

基于主成分分析与人工神经网络的风电功率预测

周松林, 茆美琴, 苏建徽

(教育部光伏系统工程研究中心(合肥工业大学), 安徽省 合肥市 230009)

Prediction of Wind Power Based on Principal Component Analysis and Artificial Neural Network

ZHOU Songlin, MAO Meiqin, SU Jianhui

(Research Center for Photovoltaic System Engineering (Hefei University of Technology),
Ministry of Education, Hefei 230009, Anhui Province, China)

ABSTRACT: A wind power prediction model based on integration of principal component analysis (PCA) with back-propagation (BP) neural network is proposed. The PCA is used to preprocess original multi-dimensional input variables and principal components of input variables are chosen as the input of BP neural network, by this way either the dimensions of input variables can be reduced or correlativity among input variables can be eliminated, thus both convergence and stability of neural network can be improved. Simulation results show that the accuracy of wind power prediction by the proposed PCA-BP model is better than that by common neural network models and the proposed model possesses better generalization performance.

KEY WORDS: wind power prediction; principal component analysis; back-propagation (BP) neural network; generalization performance

摘要:提出了主成分分析与前馈神经网络相结合的风电功率预测模型。采用主成分分析法对原始多维输入变量进行预处理,选择输入变量的主成分作为神经网络的输入,既减少了输入变量的维数,又消除了各输入变量的相关性,从而简化了网络的结构,提高了网络收敛性和稳定性。仿真结果表明,相对于一般神经网络模型,基于主成分分析的神经网络模型预测精度更高、泛化性能更好。

关键词:风电功率预测;主成分分析;前馈神经网络;泛化性能

0 引言

随着我国能源结构的调整,风能的开发利用已得到高度重视,从技术成熟性和经济可行性看,风电在可再生能源中具有很好的前景^[1]。随着风能的

加速发展,越来越多的大型风电场纳入统调电网,风电在电网的比重越来越大,风电的强随机性增大了电力调度的难度。风速及风电功率的准确预测可以减少电力系统运行成本,有利于调度部门及时调整计划,从而减轻风电对电网的不利影响。

风速及风电功率的预测精度取决于预测算法、预测对象本身的特性、数据预处理技术等。目前,风速预测常用方法有时间序列法^[2]、卡尔曼滤波法^[3]、神经网络法^[4-6]、小波-神经网络法^[7-8]、支持向量机^[9-10]、混沌预测法^[11-12]、组合预测法^[13]。预测对象本身的特性对预测精度有较大的影响,预测对象变化越快,精度越低。数据预处理方法在很大程度上影响预测效果,常用的数据预处理方法有归一化、小波分解、主成分分析、因子分析^[14]等方法。

主成分分析法(principal component analysis, PCA)是一种数据压缩和特征提取的多变量统计分析技术,它能够将多个相关变量转化为少数几个不相关的综合变量,且这些不相关的综合变量包含了原变量提供的大部分信息^[15]。因此,本文通过相关性分析选择预测因子,采用主成分分析法,对预测因子数据进行特征提取,获得几个综合指标,使得新的输入变量维数低、分量间相关性小。在此基础上,再应用BP神经网络建立风电功率预测模型,较好地风电功率进行了预测。

1 主成分分析原理

主成分分析是一种多元统计分析方法。该方法通过构造原变量的一系列线性组合形成新变量,使这些新变量在彼此互不相关的前提下尽可能多地反映原变量的信息。数据信息主要反映在数据变量

基金项目: 国家重点基础研究发展计划项目(973 项目)(2009CB219708)。

The National Basic Research Program of China (973 Program)(2009CB219708)。

的方差上，方差越大，包含信息越多。通常用累计方差贡献率来衡量。主成分分析是对多个样本的输入变量形成的数据矩阵求取相关矩阵，根据相关矩阵的特征值，获得累计方差贡献率，再根据相关矩阵的特征向量，确定主成分。具体步骤如下：

1) 原始数据标准化。为消除由于原变量的量纲不同、数值差异过大带来的影响，对原变量作标准化处理。假设有 m 个指标 X_1, X_2, \dots, X_m 分别表示每个对象的各个特性，如果有 N 个对象，可以用 $N \times m$ 矩阵表示，即

$$\mathbf{X}_{N \times m} = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & & \vdots \\ x_{N1} & \cdots & x_{Nm} \end{bmatrix} \quad (1)$$

