

基于主成分分析和遗传优化 BP 神经网络的光伏输出功率短期预测

许童羽, 马艺铭, 曹英丽, 唐瑞, 陈俊杰

(沈阳农业大学信息与电气工程学院, 辽宁 沈阳 110161)

摘要: 针对光伏系统输出功率的波动性和间歇性特点, 提出一种基于主成分分析(PCA)和遗传算法(GA)优化的 BP 神经网络功率短期预测方法。通过历史功率数据和实时气象因素对输出功率进行直接预测, 利用主成分分析法将多个原始变量降维成少数彼此独立的变量, 作为神经网络的输入。同时利用遗传算法的全局搜索特性在解空间中定位一个较好的空间, 优化 BP 的初始权值阈值, 克服了传统 BP 神经网络易陷入局部极小点、学习收敛速度慢的问题。通过建立不同预测模型进行对比, 验证了所提算法和模型的有效性。

关键词: 主成分分析; 遗传算法; 功率预测; BP 神经网络; 光伏系统

Short term forecasting of photovoltaic output power based on principal component analysis and genetic optimization of BP neural network

XU Tongyu, MA Yiming, CAO Yingli, TANG Rui, CHEN Junjie

(College of Information and Electrical Engineering, Shenyang Agricultural University, Shenyang 110161, China)

Abstract: In view of the fluctuation and the intermittence of the output power of the photovoltaic system, based on principal component analysis (PCA) and genetic algorithm (GA) optimization, a short term forecasting method of BP neural network power is proposed. Direct forecasting of output power is done by historical power data and real time meteorological factors. It uses principal component analysis to reduce the dimension of multiple original variables into a few independent variables, so that it can optimize the initial weights of back-propagation's threshold and overcome the traditional BP neural network easy to fall into local minimum point, and the problems of slow convergence speed. The results of the comparison for different forecast models validate the effectiveness of the algorithm and proposed model.

This work is supported by Natural Science Foundation of Liaoning Province (No. 2013020141).

Key words: principal component analysis; genetic algorithm; power forecasting; BP neural network; photovoltaic system

0 引言

随着国家对光伏补贴政策的推进, 光伏发电大量并入电网, 分布式户用型太阳能发电正在兴起。而我国北方地区秋冬季节的太阳能资源较丰富, 有很大的利用价值^[1]。因此, 通过对北方户用型光伏系统进行输出功率短期预测, 能够更合理地制定发电计划与负荷分布规划, 对光伏系统的发展有重大意义^[2-4]。

目前国内外一些专家学者针对光伏系统输出功率预测问题进行了研究, 通常来说, 短期功率预测是指预测未来一天或几天内各时刻光伏系统的输出

功率, 主要有两种方式。第一种是通过分析精细的气象数据和建立光伏发电系统数学模型进行间接预测, 需要气象学上的研究作为支撑, 在拥有较多光照强度测量站的欧洲比较流行。第二种是基于历史数据和天气预报数据的直接预测, 在国内研究较多^[5]。文献[6]采用组合权重法求误差值, 根据误差值确定相似日进行功率预测, 一定程度提高预测精度, 但该方法输入量的确定较复杂, 模型建立较困难。文献[7]基于遗传算法和模糊径向基神经网络进行预测, 提出平滑控制策略对光伏功率进行调节, 但数据选择较复杂、预测时间长。文献[8]将 BP 神经网络、RBF 神经网络和 RNN 神经网络进行对比, 建立三个模型预测太阳辐照强度进而计算发电功率。

但此方法过于依赖数学模型和详细气象数据, 需要对气象数据进行长期准确的监测, 对于不同类型的光伏系统模型参数的确定也较困难。

本文提出了一种基于主成分分析与遗传优化的 BP 神经网络光伏发电输出功率短期预测模型, 该模型基于光伏发电历史功率数据和天气预报数据, 无需大量详细的历史气象数据, 同时利用主成分分析对数据降维, 减少了输入量的个数, 用遗传算法优化了 BP 神经网络的权值阈值, 一定程度上克服了神经网络算法全局搜索能力较弱、收敛速度慢等问题, 能较大幅度提高预测精度, 为光伏系统输出功率短期预测提供了一种行之有效的方法。

