

Modeling and Prediction of Wind Power Statistics

Yucheng Yin

College of Science, Shaoyang University, Shaoyang Hunan
Email: 14394490@qq.com

Received: Jun. 12th, 2017; accepted: Jun. 26th, 2017; published: Jun. 30th, 2017

Abstract

In this paper, time series, artificial neural network, gray forecast are adopted to predict the power of wind turbines. By means of setting up three types of rational prediction models and performing error analysis, this paper presents an improvement scheme to make predictions more accurate. Finally, the models are applied to several wind turbines. The analysis demonstrates that the original measurement data often contain various errors, so the prediction accuracy of wind power cannot be infinitely increased.

Keywords

Wind Power, The Time Series, Artificial Neural Network, Grey Prediction, Combination Forecast

风电功率统计建模及预测

尹煜城

邵阳学院理学院, 湖南 邵阳
Email: 14394490@qq.com

收稿日期: 2017年6月12日; 录用日期: 2017年6月26日; 发布日期: 2017年6月30日

摘要

本文针对风电功率的预测问题, 分别采用时间序列法、人工神经网络、灰色预测法对未来机组输出的电功率建立了三种合理预测模型, 并通过对各种模型的误差分析, 进一步提出了改进的方案使其预测更加的准确, 在最后将模型推广到n台风电机组并给出了合理的模型。我们通过分析知道阻碍风电功率实时预测精度进一步改善的主要因素是因为实测数据本身就存在各种不可避免的误差, 得出风电功率的预测

精度无法得到无限提高的结论。

关键词

风电功率, 时间序列, 人工神经网络, 灰色预测, 组合预测

Copyright © 2017 by author and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

我国近年风电发展迅猛, 截止到 2015 年, 我国风电新增装机容量 30.5 GW, 同比上升 26.61%, 连续六年位居全球新增装机容量首位; 累计装机容量达到 145.1 GW, 其中并网容量达到 129 GW, 占全部发电装机容量 8.6%。2015 年, 风电总发电量 1863 亿千瓦时, 占全部发电量的 3.3%。与风资源分布特性及电网负荷布局相匹配, 我国的风电发展以并网型为主, 呈现典型的集群开发、弱电网接入、长距离外送的特点。风电的波动性、随机性给电网的有功平衡和电压调整带来了压力, 增加了电网规划和调度的难度, 也成为制约风电进一步发展的瓶颈[1]。掌握风电波动性在不同时间、空间尺度上的内在规律是解决大规模风电并网运行难题的关键基础。现今风力发电主要利用的是近地风能, 但近地风具有波动性、间歇性、低能量密度等特点, 因而风电功率也是波动的。大规模风电场接入电网运行时, 大幅度地风电功率波动会对电网的功率平衡和频率调节带来不利影响。因此, 我们要实现对风电场发电功率的尽可能准确地预测, 这样, 电力调度部门就能够根据风电功率变化预先安排调度计划, 保证电网的功率平衡和运行安全。

2. 问题提出

实时预测是风电功率预测的一种, 它要求滚动地预测每个时点未来 4 小时内的 16 个时点(每 15 分钟一个时点)的风电功率数值。根据国家能源局颁布的《风电场功率预测预报管理暂行办法》中的要求, 实时预测的误差不能大于 15% [2]。

实验风电场由 58 台风电机组构成, 每台机组的额定输出功率为 850 kW。实验中统计出了 2016 年 5 月 10 日至 2016 年 6 月 6 日时间段内该风电场中指定的四台风电机组(A、B、C、D)输出功率数据(分别记为 P_A , P_B , P_C , P_D ; 另设该四台机组总输出功率为 P_4)及全场 58 台机组总输出功率数据(记为 P_{58})。

根据风电功率预测精度要求现提出以下三个问题:

问题一: 风电功率实时预测方法及误差分析。

问题二: 研究风电机组的汇聚对于预测结果误差的影响。

问题三: 提高风电功率实时预测精度的探索。

通过解决上述三个问题, 找出阻碍风电功率实时预测精度进一步改善的主要因素。研究风电功率预测精度是否能无限提高。

3. 问题分析

1) 问题一: 风电功率实时预测方法及误差分析。

该问题是电功率的实时预测及误差分析，其主要研究目的是建立一定的数学模型来尽可能准确地做出风电功率的实时预测，并使的预测结果的误差在满足国家相关规定的基础上尽可能小，以便提供给电力调度部门，方便其优化调度安排。