首先进行中心标准化处理生成标准矩阵 \mathbf{Y} ，即

$$x_{ij}^* = (x_{ij} - \bar{x}_j) / s_j \quad (2)$$

式中， $i=1,2,\dots,N$ ； $j=1,2,\dots,m$ ； \bar{x}_j ， s_j 分别为指标变量 x_j 的均值和方差。

2) 建立相关矩阵 \mathbf{R} ，并计算其特征值和特征向量，即

$$\mathbf{R} = \mathbf{X}^{*T} \mathbf{X}^* / (N - 1) \quad (3)$$

式中 \mathbf{X}^* 为标准化后的数据矩阵。求得自相关矩阵 \mathbf{R} 的特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ 及相应的特征向量 u_1, u_2, \dots, u_m 。

3) 确定主成分个数。方差贡献率和累计方差贡献率分别为

$$\eta_i = 100\% \lambda_i / \sum_i \lambda_i \quad (4)$$

$$\eta_\Sigma(p) = \sum_i^p \eta_i \quad (5)$$

选取主成分的个数取决于累计方差贡献率，通常累计方差贡献率大于 75%~95% 时对应的前 p 个主成分便包含 m 个原始变量所能提供的绝大部分信息，主成分个数就是 p 个。

4) p 个主成分对应的特征向量为 $\mathbf{U}_{m \times p} = [u_1, u_2, \dots, u_p]$ ，则 n 个样本的 p 个主成分构成的矩阵为

$$\mathbf{Z}_{N \times p} = \mathbf{X}_{N \times m}^* \mathbf{U}_{m \times p} \quad (6)$$

2 风电功率预测因子的选择及主成分提取

2.1 预测因子的选择

本文算例数据来源为某风电场提供的额定容量为 900 KW 的机组在 2007 年 6 月份的历史输出功率以及在此期间内的风速、风向、温度、湿度和空气密度等气象数值。预测模型如图 1 所示。

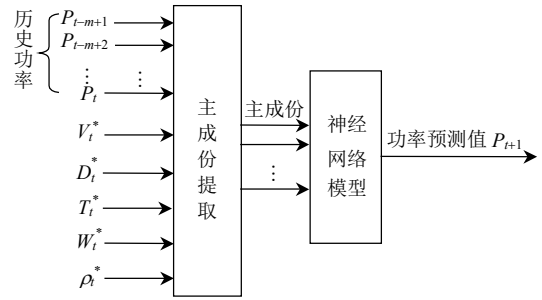


图 1 风电功率预测模型

Fig. 1 The predict model of wind power

功率预测有物理方法和统计方法 2 种^[16]，其中统计方法又有 2 种方式：一种是基于历史功率影响因素的直接预测；另一种是基于风速预测值和实测功率曲线的间接预测，即先预测风速和风向，再结合功率曲线计算风电机组发电功率。本文采用直接预测方法。

图 1 中原始预测因子包括：预测点之前的 m 个历史功率和预测点前一刻的风速、风向、温度、湿度和空气密度；输出是 $t+1$ 时刻的风电机组的发电功率。时间序列能够预测的原因之一在于其相邻数值间的自相关性。本文通过计算功率时间序列自相关系数来确定 m 。本算例功率时间序列的自相关系数曲线如图 2 所示。可见，当自相关系数降至 $1 - e^{-1}$ 时，其对应的延时步长 5 可作为 m 值。因此，该功率预测模型的预测因子(样本输入维数)是 10 维。

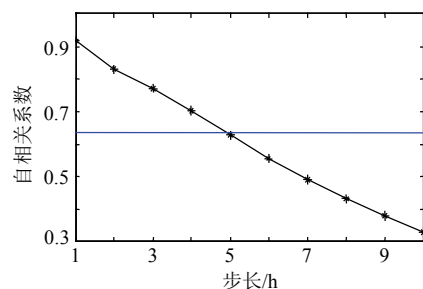


图 2 功率时间序列自相关系数

Fig. 2 The autocorrelation coefficient of power time series

由于预测因子之间存在一定的相关性，若将它们同时作为预测模型的输入会造成信息重叠，增加模型训练复杂度，降低泛化性能。对这些因素进行主成分提取，可以得到少数几个包含原始变量绝大部分信息并且互不相关的综合变量，最后以综合变量作为前馈神经网络输入，对未来输出功率进行一步预测。