1 光伏系统发电功率影响因素分析

本文收集整理了辽宁省沈阳市沈阳农业大学信电学院光伏发电监测系统的发电功率与实时监测的气象信息, 包括实时发电功率、太阳辐照强度、环境温度、风速、风向等指标。分析各因素之间的相关关系, 以确定预测模型的输入量。相关系数 r 的计算公式为

$$r = \frac{\sum_{i=1}^N (x_i - \hat{x})(y_i - \hat{y})}{\sqrt{\sum_{i=1}^N (x_i - \hat{x})^2} \sqrt{\sum_{i=1}^N (y_i - \hat{y})^2}} \quad (1)$$

由公式计算出发电功率与气象因素相关关系如表 1。

表 1 光伏发电功率与气象因素相关系数

Table 1 Correlation of meteorological elements and PV power

相关系数	发电功率	辐照强度	环境温度	风速	风向
发电功率	1	0.985	0.321	0.017	0.036

由表 1 可知, 发电功率与太阳辐照强度和温度的相关系数较高, 在功率预测时需要考虑这两项因素, 文献[9]中对其相关关系进行了详细分析, 在此不赘述; 与风速和风向的相关性较弱, 在预测时可忽略。

此外, 影响功率的其它因素如系统转换效率、光伏组件安装角等已经隐含在历史输出功率数据中, 在预测时可不予考虑^[10]。综上, 选择历史输出功率数据、太阳辐照强度和环境温度作为预测模型的输入量。

2 基于主成分分析和遗传优化 BP 神经网络的功率短期预测模型

2.1 主成分分析原理

光伏输出功率的影响因素较多, 各变量间存在

一定相关性, 增加了预测难度。采用主成分分析的方法对变量进行降维, 消除了变量间的相关关系, 利用少数几个变量代替原来的多个变量, 同时又基本涵盖原始变量的全部内容。主成分分析的原理如下。

设研究的每组对象涉及 p 个变量, 共有 n 组对象, 则构成一个 $n \times p$ 阶的数据矩阵。

(1) 将原始数据矩阵标准化变换, 得到矩阵 \mathbf{X} 。

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad (2)$$

(2) 建立变量的相关系数阵 \mathbf{R} 。

(3) 求 \mathbf{R} 的特征根 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$ 及其对应的单位特征向量 e_1, e_2, \dots, e_p 。

(4) 确定主成分数量。定义主成分的贡献率

$$\text{为 } \frac{\lambda_i}{\sum_{k=1}^p \lambda_k}, \quad (i=1, 2, \dots, p)。 \text{ 累计贡献率为 } \frac{\sum_{k=1}^i \lambda_k}{\sum_{k=1}^p \lambda_k},$$

($i=1, 2, \dots, p$)。一般取累计贡献率达到 85% 以上的特征值 $\lambda_1, \lambda_2, \dots, \lambda_m$ 对应的第 1、第 2、...、第 m ($m \leq p$) 个主成分。

(5) 主成分表达式为

$$\begin{cases} y_1 = e_{11}x_1 + e_{12}x_2 + e_{13}x_3 + \cdots + e_{1p}x_p \\ y_2 = e_{21}x_1 + e_{22}x_2 + e_{23}x_3 + \cdots + e_{2p}x_p \\ \vdots \\ y_m = e_{m1}x_1 + e_{m2}x_2 + e_{m3}x_3 + \cdots + e_{mp}x_p \end{cases} \quad (3)$$

式中: e_p 为原始变量相关矩阵第 i 个特征值所对应的 p 维特征向量; $[x_1 \ x_2 \ \cdots \ x_p]^T$ 是 p 维的初始输入变量。

2.2 遗传算法原理

遗传算法是模拟生物进化规律的计算模型, 具备不依赖梯度信息、鲁棒性强等优势。其数学规划模型如式(4), 原理流程见图 1。

$$\begin{cases} \max f(x) \\ x \subset R \\ R \subset U \end{cases} \quad (4)$$

式中: U 是基本空间; R 是 U 的子集。

遗传算法构成要素如下:

(1) 染色体编码方法。使用固定长度的二进制来表示个体, 初始个体基因值可用 $\{0, 1\}$ 之间均匀分布的随机值生成。

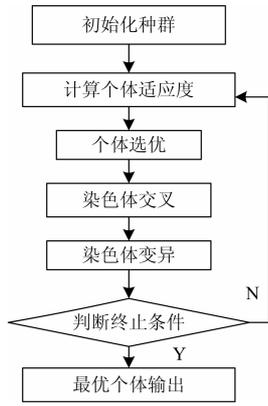


图1 遗传算法原理流程图
Fig.1 Principle graph of GA

(2) 个体适应度评价。按与个体适应度成正比的概率来决定当前群体中每一个体遗传到下一代群体中的机会多少。

(3) 遗传算子。包括选择、交叉和变异。选择运算采用轮盘赌法，交叉运算方法采用实数交叉法，第 k 个染色体 C_k 和第 l 个染色体 C_l 在 j 位交叉操作方法为

$$\begin{cases} C_{kj} = C_{kj}(1-b) - C_{lj}b \\ C_{lj} = C_{lj}(1-b) + C_{kj}b \end{cases} \quad (5)$$