该问题属于预测类的数学问题，且是直接利用历史数据，使用一定的数学模型进行预测。常见的方法有人工神经网络法、时间序列法(AMAR)、遗传算法、灰色分析预测法、卡尔曼滤波法、及其它算法。

问题一要求至少用三种预测方法对 P_A P_B P_C P_D P_4 P_{58} 这六个量在未来16个时点的风电功率数值进行预测，并对结果进行误差分析，确定实时预测的相对误差不能大于15%。由于实验中所要求的16个预测结果是滚动预测所得，一般来说，风电功率的预测值与实测值之间存在相对较大的误差，这就需要对结果进行误差分析后再根据分析结果对模型进行进一步优化。

基于以上考虑，我们可以分别用时间序列法建立数学模型一，用灰色分析预测法建立数学模型二，用人工神经网络法建立数学模型三，对结果进行预测，并将预测结果进行比较，同时分别对各模型所的预测结果进行分析。

2) 问题二：研究风电机组的汇聚对于预测结果误差的影响。

本问题要求分析风电机组的汇聚对与预测数据误差的影响。在我国主要采用集中开发的方式开发风电，各风电机组功率汇聚通过风电场或风电场群(多个风电场汇聚而成)接入电网。众多风电机组的汇聚会改变风电功率波动的属性，从而可能影响预测的误差。故而对风电机组的汇聚和其相应的预测数据误差进行分析，得出二者之间的关系，将对我们分析预测大规模的风电场群的风电功率提供参考[3]。

即问题二实质上是研究风电机组的台数与对应的风电功率预测值的相对误差之间的关系。

因此，我们可以用问题一中预测结果的相对误差，比较单台风电机组功率(P_A , P_B , P_C , P_D)预测的相对误差与多机总功率(P_4 , P_{58})预测的相对误差，再用时序分析法建立模型四来拟合确定台数的风电功率预测值的相对误差与对应风电机组的台数之间的函数关系，进而对风电机组汇聚给风电功率预测误差带来的影响做预期。

3) 问题三：提高风电功率实时预测精度的探索。

从问题一和问题二的结果我们可以看出模型一、模型二、模型三所得出的预测结果都存在一定程度的误差，而提高风电功率实时预测的准确程度对改善风电联网运行性能有重要意义。因而在模型一、模型二、模型三的基础上，构建有更高预测精度的实时预测方法是非常必要的。

通过对问题一的求解，我们可以分别得到模型一、模型二、模型三的拟合优度，再利用三个拟合优度的比值来确定三个模型所得预测值的权重，进而用组合预测的方法得出模型五，使得预测精度进一步提高。

4. 建立模型与预测精度研究

4.1. 模型一(非平稳时间序列模型)

