基于相似日理论和BP神经网络的光伏发电功率预测

光伏发电系统的输出功率受到太阳辐照强度、辐照时长和气温等多种气象因素的影响，具有随机性和波动性。对各种气象影响因素进行合理的选取和处理，寻求光伏发电系统输出功率与各种相关因素之间的内在联系，对光伏发电系统未来的输出功率进行科学的预测。由于具有相似气象条件下的光伏阵列输出功率具有较大的关联性，提出选择光伏阵列输出功率相似日的方法，设计基于相似日原理的BP神经网络光伏阵列输出功率预测模型，利用我国某地光伏发电系统的实测数据对模型进行验证。结果表明，模型有较好的预测准确度，具有一定的实用性及可行性。

倪春花[1] 陈国恩[2] 朱 伟[2] 王丙文[3] 吴在军[1]/1.东南大学电气工程学院 2.国网嘉兴供电公司 3.国电南瑞科技股份有限公司

倪春花/硕士研究生

关键词/Keywords

光伏电站·
功率预测·
相似日·
BP神经网络·

0 引言

光伏发电由于其可再生、环保灵活的特性得到广泛关注和快速发展[1]。但由于光伏发电系统的输出受到太阳辐照强度和众多天气因素的影响，其发电量的变化是一个非平稳的随机过程，对大电网而言是一个不可控源，其发电的随机波动性会对电网造成冲击，影响电网运行的可靠性与稳定性[2]。加强光伏发电功率预测的研究，预先获得光伏发电系统的日发电量曲线，从而协调电力系统制定发电计划，减少光伏发电的随机性对电力系统的影响，对光伏电站和电网都有着重要的经济意义[3-4]。

国家863高技术研究发展计划（2014AA052002），国家自然科学基金（51307023），新世纪优秀人才支持计划资助（NCET－13－0129），江苏省2015年度普通高校研究生实践创新计划项目（SJLX15_0053）。

按照不同的分类标准，光伏发电功率预测方法分类不同[5]。按预测时间尺度不同，可分为短期功率预测和超短期功率预测。超短期功率预测的时间尺度为0.5～6 h，短期功率预测的时间尺度一般为1～2天[6-7]。按预测物理量的不同，可分为直接预测和间接预测。直接预测是直接对光伏发电系统的输出功率进行预测，间接预测方式首先对地表太阳辐照强度进行预测，然后根据光伏发电系统出力模型得到系统的输出功率。按预测模型原理不同，可分为物理方法和统计方法。物理方法将气象预测数据作为输入，采用物理方程进行预测。统计方法对历史数据进行统计分析，找出其内在规律并用于预测。文献[8-10]考虑了多种气象因素，建立了基于神经网络、提前一天的光伏发电功率预测模型；文献[11]考虑诸多影响因子的复杂性和非线性，建立了光伏系统出力的支持向量机回归模型；文献[12]提出一种基于马尔科夫链，通过统计光伏电站历史功率数据建模的方法直接预测光伏电站出力。这些预测方法对气象因素的处理仍不够精细，当气象条件发生剧烈变化时，模型预测准确度有待提高。

本文提出了光伏发电系统相似日的选择方法，由于相似日的输出功率曲线具有很高的相似性，将与预测日具有相似气象条件的历史日作为预测样本，建立BP神经网络预测模型。根据某光伏电站的实测数据，对预测模型进行训练和预测。结果表明，预测模型有较好准确

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

度，具有可行性与实用性。

1 选取相似日

1.1 光伏电站输出功率特性

影响光伏发电系统输出功率的因素很多，如太阳辐射强度、阵列的转换效率、太阳能电池板安装角度、大气压和温度等。如果将所有可能的影响因素都作为预测模型的输入变量来考虑，那么会增加模型的复杂度和历史数据收集的难度。比较合理的处理方式是选取与输出功率关联性强的因素并对其做必要的预处理后将其作为预测模型的输入变量。