其中， b 是 $[0,1]$ 间的随机数。

变异运算选取第 k 个个体的第 j 个基因进行变异，操作方法为

$$C_{ij} = \begin{cases} C_{ij} + (C_{ij} - C_{\max}) \times f(g) & r \geq 0.5 \\ C_{ij} - (C_{\min} - C_{ij}) \times f(g) & r < 0.5 \end{cases} \quad (6)$$

其中： C_{\max} 为基因 C_{ij} 的上界； C_{\min} 为基因 C_{ij} 的下界； $f(g) = r_2(1 - g/G_{\max})$ ， r 为 $[0,1]$ 间一个随机数； g 为当前迭代次数； G_{\max} 是最大进化次数。

(4) 运行参数。包括群体大小 N ，遗传算法终止进化代数 T ，交叉概率 p_c ，变异概率 p_m 等。

2.3 预测模型设计

利用遗传算法全局搜索能力强的特点，用 Matlab 建立与训练 GA-BP 神经网络预测模型的步骤如下所述。

(1) 清空环境变量，将数据归一化。

(2) 创建一个新的前向神经网络，设隐藏神经元数为 9 个，学习率为 0.05，设置动量因子、最小均方根误差等训练参数，进行 BP 网络训练和预测。

(3) 遗传算法参数初始化。确定迭代次数为 10，种群规模为 10，交叉概率 $p_c = 0.75$ ，变异概率 $p_m = 0.25$ 。

(4) 种群初始化。随机产生一个种群，计算染色

体的适应度，记录每一代迭代中最好的适应度和平均适应度。

(5) 迭代求解最佳初始阈值权值。通过选择、交叉、变异，找到最小和最大适应度染色体及其在种群中的位置，代替上一次进化中最好染色体，替换最差个体和相应适应度值，记录每一代迭代中最好适应度和平均适应度，达到最大迭代数时终止。

(6) 把最优初始权值阈值赋给 BP 网络。进行 GA_BP 网络训练和预测。

根据以上分析构建的基于主成分的 GA-BP 功率短期预测模型结构如图 2。输入量共 5 个， y_1, y_2, \dots, y_5 为主成分分析后选择的主成分；输出量共 10 个， p_1, p_2, \dots, p_{10} 为预测日每小时平均输出功率值。

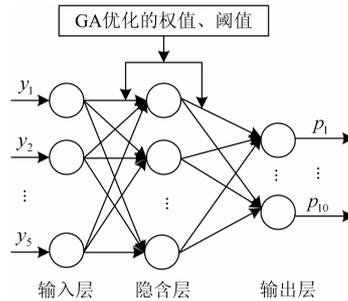


图2 基于主成分分析的 GA-BP 功率短期预测模型
Fig.2 Short-term PV power forecasting model by GA-BP neural network based on PCA

3 预测实例分析

3.1 基于 PCA 的数据预处理

预测数据来源为辽宁省沈阳市沈阳农业大学 40 kWp 户用式光伏发电系统，以秋冬季节为例，对其历史输出功率及监测到的相关气象数据进行主成分分析。采集历年 9 月、10 月、11 月的共 200 组监测数据，采样间隔为 6 min，考虑到北方地区秋冬季节日出较晚、日落较早，且夜晚没有光照，光伏系统输出功率值为 0，所以每组数据中包含的变量有：预测日前一天 6:32-16:26，每小时的平均输出功率；预测日前一天 6:32-16:26，每小时的平均辐射强度；预测日前一天环境温度最大值、最小值、差值；预测日环境温度最大值、最小值、差值。共计 26 个输入量。