4.1.1. 数据预处理

首先取 58 台 5 月 30 号的 96 个样本数据序列 $\{x_t\}$ 得到如下图 1。

由图 1 表明：该样本构成的时间序列为非平稳时间序列。

由此对该样本值进行有序差分变换

$$\text{差分算子 } \nabla = 1 - B$$

对 96 个样本值进行一阶差分可得到如下 $\{Y_t\}$ 序列图 2。

有图 2 表明：该序列已平稳，则原时间序列可表示为

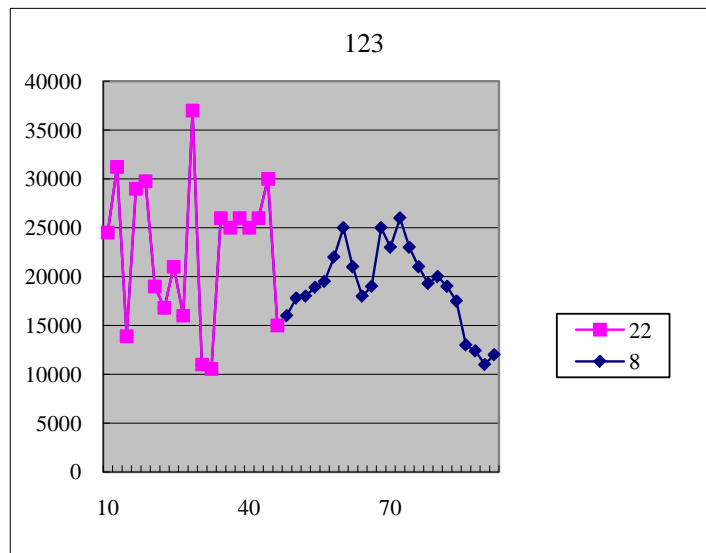


Figure 1. Sequence curve

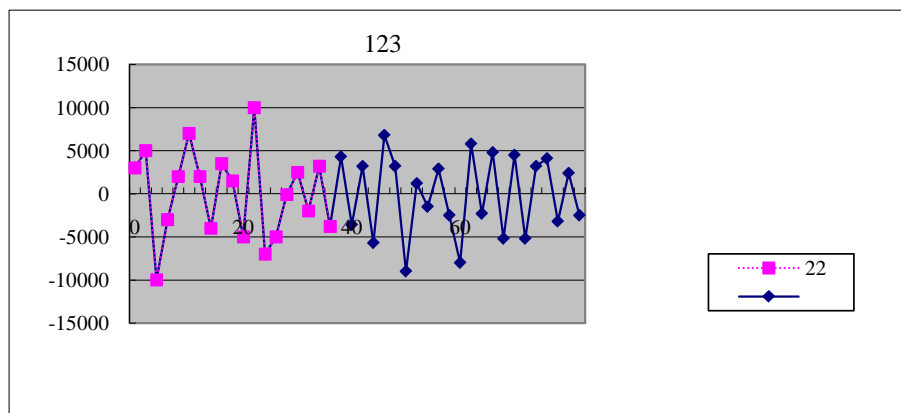
图 1. $\{x_t\}$ 序列曲线

Figure 2. Data difference processing sequence curve

图 2. 数据差分处理序列曲线

$$\phi(B)\nabla^d y_t = \theta(B)a_t$$

即自回归 - 滑动平均模型 $ARIMA(p, d, q)$ 。

4.1.2. 平稳随机时间序列模型的识别

计算偏自相关系数及自相关系数

自相关函数 $\check{\rho}_k$

$$\check{\rho}_k = \frac{\check{\gamma}_k}{\gamma_0}$$

$$\check{\gamma}_k = \frac{1}{N} \sum_{t=1}^{N-k} y_t y_{t+k}$$

偏自相关函数 ϕ_{kk}

$$d = E \left[\left(y_t - \sum_{i=1}^k \alpha_{ki} y_{t-i} \right)^2 \right]$$

当 $l=k$ 时, 第 k 系数 α_{kk} 为平稳序列 $\{Y_t\}$ 的偏自相关系数。
运用统计学软件我们得到该平稳序列 $\{Y_t\}$ PACF、ACF。

4.1.3. 模型识别与模型参数估计

1) 模型识别

我们运用经典的 Box-Jenkins 模型识别方法。

对于 AR 模型, 其偏自相关函数满足下式

$$\begin{cases} \alpha_{kj} = \varphi_j, \square 1 \leq j \leq p \\ \alpha_{kj} = 0, \square j = p+1, p+2, \dots, k \end{cases}$$

对于 MA 模型, 其自相关函数满足下式

$$\rho_k = \begin{cases} 1, & k=0 \\ \frac{-\theta_k + \theta_{k+1}\theta_1 + \dots + \theta_q\theta_{q-k}}{1 + \theta_1^2 + \dots + \theta_q^2}, & 1 \leq k \leq q \\ 0, & k > q \end{cases}$$