对于由固定安装角度太阳能电池板组成的光伏发电系统，单位面积的光伏阵列输出功率为

$$P_s=\eta SI[1-0.005(T_0+25)] \tag{1}$$

式中，I为光照强度，$\mathrm{kW/m^2}$；η为光伏电池转换效率；T_0为环境温度，℃；S为光伏阵列的面积，$\mathrm{m^2}$。

1.2 天气类型编码

对于既定的光伏发电系统来说，因为所有的历史输出功率时间序列都来自同一套发电系统，故历史发电量数据中已包含光伏阵列的系统信息，预测模型的输入变量可不用考虑光伏阵列的光电转换效率η、光伏阵列的面积S等隐含在历史输出功率数据中的因素。但太阳辐照强度I和环境温度T_0等外界影响因素的变化需在模型输入变量中考虑。

当天气类型发生变化时太阳辐射强度I也会发生变化，光伏阵列输出功率差异很大。为保证天气类型突变时的模型预测准确度，可在预测模型输入变量中考虑天气类型的影响。一般天气预报中给出的天气参数为较模糊的类型描述：如晴天、晴天到多云、阴天、阴天有小雨和小雨转大雨等。根据不断完善的光伏监控系统数据库，对大量有效的历史发电量数据进行统计分析，将模糊的描述性天气类型映射为0～1间的具体数值作为预测模型的输入变量参考。如表1所示。

表1 天气类型编码表

天气类型	量化值	天气类型	量化值
晴	0.9	多云	0.7
雾	0.7	霾	0.7
多云转雷阵雨	0.5	多云转小雨	0.2
多云转大雨	0.1	阴转多云	0.5
雷雨	0.4	雷阵雨	0.4
阵雨	0.4	阴	0.3
阴有小雨	0.2	小雨	0.15
小到中雨	0.1	中到大雨	0.05
中雨	0.1	大雨	0.05
中雪	0.05	暴雨	0.005

1.3 相似日选取原理

1.3.1 构造气象特征向量

根据前文对光伏阵列发电功率特性的分析，选取气温和天气类型（反映太阳辐射强度）作为影响光伏阵列输出功率的气象影响因素，构造每天气象特征向量[13]，即

$$\boldsymbol{Y}=[\ T_{\max}\quad T_{\min}\quad w\ ] \tag{2}$$

式中，$T_{\max}$为最高温；$T_{\min}$为最低温；w为天气类型。

1.3.2 计算相似度

1）首先采用“极差法”对各分量进行归一化，即

$$x_i(j)=\frac{y_i(j)-m(j)}{M(j)-m(j)} \tag{3}$$

式中，$y_i(j)$是第i日的第j个分量；$m(j)$、$M(j)$分别为第j个分量的最小值和最大值，$j=1$，2，3。

归一化后基准日和第j日的特征向量为

$$\boldsymbol{x}_0=[x_0(1),x_0(2),x_0(3)]^{\mathrm{T}} \tag{4}$$

$$\boldsymbol{x}_j=[x_j(1),x_j(2),x_j(3)]^{\mathrm{T}} \tag{5}$$

2）计算$\boldsymbol{x}_0$和$\boldsymbol{x}_j$在第k个因素的关联系数[14]，即

$$\varepsilon_j(k)=\frac{\min\limits_j\min\limits_k\left|x_0(k)-x_j(k)\right|+\rho\max\limits_j\max\limits_k\left|x_0(k)-x_j(k)\right|}{\left|x_0(k)-x_j(k)\right|+\rho\max\limits_j\max\limits_k\left|x_0(k)-x_j(k)\right|} \tag{6}$$