2.2 主成分提取

计算标准化后的输入因子的相关系数矩阵 \mathbf{R} 、 \mathbf{R} 的特征根、相应的标准特征向量以及贡献率，并

依据累积贡献率提取主成分。计算结果如表 1，图 3 是方差相对于成分个数的散点图。

表 1 方差及主成分贡献率
Tab. 1 Eigenvalue and principle component contribution rates

成分	特性向量		
	特征值	方差贡献率/%	方差累计贡献率/%
1	5.191	79.800 8	79.800 8
2	0.700	10.765 2	90.566 0
3	0.288	4.428 8	94.994 8
4	0.112	1.734 0	96.728 8
5	0.086	1.328 2	98.05 69
6	0.051	0.796 3	98.853 2
7	0.034	0.524 2	99.377 4
8	0.028	0.433 5	99.810 9
9	0.006	0.104 5	99.915 5
10	0.006	0.084 5	100.000 0

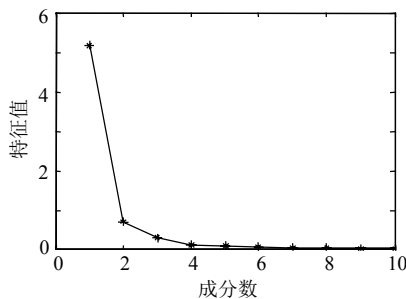


图 3 方差相对于成分个数的散点图
Fig. 3 The scatter of variance relative to the number of component

主成分(principal component, PC)的选择方法是先定义一个必须由 PC 集合达到的累计百分比方差的上限，这个上限取决于实际处理的问题，可以在 75%~95% 的范围变化^[17]。另一种方法是检查方差相对于成分个数的散点图，并选择图形近似水平的点。第 3 种方法是交叉验证，选取使验证误差最小的 PC 个数。本文综合运用这 3 种方法来确定 PC 个数。先观察表 1 和图 3，按照前 2 种方法可以确定 PC 个数为 3 或 4 个；再利用第 3 种方法进行交叉验证，发现当取 3 个 PC 得到的验证误差小于 4 个 PC 的情况。进一步对提取的前 3 项主成分进行相关性分析，发现 3 项主成分之间的相关性显著减小，因此可以用这 3 项主成分的特征值对应的特征向量，即经过变换后的综合指标，标准化后作为神经网络预测模型的输入。

下文采用本文算例进一步分析 PCA 技术对前馈神经网络性能的影响。图 4 给出了基于 PCA 和非 PCA 的神经网络性能比较曲线。可以看出，基于 PCA 的方法和非 PCA 方法有显著不同，非 PCA 方法训练误差持续减小，但验证误差初始减小，然后在 400 次迭代后开始增加，这表明网络在这一点开始过拟

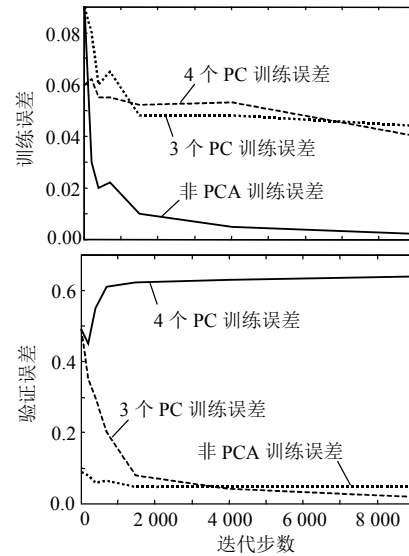


图 4 PCA 对神经网络性能的影响

合；相反，基于 PCA 的方法几乎消除了过拟合，训练误差和验证误差都减小并稳定到一个常值，其中，3 个 PC 的网络对应的 2 种误差下降更明显，最终稳定值也小于 4 个 PC 的网络。这说明采用 PC 作为网络输入，消除了输入因子多重相关性的影响，而正是这种多重相关性导致网络结构复杂，容易出现过拟合，从而产生大的预报方差。

针对单隐层神经元数目对 PCA 神经网络和非 PCA 神经网络的影响，本文采用上述 2 种模型对分别具有 2—8 个隐层神经元的前馈网络的预测误差进行统计，结果如图 5 所示。显然，非 PCA 网络的预测误差随着神经元从 2 增加到 6 而减小，之后又增加，这意味着 6 个神经元是最佳的，而用 PC 作为输入的网络，4 个隐层神经元就达到最佳，且随着隐层神经元数目的增加，预测误差增加较平缓，最终很快稳定。这表明基于 PCA 的网络具有较强的恢复能力。