首先，剔除样本数据中的畸变数据，用上一时刻数据代替。

然后，将样本进行标准化变换，有

$$z_{ij} = (x_{ij} - x_i) / s_i \quad (7)$$

其中： z_{ij} 是标准化变换后的变量值； x_{ij} 为原变量值； x_i 是各变量的算数平均值； s_i 是标准差。

接下来, 进行主成分分析。得到的协方差矩阵特征值、方差贡献率和累计贡献率如表 2。

表 2 主成分特征值与方差贡献率

Table 2 Eigenvalue and variance contribution rate of principal components

编号	特征值	方差贡献率/(%)	累计贡献率/(%)
1	13.3789	51.4572	51.4572
2	4.8579	18.6841	70.1414
3	2.8287	10.8796	81.0210
4	1.1385	4.3787	85.3998
5	1.0152	3.9045	89.3043
6	0.6383	2.4552	91.7595
7	0.4745	1.8251	93.5846
8	0.3713	1.4282	95.0128
9	0.2636	1.0137	96.0265
⋮	⋮	⋮	⋮
26	0.0003	0.0012	100.0000

由表 2 可知, 前 5 个主成分的累计贡献率已达到 89.3%, 能很好地概括原始变量, 所以选择前 5 个主成分作为输入神经网络预测的输入量。其特征值对应特征向量如表 3 所示, 通过式(3)即可写出前 5 个主成分 y_1, y_2, \dots, y_5 的表达式。

表 3 前 5 个主成分特征值对应的特征向量

Table 3 Characteristic vector of eigenvalue of the first five principal components

	y_1	y_2	y_3	y_4	y_5
e_1	0.2143	0.16	-0.2067	0.0097	0.1231
e_2	0.2179	0.0553	-0.2912	0.013	0.1397
e_3	0.2039	-0.0687	-0.2846	0.0184	0.1816
e_4	0.2251	-0.123	-0.2177	0.0732	-0.1152
e_5	0.2305	-0.1688	-0.0657	-0.0339	-0.2145
e_6	0.2338	-0.157	0.0841	-0.0025	-0.248
e_7	0.2251	-0.0891	0.2679	0.0671	-0.1225
e_8	0.2278	-0.0161	0.2876	-0.01	0.0421
e_9	0.2185	0.1214	0.2397	-0.0349	0.2074
⋮	⋮	⋮	⋮	⋮	⋮
e_{26}	0.057	0.1647	-0.0147	0.7147	-0.0968

最后, 求出 200 组数据的前 5 个主成分的值, 进行非标准化变换后建模。其中前 150 组作为训练样本集, 后 50 组作为测试样本集。

3.2 光伏发电功率预测及结果分析

为保证实验的科学性, 选取在模型训练时没有使用的两天, 即 2015 年 10 月 13 日(晴天)和 2015 年 10 月 16 日(多云)进行光伏系统短期输出功率预测, 分别验证本文所提出的基于主成分分析和遗传优化 BP 神经网络功率预测模型在稳定型天气和突变型天气下的预测精度。同时, 为了验证主成分分

析方法提出的必要性和有效性, 证明遗传算法在改进 BP 神经网络上的优势, 建立了对比预测模型。模型的组合方式见表 4。

表 4 模型组合表

Table 4 Combination of models

模型	主成分分析	遗传算法优化
Model 1	×	×
Model 2	√	×
Model 3	×	√
Model 4	√	√

图 3~图 5 为晴天条件下 4 个模型的预测对比, 图 6 为突变型天气下 4 个模型的预测对比。采用平均绝对百分比误差 E_{MAPE} 和均方根误差 E_{RMSE} 对各模型的预测性能进行评估, 评估结果见表 5。分析可知, 晴天条件下, 与 Model 1 相比, Model 2 采用主成分分析将原来的 26 个输入量减少为 5 个输入量, 模型训练所用时间缩短了三分之二, 预测精度也有较大提高, E_{MAPE} 从 14.79% 降到 8.43%; Model 3 采用了遗传算法优化, 预测值的离散程度下降,