式中，ρ为分辨系数，其值一般取0.5。

3）综合各影响因素的关联系数，定义$\boldsymbol{x}_0$和$\boldsymbol{x}_j$的相似度，即

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

$$F_j = \prod_{k=1}^{3} \varepsilon_j(k) \tag{7}$$

采用连乘方式定义相似度，可以简单、自动地识别主导因素，并解决各因素权重设定问题。

1.3.3 选取相似日

选择第i日的相似日的具体步骤如下。

1）从最临近历史日开始，逆向逐日计算第j日、第i日的相似度值F_j。

2）选取最近N日中相似度最高的m日或者相似度$F_j \geqslant r$（r为一定的数值，如$r=0.85$）的m日作为第i日的相似日，构成预测模型训练样本集。

2 预测模型设计与训练

2.1 预测模型结构设计

BP神经网络是基于误差反向传播的多层前馈神经网络，具有强大的非线性映射能力和较好的泛化能力，适用于受自然环境影响大，随机性大的光伏发电功率预测。故采用BP神经网络对光伏发电系统输出功率进行预测。设计模型结构如图1所示。

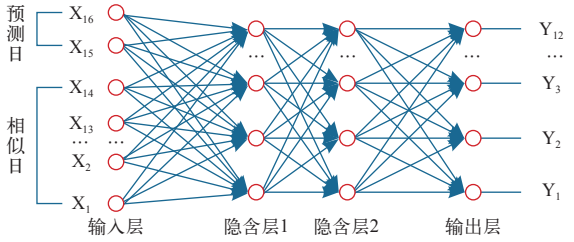

图1 相似日BP神经网络光伏发电功率预测模型

采用16个输入变量，由于处于夜间的12个时间点光伏阵列的发电量基本为0，故选取白天的12个发电时间序列（7：00～18：00）作为预测模型的12个输入变量，相似日的最高气温和最低气温作为预测模型的两个输入变量，最后两个输入变量是预测日的最高气温和最低气温。

采用12个输出变量，预测模型输出的是预测日白天12个时间点的发电功率，对应预测日的发电量序列。

采用两个隐层，以训练速度为代价获得更高的预测准确度。根据隐含层节点数的经验公式[15]取8个隐层节点。

$$\begin{gathered} m = \sqrt{n+l} + a \\ m = \log 2^n \\ m = \sqrt{nl} \end{gathered} \tag{8}$$

式中，n为输入层节点数；m为隐含层节点数；l为输出层节点数；a为1～5的常数。

2.2 预测模型训练与评估

2.2.1 样本训练集

根据相似日选取原理编写相应算法选取某预测日的m个相似日，按相似度从大到小顺序排列。利用这m个相似日的发电量数据和温度构造m个训练样本。具体过程如下。

1）第i个训练样本的输入变量。第i个相似日的12个发电量数据，第i个相似日的最高气温和最低气温，第$i+1$个相似日的最高气温和最低气温。

2）第i个训练样本的理论输出（教师数据）：第$i+1$个相似日的12个发电量数据。特别的，当$i=m$时，即第m个训练样本的理论输出为第1个相似日的12个发电量数据。

2.2.2 训练样本数据预处理

由于所能获得的数据的有限性，有些实际上对光伏发电功率有影响的变量无法在预测模型中考虑。这使得训练样本存在噪声，过大的噪声会湮没样本本来的特性，使神经网络无法学习样本的内在规律性。这种明显异常的样本需在进行网络训练之前剔除。

训练样本的原始数据中，不同变量的单位不同，数量级差异也可能很大，由神经元激活函数的特性可知，神经元的输出通常被限制在一定范围内，若直接以原始数据进行网络训练，很容易引起神经元饱和。因此需在网络训练之前对样本数据进行归一化处理，以消除样本数据形式不同带来的弊端。归一化后，样本输入数据和目标输出均落在0～1之间。归一化公式为

$$\begin{gathered} P_i^* = P_i / P_{\max} \\ T_{\text{high}}^* = T_{\text{high}} / T_{\text{h}\max} \\ T_{\text{low}}^* = T_{\text{low}} / T_{\text{l}\max} \end{gathered} \tag{9}$$