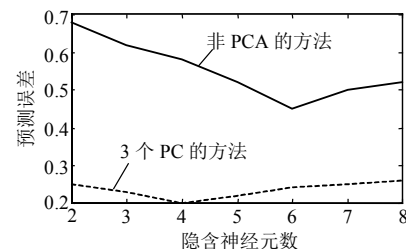


图 5 隐层神经元对神经网络性能的影响

Fig. 5 The effect of hidden layer neurons on neural network performance

3 基于 PCA-BP 方法的风电功率预测

分别用原始输入样本和经主成分分析后得到的新样本作为 BP 神经网络的输入，对风力发电功率建

立预测模型，这两种预测模型分别称为前馈网络和 PCA-BP 网络，因为 3 层网络理论上可以逼近任何非线性函数，所以这 2 个网络都采用单隐层结构。通过前文的分析确定前馈网络结构为 10-6-1，PCA-BP 网络结构为 3-4-1。2 个模型均采用用于网络训练的 300 个训练样本和用于验证预测效果的 80 个测试样本。2 种方法均做 10 次试验。图 6、7 是平均绝对误差最小时的预测曲线及绝对值误差曲线。

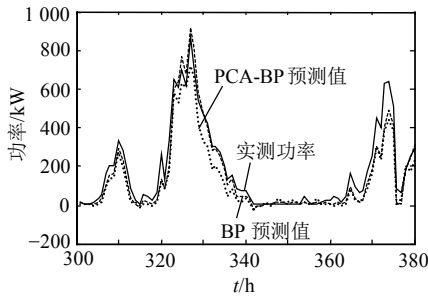


图 6 平均绝对误差最小时的功率预测曲线
Fig. 6 The prediction curves of power when the average absolute error is minimum

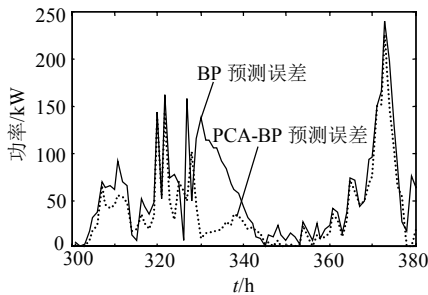


图 7 平均绝对误差最小时的功率绝对值误差曲线
Fig. 7 The absolute error curves of power when the average absolute error is minimum

从图 6、7 可看出，PCA-BP 模型相对于 BP 模型，预测曲线更接近于真实曲线，其绝对值误差较小，特别是在功率变化幅度大的位置，其优越性更为明显。主要原因为以经主成分提取后的新的综合变量代替原始多元变量作为 BP 网络的输入有 3 个优点：新变量能涵盖原始变量绝大部分信息且互不相关，避免了因原始变量的互相关造成的信息冗余；通过主成分提取能降低输入变量的维数，简化网络结构，减少训练的复杂程度和训练时间；前馈神经网络对低维、相互独立的综合输入变量有更强的模式识别能力，有助于网络训练和预测。

采用 PCA-BP 和前馈网络对上述算例进行 1—24 h 的多步预测，即将前 1 h 的预测值当作真实值加入到网络输入端，同时去除离当前预测点最远的历史数据，更新网络，重新训练，从而预测下一时刻的功率。多步预测均方根误差曲线如图 8 所示。

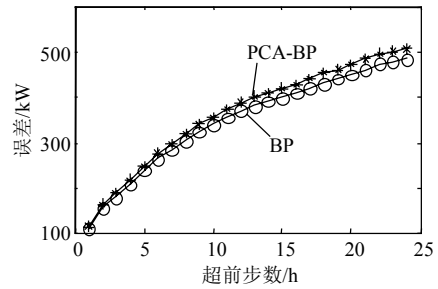


图 8 多步预测误差
Fig. 8 The error of multi-step forecasting

由图 8 可知：随着预测步长的增加，PCA-BP 网络和 BP 网络的预测误差越大；PCA-BP 网络预测误差小于前馈网络预测误差，超前步数越多，PCA-BP 网络的优势越明显。

