da, CHEN Guang-juan, et al. Support Vector Machine Models Optimized by Genetic Algorithm for Hourly Load Rolling Forecasting[J]. Transactions of China Electrotechnical Society, 2007, 22(6): 148-153.
- [11] Kindt V T, Monmarch N, Tercinet F, et al. An Ant Colony Optimization Algorithm to Solve a 2-machine Bicriteria Flowshop Scheduling Problem [J]. European Journal of Operational Research, 2002, 142(2): 250-257.
- [12] Dorigo M, Maniezzo V, Colony A. Ant System: Optimazation by a Colony of Cooperation Agents[J]. IEEE Trans on SMC Part B, 1996, 26(1): 29-41.
- [13] Joachims T. Estimating the Generalization Performance of an SVM Efficiently[A]. Proceedings of the 17th International Conference on Machine Learning[C]. San Francisco: 2000. 431-438.
- [14] 贺益君, 俞欢军, 陈德钊. 基于募集机制的连续蚁群系统及其应用[J]. 浙江大学学报(工学版), 2006, 40(5): 748-752.
HE Yi-jun, YU Huan-jun, CHEN De-zhao. Recruitment Mechanism Based Continuous Ant Colony Optimization System and Its Application[J]. Journal of Zhejiang University(Engineering Science), 2006, 40(5): 748-752.
- [15] 孙学勤, 刘丽, 付萍, 等. 一种连续空间优化问题的蚁群算法及应用[J]. 计算机工程与应用, 2005, 41(34): 217-220.
SUN Xue-qin, LIU Li, FU Ping, et al. Ant Colony Algorithm in Continuous Space[J]. Computer Engineering and Applications, 2005, 41(34): 217-220.

收稿日期: 2008-04-18; 修回日期: 2008-06-19

作者简介:

魏俊(1978-), 男, 硕士研究生, 研究方向为电力系统自动化及其计算机应用; E-mail: stevenweijun@163.com

周步祥(1965-), 男, 博士, 教授, 从事电力系统自动化及其计算机应用的研究和教学工作;

林楠(1973-)女, 硕士, 讲师, 从事电力系统自动化及其计算机应用的研究和教学工作。

国网信通公司完成跨区电网应急指挥中心建设任务

近日, 国网信通公司直属公司国电通公司顺利完成国网建设运行部跨区电网应急指挥中心建设项目, 并通过工程验收。

应急指挥中心项目建设主要由远程联网与本地会议室建设组成。远程联网通过敷设电缆、光缆, 实现本地会议室与楼内通信机房、信息机房和国调自动化机房的信息互通, 满足各应用系统的接入需求。本地会议室配置了视频会议终端设备、集中控制设备、发言扩声设备等, 可以组织召开视频会议。国电通公司在承接项目过程中, 认真研究用户要求, 详细制订工程方案, 通过周密安排、精心施工, 保质保量地完成了中心建设任务。

应急指挥中心的建成将满足国网建设运行部远程视频会商、应急指挥的需要, 同时可以实时监视运行场站与在建场站的工作情况。