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

2.2.3 神经网络训练过程

在利用BP神经网络进行光伏发电功率预测之前，首先要对BP神经网络进行训练。假设采用训练样本集的第一个样本开始对神经网络进行训练，首先由训练样本给出预测模型输入量和理论输出。然后根据激励函数计算隐含层、输出层各节点的输入输出。

如果神经网络输出层的计算输出与理论输出（教师数据）不相同，需要对神经网络的连接权值进行不断修正。权值修正采用附加动量的自适应学习速率负梯度修正法，此方法将附加动量与自适应学习速率相结合，能提高模型收敛速度，并减小BP神经网络陷入局部最小解的可能性。具体如下。

1）传统BP神经网络权值修改算法为

$$w_{ij}(t+1)=w_{ij}(t)+\Delta w_{ij}(t) \tag{10}$$

2）只加入附加动量后网络权值修改算法为

$$w_{ij}(t+1)=w_{ij}(t)+\Delta w_{ij}(t)+\alpha[w_{ij}(t)-w_{ij}(t-1)] \tag{11}$$

式中，α为动量因子，取值需在0～1之间，一般取为0.95。

3）加入自适应学习速率后网络权值修改算法为

$$w_{ij}(t+1)=w_{ij}(t)+\alpha[w_{ij}(t)-w_{ij}(t-1)]+(1-\alpha)\eta(t+1)\frac{\partial E}{\partial w} \tag{12}$$

$$\eta(t+1)=\begin{cases}K_1\eta(t) & E(t+1)<E(t)\\ K_2\eta(t) & E(t+1)>E(t)\\ \eta(t) & E(t+1)=E(t)\end{cases} \tag{13}$$

式中，t为当前迭代次数；$\eta(t)$为第t次迭代时所采用的学习速率；$K_1>1$，一般取为1.05；$K_2<1$，一般取为0.7；$E(t+1)$、$E(t)$分别为第$t+1$次和第t次迭代的误差总平方和；$\frac{\partial E}{\partial w}$为误差平方和对权值的负梯度。

网络训练总流程如图2所示。

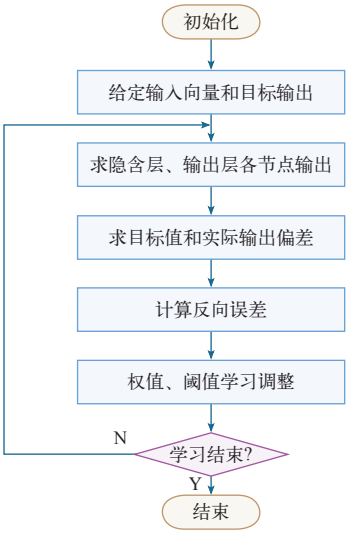

图2 BP神经网络训练流程图

2.2.4 预测结果评估

对神经网络发电功率预测模型进行准确的评估，需要采用一定的评估指标，如平均绝对偏差、均反差和平均绝对百分比误差。此外，还有其他一些指标，但最常用的还是平均绝对百分比误差（MAPE），即

$$\mathrm{MAPE}=\frac{100}{N}\sum_{i=1}^{N}\frac{\left|P_{\mathrm{f}}^{i}-P_{\mathrm{a}}^{i}\right|}{P_{\mathrm{a}}^{i}}\% \tag{14}$$

式中，P_{a}为实际功率输出；P_{f}为预测功率输出；N为序列数据总数；i为数据的序号。

3 算例分析

以江苏宿迁某光伏电站的输出功率实测值和当地气象数据为例，功率实测值每隔1 h整点采集一次，取7：00～18：00的12个整点功率输出值作为预测模型的12个输入变量。将近两年的历史发电数据和气象数据分为训练集和测试集。对于测试集中某“待预测日”，根据上文所述相似日选取方法在训练集中选取该日的20个相似日，构成训练样本集。