对于 ARMA 模型, 同时具备 AR 和 MA 模型的特征[4]。

由此我们给出三种模型的基本特征如表 1。

由上面 φ_{kk} 、 ρ_k 统计特性, 我们可以判断该模型属于 ARMA(p, q)。

2) 模型参数估计

由 1) 我们得到该模型属于 ARMA(p, q) 模型, 因此我们对该模型进行参数估计, 由该模型可推得

$$\begin{bmatrix} \widehat{\varphi}_1 \\ \widehat{\varphi}_1 \\ \vdots \\ \widehat{\varphi}_p \end{bmatrix} = \begin{bmatrix} \widehat{\gamma}_q & \widehat{\gamma}_{q-1} & \widehat{\gamma}_{q-2} & \cdots & \widehat{\gamma}_{q-p+1} \\ \widehat{\gamma}_{q+1} & \widehat{\gamma}_q & \widehat{\gamma}_{q-1} & \cdots & \widehat{\gamma}_{q-p+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \widehat{\gamma}_{q+p-1} & \widehat{\gamma}_{q+p-2} & \widehat{\gamma}_{q+p-2} & \cdots & \widehat{\gamma}_q \end{bmatrix}^{-1} \begin{bmatrix} \widehat{\gamma}_{q+1} \\ \widehat{\gamma}_{q+2} \\ \vdots \\ \widehat{\gamma}_{q=p} \end{bmatrix}$$

我们令 $\bar{y}_t = y_t - \sum_{j=1}^p \varphi_j y_{t-j}$, 则 $\bar{y}_t = a_t - \sum_{i=1}^q \theta_i y_{t-i}$, 这样就将原 ARMA(p, q) 变成 MA(q) 模型, 我们根据

MA(q) 的参数估计对 $\hat{\sigma}^2$ 、 $\hat{\theta}_1$, \dots , $\hat{\theta}_q$ 进行估计。

若 MA(q) 的阶数较低我们可直接求解。

若 MA(q) 的阶数较高, 可运用线性迭代法求解[5]。

Table 1. Characteristics of three basic models
表 1. 三种基本模型特征

类别	模型名称		
	AR(p)	MA(q)	ARMA(p, q)
自相关函数	拖尾	截尾	拖尾
偏自相关函数	截尾	拖尾	拖尾

4.1.4. 模型定阶

根据 AIC 最小信息准则法进行模型定阶，经过逐步的模型拟合，矩估计模型参数估计，我们最终得到当模型的阶数为 AR(6) 模型，拟合效果达到最优。

由此我们确定模型为 ARMA(6,1)。

4.1.5. 建立预测模型

对于 ARMA(6,1,0) 模型，由于样本个数 $h > p$ ，因此残差项 $a_t = 0$ ，这里运用统计 spss 软件可求得则预测方程为

$$\begin{aligned}\hat{y} = & 19930.39216 + AR(1) = 0.6752061645\hat{y}_{t-1} + AR(2) = 0.1657058252\hat{y}_{t-2} \\ & + AR(3) = -0.04506977184\hat{y}_{t-3} + AR(4) = -0.1007669642\hat{y}_{t-4} \\ & + AR(5) = 0.00320666507\hat{y}_{t-5} + AR(6) = 0.03881877566\hat{y}_{t-6}\end{aligned}$$

4.1.6. 实时预测

运用预测方程对 5 月 31 号的数据进行预测得到如下曲线图 5。

从图 3 中表明：基本上符合了 58 台机组 5 月 31 号的输出电功率的趋势。但预测明显存在延时性，以及预测精度不高，现在我们对其实进行误差分析[6]。

4.1.7. 误差分析

通常用三个指标来对模型进行评估：

均方根误差：

$$e_{MSE} = \sqrt{\frac{1}{n+1} \sum_{t=s}^{s+n} (\hat{x}_t - x_t)^2}$$

平均绝对误差：

$$e_{MAE} = \frac{1}{n+1} \sum_{t=s}^{s+n} |\hat{x}_t - x_t|$$

平均相对误差：

$$e_{MAE} = \frac{1}{n+1} \sum_{t=s}^{s+n} \left| \frac{\hat{x}_t - x_t}{x_t} \right|$$

在这里我们仅运用相对误差进行分析如下。

将数据代入公式我们得到表 2。

为了减小误差我们将采用组合预测的方法使误差达到最低。

4.2. 模型二(人工神经网络模型 RBF)

4.2.1. 构造神经网络结构图

一般的我们选取隐层为 3 的结构图，它能更好地对误差进行反向修正。由此我们构造如图 4。

4.2.2. 输入变量的选择

我们利用模型一建立好的时间序列模型选择输入变量。

在模型一中表明：原始的电功率序列 $\{X_t\}$ 进行 1 阶差分处理的序列 $\{Y_t\}$ 可识别为 ARMA(6,1) 模型。由此我们选择输入的变量为最近的 6 个历史数据及最近的一个残差 a_t 为组。