4 模型比较

为进一步验证本文方法在风电场功率预测中体现出的性能，将 PCA-BP 网络与支持向量机 (support vector machine, SVM)、径向基函数(radial basis function, RBF)、前馈网络同时应用于 850 MW 风电场的出力预测。根据获得的算例数据确定原始输入变量有 8 个，RBF 网络输入节点数取原始变量个数，采用从样本中选择固定隐层中心的方法，隐层神经元使用相同宽度，输出层权值使用最小二乘法求解；BP 网络输入节点数同样取 8，隐层节点数以均方根误差最小化为原则根据多次实验确定，网络结构最后确定为 8-12-1。PCA-BP 网络结构为 4-8-1；RBF、前馈和 PCA-BP 网络权值均需初始化。为减小初始化带来的输出结果不稳定，对 3 种网络均采用 10 次独立实验，预测结果取平均值。SVM 核函数均选择 RBF，采用网络搜索法确定模型参数^[18-19]。4 种预测方法的实验结果见表 2。表中：MAE 表示平均绝对值误差；RMSE 表示均方根误差。

表 2 4 种网络训练及预测结果
Tab. 2 The train and forecast results based on four network

网络模型	训练误差平均值/MW		预测误差平均值/MW		训练时间/s
	MAE	RMSE	MAE	RMSE	
PCA-BP	13.4	19.5	67.8	97.7	252.9
SVM	10.6	19.7	72.2	102.6	272.1
RBF	12.4	16.6	78.3	113.3	10.4
前馈网络	17.4	24.5	83.9	130.8	292.6

由表 2 可知：由于支持向量机基于结构风险最小化原则，具有优越的泛化性能，而 RBF 和 BP 网络模型是基于经验风险最小化，其泛化性能较差，SVM 模型比 RBF 和 BP 预测精度高，并且发现 PCA-BP 的预测误差比 SVM 小；从训练时间方面看，RBF 网络训练速度最快，BP 网络训练时间最

长, PCA-BP 由于结构简单, 输入变量少, 其耗时较 BP 要少, SVM 的耗时也较长, 所耗时间绝大部分用于参数的网格搜索。通过比较可以看出, 相对于其他方法, 结合了 PCA 技术的 BP 神经网络具有更高的预测精度和更快训练速度。

5 结论

本文利用多元统计分析理论中的主成分提取技术对多维、互相关的原始输入因子进行主成分提取, 将提取后的综合因子作为神经网络的输入对风电机组输出功率进行预测。该方法有效降低了数据相关性, 减少了输入样本维数, 大大简化了网络结构, 降低了网络训练的难度。实验结果表明对风电机组输出功率的预测, PCA-BP 网络较一般 BP 网络具有更高的预测精度、更短的学习时间。

现有的一些预测方法主要集中在确定性预测方面, 但仅仅提供确定性预测结果往往难以满足风电管理的要求^[20], 如何在确定性预测结果的基础上进一步分析风电功率预测的不确定性, 向风电管理人员提供预测结果的概率特性是今后研究的方向。