算例分析主要是对所提出的基于相似日理论的BP神经网络预测模型与传统BP神经网络预测模型的预测效果进行分析比较。

3.1 基于相似日原理的功率预测

对测试集中的2013年4月15日（晴天，最高温度为30 ℃、最低温度为17 ℃）进行光伏发电功率预测，该日期为随机选择，该日的实际发电功率数据如表2所示。

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

表2 2013年04月15日实际发电功率P_a

时刻	7：00	8：00	9：00	10：00	11：00	12：00
发电功率/MW	0.86	2.33	3.77	5.07	5.72	6.00
时刻	13：00	14：00	15：00	16：00	17：00	18：00
发电功率/MW	5.75	5.01	3.95	1.46	1.14	0.16

选出满足条件：天气类型为晴天，最高温度为30 ℃、最低温度为17 ℃的20个相似日，选取结果如表3所示，相似日功率输出如图3所示。

表3 30－17－晴相似日选取结果

ID	Date	T_h/℃	T_l/℃	Type
1	13-06--13	31	17	0.9
2	13-04-27	30	18	0.9
3	13-05-11	30	18	0.9
4	13-09-01	28	17	0.9
5	13-10-12	28	17	0.9
6	13-06-03	30	19	0.9
7	13-08-31	27	17	0.9
8	13-10-13	28	16	0.9
9	13-05-12	31	19	0.9
10	13-09-02	29	19	0.9
11	13-09-03	29	19	0.9
12	12-09-05	27	18	0.9
13	12-09-25	27	18	0.9
14	13-06-04	31	20	0.9
15	12-09-18	26	16	0.9
16	13-05-13	33	19	0.9
17	13-05-14	33	19	0.9
18	13-10-11	27	15	0.9
19	13-05-21	32	20	0.9
20	12-09-15	26	15	0.9

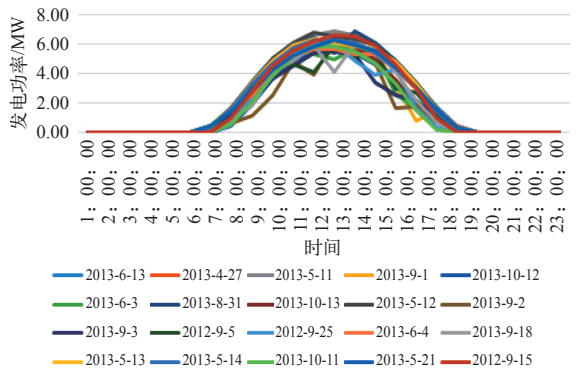

图3 晴天的相似日输出功率趋势图

由图3可以看出，相似日的输出功率曲线形状相似，变化趋势相似，幅值大致相同。用这20个相似日的发电功率数据和气象数据构成20个训练样本，对BP神经网络进行训练。训练完成后，利用相似度最高的那天的发电功率数据、气象数据和待预测日的气象数据作为模型输入，得到基于相似日原理的光伏发电预测输出功率。预测输出功率数据如表4所示。

表4 相似日预测功率P_{f_1}

时刻	7：00	8：00	9：00	10：00	11：00	12：00
发电功率/MW	1.28	2.95	4.52	5.65	6.08	6.13
时刻	13：00	14：00	15：00	16：00	17：00	18：00
发电功率/MW	6.07	5.63	4.27	2.75	1.24	0.22

3.2 基于历史日数据的传统功率预测

同样的，对测试集中的2013年4月15日（晴天，最高温度为30 ℃、最低温度为17 ℃）进行光伏发电功率预测，该日期为随机选择。该日发电功率数据参见表2。

与基于相似日原理的光伏发电功率预测不同的是，传统功率预测模型的训练样本是直接利用待预测日前20天的历史发电功率数据与气象数据构成。历史日气象数据如表5所示，发电功率曲线如图4所示。