Table 2. Error analysis of time series model
表 2. 时间序列模型误差分析

P58 电机组	真实值	预测值	绝对误差	相对误差
1	11634.84	10534.83	1100	0.094544
2	16696.88	10686.88	6010	0.359947
3	18568.59	12448.59	6120	0.329589
4	16420.41	11420.11	5000	0.304499
5	18146.06	18100.06	46	0.002535

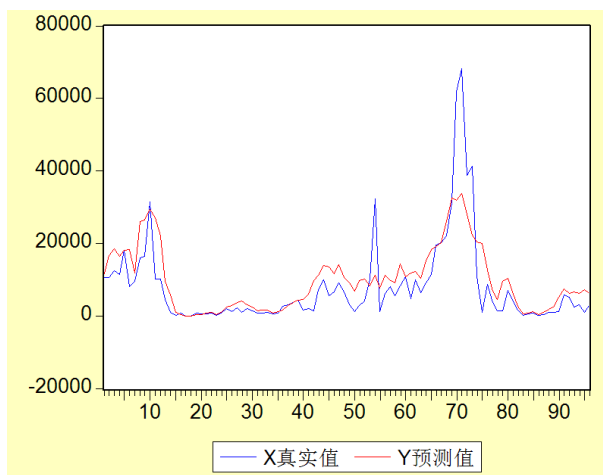


Figure 3. Comparison curve of prediction data and real value
图 3. 预测数据与真实值的比较曲线图

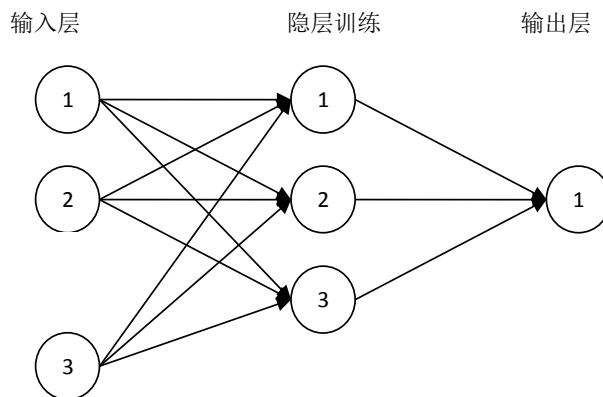


Figure 4. Neural network structure diagram
图 4. 神经网络结构图

4.2.3. 网络输入输出数据的归一化处理

对输入输出的数据 x 采用以下公式

$$y_i = \frac{x_i - \frac{1}{2}(x_{\min} - x_{\max})}{\frac{1}{2}(x_{\max} - x_{\min})}$$

$$x_i = \frac{1}{2}(x_{\max} - x_{\min})y_i + \frac{1}{2}(x_{\max} + x_{\min})$$

进行归一化处理使其落在[-1,1]区间。

4.2.4. 网络训练及预测结果输出

对输入选择的变量进行网络函数训练。

第一步我们对输入的变量从输出层得到输出值 $h_j(x)$

$$h_j(x) = \phi(\|x - c_j\|, \sigma_j) = \exp\left[-\frac{\|x - c_j\|^2}{2\sigma_j^2}\right]$$