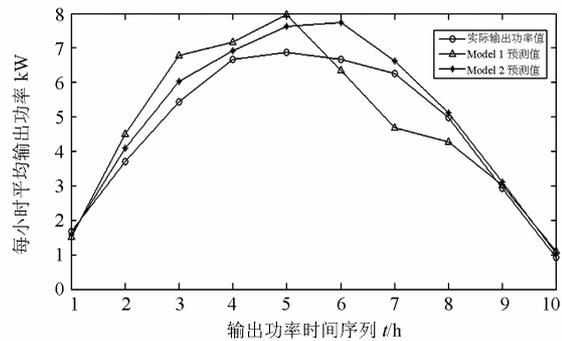


图 3 晴天条件下 Model 1 和 Model 2 预测结果对比

Fig. 3 Forecasting result of Model 1 and Model 2 in sunny conditions

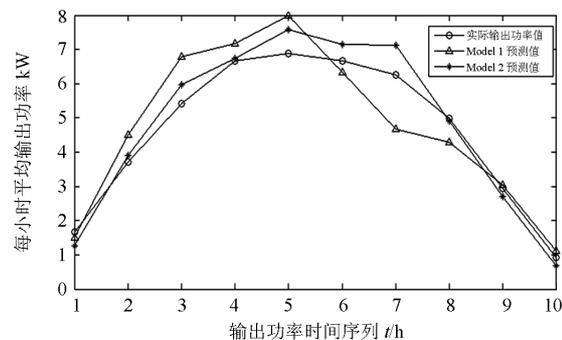


图 4 晴天条件下 Model 1 和 Model 3 预测结果对比

Fig. 4 Forecasting result of Model 1 and Model 3 in sunny conditions

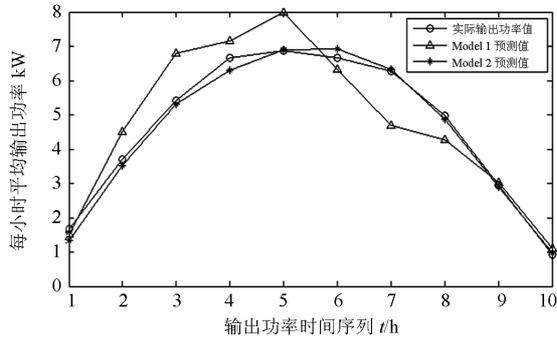


图 5 晴天条件下 Model 1 和 Model 4 预测结果对比
Fig. 5 Forecasting result of Model 1 and Model 4 in sunny conditions

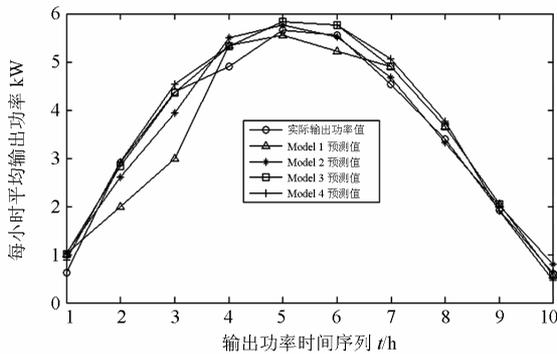


图 6 突变型天气下预测结果
Fig. 6 Forecasting result in the mutant weather

表 5 模型预测评估

Table 5 Forecasting results assessment of models

模型	稳定型天气		突变型天气	
	E_{RMSE}	E_{MAPE}	E_{RMSE}	E_{MAPE}
Model 1	0.8437	14.79%	0.3407	17.07%
Model 2	0.2489	8.43%	0.0879	14.10%
Model 3	0.4593	10.73%	0.0743	11.07%
Model 4	0.0386	4.73%	0.0680	10.88%

E_{RMSE} 从原来 0.8473 降为 0.4593, 系统的拟合效果更好, 由于初始化权值需要一定时间, 故训练时间相对增加; Model 4 在主成分分析法和遗传优化的共同下, 既提高了精度, 又缩短了训练时间, E_{MAPE} 为 4.73%, E_{RMSE} 为 0.0386, 系统拟合效果优于其他 3 个模型。

突变型天气下 4 个模型预测精度均有所降低, 主要原因是受云朵遮挡、气温不定等因素影响, 太阳辐射强度会产生较大波动, 输出功率也相应发生波动, 进而规律性减弱, 增加预测难度。但是采用

Model 4, E_{RMSE} 为 0.068, E_{MAPE} 为 10.88%, 还在可接受范围内。

4 结论

本文提出了一种将 PCA 算法和遗传算法结合的 BP 神经网络光伏输出功率短期预测方法, 将影响输出功率的多项因素通过主成分分析降维成少数几个彼此独立、互不相关的变量, 将遗传算法的全局搜索特性应用在 BP 神经网络的权值阈值优化上, 建立对比模型分别预测稳定型和突变型天气下的输出功率。结果表明, 该方法一定程度上提高了模型的预测精度和训练速度, 改善了传统 BP 网络的缺陷, 对不同天气类型均可很好地预测, 对北方地区合理制定光伏发电计划和负荷规划有重要意义。