表5 历史日选取结果

ID	Date	T_h/℃	T_l/℃	Type
1	13-04-14	26	16	0.9
2	13-04-13	25	15	0.4
3	13-04-12	24	15	0.9
4	13-04-11	19	14	0.7
5	13-04-10	20	13	0.7
6	13-04-09	20	13	0.7
7	13-04-08	21	12	0.7
8	13-04-07	18	11	0.9
9	13-04-06	15	8	0.7
10	13-04-05	20	11	0.1
11	13-04-04	22	16	0.4
12	13-04-03	20	14	0.4
13	13-04-02	20	15	0.7
14	13-04-01	19	15	0.15

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

（续）

ID	Date	T_h/℃	T_l/℃	Type
15	13-03-30	20	14	0.1
16	13-03-29	21	14	0.7
17	13-03-28	16	14	0.3
18	13-03-27	18	13	0.7
19	13-03-26	12	9	0.4
20	13-03-25	16	9	0.7

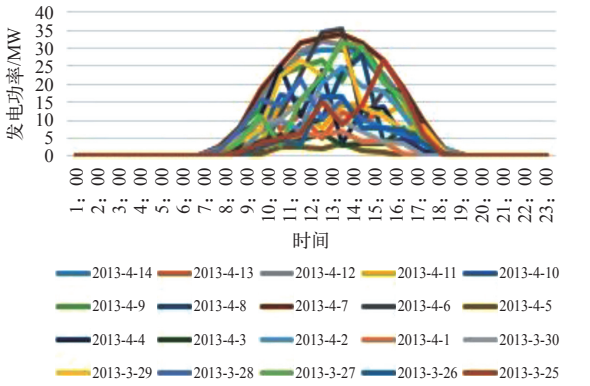

图4 历史日输出功率趋势图

由表5和图4可以看出，待预测日的20个历史日发电功率曲线形状各异，变化趋势复杂无规律，幅值差异大。用这样的20个历史日发电功率数据和气象数据构成20个样本，对BP神经网络进行训练。训练完成后，利用最临近历史日的发电功率数据、气象数据和待预测日的气象数据作为模型输入，得到基于历史日数据的传统光伏发电预测输出功率。预测输出功率数据如表6所示。

表6 历史日预测功率P_{f_2}

时刻	7：00	8：00	9：00	10：00	11：00	12：00
发电功率/MW	1.17	3.31	4.57	7.66	8.41	19.01
时刻	13：00	14：00	15：00	16：00	17：00	18：00
发电功率/MW	28.51	30.01	20.05	17.96	6.35	1.01

3.3 预测结果分析比较

两种预测模型的预测结果曲线如图5所示，预测结果的平均绝对百分比误差如表7所示。

可见，此算例中的基于历史日数据的传统功率预测模型已基本失效。这是因为历史日的气温、天气类型与待预测日的差异很大，变化趋势复杂，此模型未能合理考虑天气因素对光伏发电

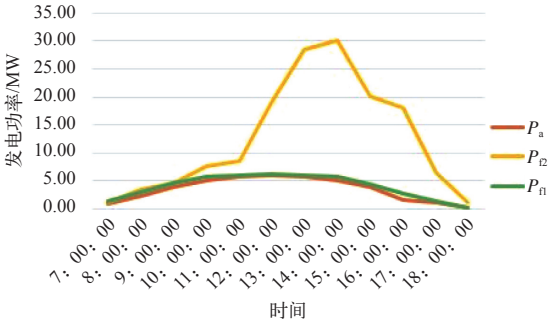

图5 不同预测模型预测结果误差对比

表7 预测误差比较（%）

预测模型	MAPE
基于相似日（P_{f_1}）	14.297
基于历史日（P_{f_2}）	319.55

功率输出的影响。而在所提出的基于相似日理论的BP神经网络预测模型中，对主要气象因素的处理比较精细，保证了天气类型一致和气温相近，有效地提高了光伏发电功率预测准确度。虽然与实际发电量有一定差距，但对于光伏发电系统配合电力系统调度中心提前制定发电计划仍有较高的参考价值。