第二部从隐含层输出函数为

$$f(x) = \sum_{j=1}^3 h_j(x)w_j$$

由此我们根据 LAMP 算法可得到如下曲线图 5。

4.2.5. 模型误差分析

我们同样取 P58 台发电机组的 5 个数据进行误差分析得到如下表 3。

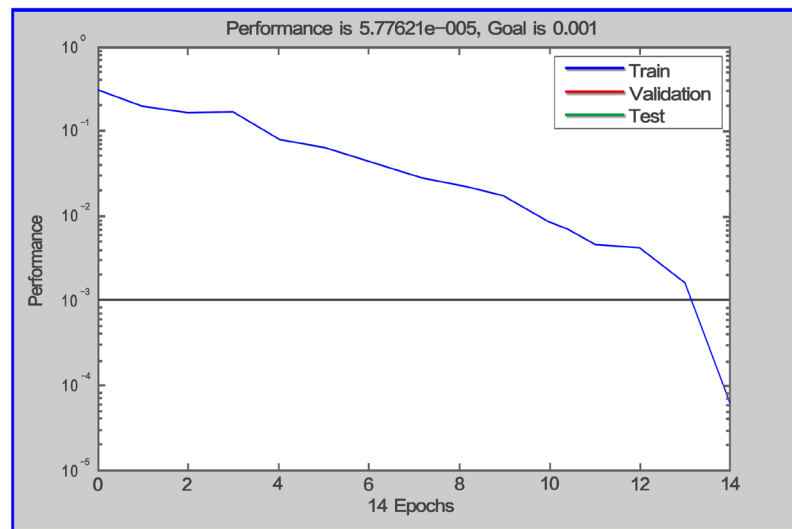


Figure 5. Predicted value curve graph

图 5. 预测值曲线图

Table 3. Error analysis of neural network

表 3. 神经网络误差分析

P58	相对误差
1	0.26102
2	0.126504
3	0.113031
4	0.069922
5	0.0538867

4.3. 模型三

灰色预测矩阵

同理我们仅取 P58 台机组的 6 个历史数据运用灰色预测法进行预测。

灰色预测基本原理如下：

首先，进行一次累加

$$x^{(1)} = [x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)]$$

$$x^{(1)}(i) = \left\{ \sum_{j=1}^i x^{(0)}(j) \mid i=1, 2, \dots \right\}$$

其中，显然有 $x^{(1)}(1) = x^{(0)}(1)$ ， $x^{(1)}$ 表示累加后的数列[7]。

然后，参数估计 $A = (B^T B)^{-1} B^T Y_n = \begin{bmatrix} \hat{a} \\ \hat{u} \end{bmatrix}$

$$B = \begin{pmatrix} -\frac{1}{2}[x^{(1)}(1) + x^{(1)}(2) \cdots 1] \\ -\frac{1}{2}[x^{(1)}(2) + x^{(1)}(3) \cdots 1] \\ \vdots \\ -\frac{1}{2}[x^{(1)}(n-1) + x^{(1)}(n) \cdots 1] \end{pmatrix} Y_n = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}$$

其次，累加数列 $x^{(1)}$ 的灰色预测模型

$$\hat{x}^{(1)}(k+1) = \left[x^{(0)}(1) - \frac{\hat{u}}{\hat{a}} \right] e^{-\hat{a}(k+1)} + \frac{\hat{u}}{\hat{a}} \quad (k=0, 1, 2, \dots)$$

由此，求原始数列的灰色预测模型

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) = (1 - e^{\hat{a}}) \left(\hat{x}^{(0)}(1) - \frac{\hat{u}}{\hat{a}} \right) e^{-\hat{a}k} \quad (k=1, 2, 3, \dots)$$

最后，进行模型精度的方差检验

历史数据残差为：

$$E(k) = x^{(0)}(k) - \hat{x}^{(0)}(k)$$

相对残差为：

$$e(k) = \left[x^{(0)}(k) - \hat{x}^{(0)}(k) \right] / x^{(0)}(k)$$

预测结果：

对模型的预测结果。

得到真实观测与拟合曲线的对比基本重合。

同理我们对模型进行误差分析得到如下表 4。

1) 问题二

我们分别运用三种模型对 PA、PB、PC、PD、P4、P58 进行误差分析，我们仅取相对误差进行分析如下表 5。

表 5 表明：多机组的误差总体上小于单机组的误差。

Table 4. Error analysis of grey model prediction
表 4. 灰色模型预测误差分析

P58 电机组	真实值	模拟值	绝对误差	相对误差
1	28560.56	3.04E + 04	1805.44	0.06
2	17980.03	2.18E + 04	3855.97	0.21
3	19235.06	1.76E + 04	1605.06	0.08
4	21874.5	1.88E + 04	3072.50	0.14
5	29787.38	2.84E + 04	1416.38	0.05

Table 5. Relative error analysis of PA, PB, PC, PD, P4 and P58
表 5. PA、PB、PC、PD、P4、P58 相对误差分析

	相对误差分析					
	PA	PB	PC	PD	P4	P58
模型一	0.023	0.0156	0.142	0.0245	0.0206	0.013
模型二	0.0102	0.1203	0.106	0.102	0.0135	0.0112
模型三	0.125	0.156	0.179	0.194	0.113	0.06