参考文献

- [1] 孔波利, 崔丽艳, 丁钊, 等. 基于风光混合模型的短期功率预测方法研究[J]. 电力系统保护与控制, 2015, 43(18): 62-66.
KONG Boli, CUI Liyan, DING Zhao, et al. Short term power prediction based on hybrid wind-PV forecasting model[J]. Power System Protection and Control, 2015, 43(18): 62-66.
- [2] 姚仲敏, 潘飞, 沈玉会, 等. 基于 GA-BP 和 POS-BP 神经网络的光伏电站出力短期预测[J]. 电力系统保护与控制, 2015, 43(20): 83-89.
YAO Zhongmin, PAN Fei, SHEN Yuhui, et al. Short-term prediction of photovoltaic power generation output based on GA-BP and POS-BP neural network[J]. Power System Protection and Control, 2015, 43(20): 83-89.
- [3] 李霄, 王昕, 郑益慧, 等. 基于改进最小二乘支持向量和预测误差校正的短期风电负荷预测[J]. 电力系统保护与控制, 2015, 43(11): 63-69.
LI Xiao, WANG Xin, ZHENG Yihui, et al. Short-term wind load forecasting based on improved LSSVM and error forecasting correction[J]. Power System Protection and Control, 2015, 43(11): 63-69.
- [4] 崔红芬, 汪春, 叶季蕾, 等. 多接入点分布式光伏发电系统与配电网交互影响研究[J]. 电力系统保护与控制, 2015, 43(10): 91-97.
CUI Hongfen, WANG Chun, YE Jilei, et al. Research of interaction of distributed PV system with multiple access points and distribution network[J]. Power System Protection and Control, 2015, 43(10): 91-97.
- [5] 茆美琴, 周松林, 苏建徽. 基于风光联合概率分布的微电网概率潮流预测[J]. 电工技术学报, 2014, 29(2): 55-63.
MAO Meiqin, ZHOU Songlin, SU Jianhui. Probabilistic

- power flow forecasting of microgrid based on joint probability distribution about wind and irradiance[J]. Transactions of China Electrotechnical Society, 2014, 29(2): 55-63.
- [6] 杨锡运, 刘欢, 张彬, 等. 组合权重相似日选取方法及光伏输出功率预测[J]. 电力自动化设备, 2014, 34(9): 118-122.
YANG Xiyun, LIU Huan, ZHANG Bin, et al. Similar day selection based on combined weight and photovoltaic power output forecasting[J]. Electric Power Automation Equipment, 2014, 34(9): 118-122.
- [7] 叶林, 陈政, 赵永宁, 等. 基于遗传算法-模糊径向基神经网络的光伏发电功率预测模型[J]. 电力系统自动化, 2015, 39(16): 16-22.
YE Lin, CHEN Zheng, ZHAO Yongning, et al. Photovoltaic power forecasting model based on genetic algorithm and fuzzy radial basis function neural network[J]. Automation of Electric Power Systems, 2015, 39(16): 16-22.
- [8] YONA A, SENJYU T, SABER A Y, et al. Application of neural network to one-day-ahead 24 hours generating power forecasting for photovoltaic system[C] // 2007 International Conference on Intelligent Systems Applications to Power Systems (ISAP), Toki Messe, Niigata, 2007.
- [9] 张雪莉, 刘其辉, 马会萌, 等. 光伏电站输出功率影响因素分析[J]. 电网与清洁能源, 2012, 28(5): 75-81.
ZHANG Xueli, LIU Qihui, MA Huimeng, et al. Analysis of influencing factors of output power of photovoltaic power plant[J]. Power System and Clean Energy, 2012, 28(5): 75-81.
- [10] 姚致清, 于飞, 赵倩, 等. 基于模块化多电平换流器的大型光伏并网系统仿真研究[J]. 中国电机工程学报, 2013, 33(36): 27-33.
YAO Zhiqing, YU Fei, ZHAO Qian, et al. Simulation research on large-scale PV grid-connected systems based on MMC[J]. Proceedings of the CSEE, 2013, 33(36): 27-33.

收稿日期: 2016-01-13; 修回日期: 2016-04-05

作者简介:

许童羽(1967-), 男, 博士, 教授, 博士生导师, 主要研究方向为农村户用光伏系统分析; E-mail: yatongmu@163.com

马艺铭(1992-), 女, 硕士研究生, 主要研究方向为地方电力系统及其自动化、光伏发电系统输出功率分析; E-mail: 547566468@qq.com

曹英丽(1977-), 女, 通信作者, 博士, 副教授, 主要研究方向为电网信息化研究。E-mail: caoyingli@163.com

(编辑 姜新丽)