由于功率数据和气象数据的有限性，影响光伏发电功率因素的复杂性等原因，通过合理选取影响因素的个数和表现形式，增加训练样本个数，提高训练样本质量，预测准确度还有提高的可能。

4 结束语

光伏发电系统输出功率受多种气象因素影响，具有一定随机性和波动性，对大电网而言是不可控源。随着光伏并网容量不断增大，其对大电网的影响越发不容忽视。通过量化天气类型和定义相似度，提出了相似日选取原理。利用相似日发电功率曲线的相似度较高，采用附加动量与自适应学习率相结合的BP神经网络权值学习算法，提出基于相似日原理的BP神经网络预测模型，对光伏发电功率进行日前预测，协调电力系统制定发电计划，减少光伏发电的随机化对电力系统的影响。

最后利用我国某地光伏发电系统的实测数据

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

对模型进行验证。结果表明，模型有较好的预测准确度，具有一定的实用性及可行性。

参考文献

[1] 胡云岩，张瑞英，王军．中国太阳能光伏发电的发展现状及前景 [J]．河北科技大学学报，2014，35 (1)：69-72．

[2] 陈炜，艾欣，吴涛，等．光伏并网发电系统对电网的影响研究综述 [J]．电力自动化设备，2013，33 (2)：26-32，39．

[3] 丁明，王伟胜，王秀丽，等．大规模光伏发电对电力系统影响综述 [J]．中国电机工程学报，2014，34 (1)：1-14．

[4] 卢静，翟海青，刘纯，等．光伏发电功率预测统计方法研究 [J]．华东电力，2010，38 (4)：563-567．

[5] 邓雅，胡书举，孟岩峰，等．光伏发电系统功率预测研究方法综述 [J]．电气制造，2013 (6)：50-53．

[6] 崔洋，孙银川，常倬林．短期太阳能光伏发电预测方法研究进展 [J]．资源科学，2013，35 (7)：1474-1481．

[7] Yona A，Senjyu T，Funabashi T．Application of recurrent neural network to short-term-ahead generating power forecasting for photovoltaic system [C]．IEEE Power Engineering Society General Meeting，2007．

[8] 陈昌松，段善旭，殷进军．基于神经网络的光伏阵列发电预测模型的设计 [J]．电工技术学报，2009，24 (9)：153-158．

[9] 张岚，张艳霞，郭嫦敏，等．基于神经网络的光伏系统发电功率预测 [J]．中国电力，2010，43 (9)：75-78．

[10] Adel M，Alessandro M P．Performance prediction of 20 kW grid-connected photovoltaic plant at trieste (Italy) using artificial neural network [J]．Energy Conversion and Management，2010，51 (12)：2431-2441．

[11] 栗然，李广敏．基于支持向量机回归的光伏发电出力预测 [J]．中国电力，2008，41 (2)：74-78．

[12] 丁明，徐宁舟．基于马尔可夫链的光伏发电系统输出功率短期预测方法 [J]．电网技术，2011，35 (1)：152-157．

[13] 王晓兰，葛鹏江．基于相似日和径向基函数神经网络的光伏阵列输出功率预测 [J]．电力自动化设备，2013，33 (1)：100-103，109．

[14] 傅美平，马红伟，毛建容．基于相似日和最小二乘支持向量机的光伏发电短期预测 [J]．电力系统保护与控制，2012，40 (16)：65-69．

[15] Benardos P G，Vosniakos G C．Optimizing feed forward artificial neural network architecture [J]．Engineering Applications of Artificial Intelligence，2007，20 (3)：365-382．

（收稿日期：2015－06－10）EA

公　　示

根据国家新闻出版广电总局《关于开展新闻记者证2015年度核验工作的通知》（新广出办发［2015］140号）要求，我单位《电气应用》杂志已对领记者证人员进行年度核验，现将我单位领新闻记者证人员张丰收进行公示。国家新闻出版广电总局举报电话：010-83138953。

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net