参考文献

- [1] 杨秀媛, 肖洋, 陈树勇. 风电场风速和发电功率预测研究[J]. 中国机电工程学报, 2005, 25(11): 1-5.
Yang Xiuyuan, Xiao Yang, Chen Shuyong. Wind speed and generated power forecasting in wind farm[J]. Proceedings of the CSEE, 2005, 25(11): 1-5(in Chinese).
- [2] 潘迪夫, 刘辉, 李燕飞. 风电场风速短期多步预测改进算[J]. 中国机电工程学报, 2008, 28(26): 87-91.
Pan Difu, Liu Hui, Li Yanfei. Optimization algorithm of short-term multi-step wind speed forecast[J]. Proceedings of the CSEE, 2008, 28(26): 87-91(in Chinese).
- [3] Bossanyi E A. Short-term wind prediction using Kalman filters[J]. Wind Engineering, 1985, 9(1): 1-8.
- [4] Mohamed A M, Shafiqur R, Talal O H. A neural networks approach for wind speed prediction[J]. Renewable Energy, 1998, 13(3): 345-354.
- [5] 范高锋, 王伟胜, 刘纯. 基于神经网络的风电功率短期预测系统[J]. 电网技术, 2008, 32(22): 82-86.
Fan Gaofeng, Wang Weisheng, Liu Chun. Artificial neural network based wind power short term prediction system[J]. Power System Technology, 2008, 32(22): 82-86(in Chinese).
- [6] 白永祥, 房大中, 侯佑华, 等. 内蒙古电网区域风电功率预测系统[J]. 电网技术, 2010, 34(10): 158-162.
Bai Yongxiang, Fang Dazhong, Hou Youhua, et al. Regional wind power forecasting system for Inner Mongolia power grid[J]. Power System Technology, 2010, 34(10): 158-162(in Chinese).
- [7] 张彦宁, 康龙云, 周世琼, 等. 小波分析应用于风力发电预测控制系统中的风速预测[J]. 太阳能学报, 2008, 29(5): 520-524.
Zhang Yanning, Kang Longyun, Zhou Shiqiong, et al. Wavelet analysis applied to wind speed prediction in predicate control system of wind turbine[J]. Acta Energiac Solaris Sinica, 2008, 29(5): 520-524(in Chinese).
- [8] 王丽婕, 冬雷, 廖晓钟, 等. 基于小波分析的风电场短期发电功率预测[J]. 中国电机工程学报. 2009, 29(28): 30-33.
Wang Lijie, Dong Lei, Liao Xiaozhong, et al. Short-term power prediction of a wind farm based on wavelet analysis[J]. Proceedings of the CSEE, 2009, 29(28): 30-33(in Chinese).
- [9] Pai Pingfeng, Hong Weichiang. Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms[J]. Electric Power Systems Research, 2005, 74(3): 417-425.
- [10] 戚双斌, 王维庆, 张新燕. 基于支持向量机的风速与风功率预测方法研究[J]. 华东电力, 2009, 37(9): 1600-1603.
Qi Shuangbin, Wang Weiqing, Zhang Xinyan. Wind speed and wind power prediction based on SVM[J]. East China Electric Power, 2009, 37(9): 1600-1603(in Chinese).
- [11] 王丽婕, 廖晓钟, 高爽, 等. 并网型大型风电场风力发电功率-时间序列的混沌属性分析[J]. 北京理工大学学报, 2007, 27(12): 1077-1080.
Wang Lijie, Liao Xiaozhong, Gao Shuang, et al. Chaos characteristics analysis of wind generation time series for a grid connecting wind farm[J]. Transactions of Beijing Institute of Technology, 2007, 27(12): 1077-1080(in Chinese).
- [12] 罗海洋, 刘天琪, 李兴源, 等. 风电场短期风速的混沌预测方法[J]. 电网技术, 2009, 33(9): 67-71.
Luo Haiyang, Liu Tianqi, Li Xingyuan, et al. Chaotic forecasting method of short-term wind speed in wind farm[J]. Power System Technology, 2009, 33(9): 67-71(in Chinese).
- [13] 刘纯, 范高锋, 王伟胜, 等. 风电场输出功率的组合预测模型[J]. 电网技术, 2009, 33(13): 74-79.
Liu Chun, Fan Gaofeng, Wang Weisheng, et al. A combination forecasting model for wind farm output power[J]. Power System Technology, 2009, 33(13): 74-79(in Chinese).
- [14] 王学仁, 王松桂. 实用多元统计分析[M]. 上海: 上海科学技术出版社, 1990: 270-344.
- [15] 王华, 王连华, 葛岭梅. 主成分分析与 BP 神经网络在煤耗氧速度预测中的应用[J]. 煤炭学报, 2008, 33(8): 920-925.
Wang Hua, Wang Lianhua, Ge Lingmei. Application of principal component analysis and BP neural network in the rate of coal oxygen consumption prediction[J]. Journal of China Coal Society, 2008, 33(8): 920-925(in Chinese).
- [16] Lexiadis M A, Dokopoulou S P, Samanoglou S H, et al. Short term forecasting of wind speed and related electrical power[J]. Solar Energy, 1998, 63(1): 61-68.
- [17] Sandhya S. 史晓霞译. 神经网络在应用科学和工程中的应用—从基本原理到复杂的模式识别[M]. 北京: 机械工业出版社, 2009: 7.
- [18] 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: 173-192.
- [19] Vapnik V. The nature of statistical learning theory[M]. New York: Springer-Verlag, 1995: 84-93.
- [20] Jeremie J, Lionel F, George K. Probabilistic short-term wind power forecasting based on kernel density estimators[C]//European Wind Energy Conference. Milan, Italy: European Wind Energy Association, 2007: 7-10.



周松林

收稿日期: 2011-04-26.

作者简介:

周松林(1975), 男, 博士研究生, 副教授, 研究方向为分布式发电微网系统, E-mail: zsl040908@163.com;

茆美琴(1961), 女, 教授, 博士, 研究方向为电力电子技术可在再生能源发电系统中的应用。

(责任编辑 杜宁)