因此我们期望为了尽可能准确预测，尽量使多台电机共同运作，以使能够更好的对未来发电输出功率进行预测。

2) 问题三

为了进一步提高风电功率的预测精度，我们建立了组合预测模型其表达式如下。

设三种模型分别表示为 P_1 、 P_2 、 P_3 则组合预测模型为

$$P = \lambda_1 P_1 + \lambda_2 P_2 + \lambda_3 P_3$$

$$\lambda_1 + \lambda_2 + \lambda_3 = 1$$

3 种预测模型的方差、误差分别为 σ_1 、 σ_2 、 σ_3 、 e_1 、 e_2 、 e_3 则

组合预测误差的方差为 $\text{var}(e_c) = \sum_{i=1}^3 \lambda_i^2 \sigma_{ii}$ 。

当三种预测方法的预测误差分别服从零均值正太分布时，可用以下式估计

$$\sigma_{ii} = \frac{1}{n} \sum_{t=1}^n e_{it}^2 \quad (i=1,2,3)$$

由此我们可得出 $\lambda_i (i=1,2,3)$ 的估计值为

$$\hat{\lambda}_i = \left(\sum_{t=1}^n e_{it}^2 \right)^{-1} \left[\sum_{j=1}^3 \left(\sum_{t=1}^n e_{jt}^2 \right)^{-1} \right]^{-1}$$

由此证明 $\text{var}(e_c)_{\min} \leq \min(\sigma_{11}, \sigma_{12}, \sigma_{13})$ ，表明组合预测的方法优于单一预测方法[8]。

5. 结论

1) 模型一的评价

优点：对于短、近期预测比较显著。

缺点：延伸到更远的将来，就会出现很大的局限性，导致预测值偏离实际较大而使决策失误。

不适用于对长远数据的预测。

2) 模型二的评价

优点: BP 神经网络是一种有效的非线性建模方法; 具有很强的容错性和很快的处理数据能力。

缺点: 网络结构的选择尚无一种统一而完整的理论指导, 理解起来比较费劲, 对模型学习速度慢。

3) 模型三的评价

优点: 算法简单、可利用较少数据建模。

缺点: 对于复杂的非线性系统来说预测的效果不是很理想。

4) 总结:

对于风电机组群问题我们同样采用问题三模型, 即 P_n 时我们采用组合预测模型。这样可以尽量提高预测精度, 但是不可能得到无限提高。

基金项目

2013 湖南省教育厅课题 13C862。

参考文献 (References)

- [1] 张新房, 徐大平, 吕跃刚, 柳亦兵. 风力发电技术的发展及若干问题[J]. 现代电力, 2003, 20(5).
- [2] 雷亚洲, 王伟胜, 任永华, 等. 含风电场电力系统的有功有化潮流[J]. 电网技术, 2002, 26(6): 18-21.
- [3] 楼顺天, 施阳. 基于 MATLAB 的系统分析与设计——神经网络[M]. 西安: 电子科技大学出版社, 2000.
- [4] 董安正, 赵国潘. 人工神经网络在短期资料风速估计方面的应用[J]. 工程力学, 2003, 20(5): 10-13.
- [5] 潘迪夫, 刘辉, 李燕飞. 给予时间序列分析和卡拉曼滤波法的风电场风速预测优化模型[J]. 电网技术, 2008, 32(7): 82-86.
- [6] 丁明, 张立军, 吴仪纯. 给予时间序列分析的风电场风俗预测模型[J]. 电力自动化设备, 2005, 25(8): 32-34.
- [7] 刘玉. 基于实测数据分析的大型风电场风电功率预测研究[J]. 黑龙江电力, 2011, 33(1): 11-15.
- [8] 刘纯, 范高峰, 王伟胜, 戴慧珠. 风电场输出功率的组合预测模型[J]. 电网技术, 2009(13): 74-79.

期刊投稿者将享受如下服务:

1. 投稿前咨询服务 (QQ、微信、邮箱皆可)
2. 为您匹配最合适的期刊
3. 24 小时以内解答您的所有疑问
4. 友好的在线投稿界面
5. 专业的同行评审
6. 知网检索
7. 全网络覆盖式推广您的研究

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: sa@hanspub.org