

# 基于统计聚类分析的短期风电功率预测

方江晓<sup>1</sup>, 周 晖<sup>1,2</sup>, 黄 梅<sup>1</sup>, T.S.Sidhu<sup>2</sup>

(1. 北京交通大学电气工程学院, 北京 100044;

2. Department of Electric Power Engineering, University of Western Ontario, London, Canada, N6A5B9)

**摘要:** 考虑到短期风电功率预测模型建立时, 样本的选取对预测模型的精度有较大影响, 提出了运用聚类方法对历史风速数据进行处理, 实现了历史数据的自动分类。根据预测日的平均风速和最大风速等特征参数, 按照相似度最大的原则, 选择合适的类别作为预测建模用的训练样本。运用时间序列方法, 建立风速预测模型, 与未经过预处理的相比, 所建立预测精度得到了提高, 验证了运用聚类进行数据预处理的正确性。运用风力发电机的出力曲线, 得到了未来日的风电功率的预测值, 为含风电系统的电力系统运行计划的制定, 提供了基础数据支持。

**关键词:** 风电预测; 聚类分析; 最大相似度; 时间序列模型

## Short-term wind power prediction based on statistical clustering analysis

FANG Jiang-xiao<sup>1</sup>, ZHOU Hui<sup>1,2</sup>, HUANG Mei<sup>1</sup>, T.S.Sidhu<sup>2</sup>

(1. School of Electrical Engineering, Beijing Jiaotong University, Beijing 100044, China;

2. Department of Electric Power Engineering, University of Western Ontario, London, Canada, N6A5B9)

**Abstract:** Considering that in the process of establishing short-term wind power prediction model, the sample selection would affect the predication accuracy of wind power model, it is necessary to process the history wind speed data prior to modeling. Data classification is automatically accomplished through the statistical clustering approach. With the criterion of maximal similarity, we select a group of data as our trained samples according to the average and maximum wind speed of prediction day. Then we establish the prediction model of wind speed based on ARIMA process. Compared with the conventional ARIMA process, the prediction accuracy using statistic clustering approach we proposed is improved. An example is used to verify the correctness of our assumption. Finally, with power curve of a wind turbine, anticipated wind power is easily gotten, which offers valuable reference for drawing out operation schedule of power system integrated with wind power.

**Key words:** prediction of wind power; statistical clustering; maximal similarity; ARIMA process

中图分类号: TM715 文献标识码: A 文章编号: 1674-3415(2011)11-0067-07

## 0 引言

风电预测, 是风电接入研究的一个关键问题。提高风电预测的准确性, 可以减少含风电接入电力系统调度计划的不确定性, 这样既提高了系统运行的安全性, 也提高了经济性<sup>[1]</sup>。

目前风电功率的预测, 在国内外都是一个热点研究问题, 现已有成熟的风电功率预测系统应用于风电场中, 风电预测系统多采用组合预测模型来预测功率, 其预测效果较好, 但依赖于单一预测模型的精度和组合的算法, 因此单一预测方法依然是研究的基础和重点。其中时间序列法, 卡尔曼滤波法 (Kalman-filter), 人工神经网络法 (ANN), 小波

分析法 (Wavelet analysis), 以及模糊预测方法 (如 ANFIS), 支持向量机 (SVM) 等均在风电功率预测中有所应用, 其预测的相对误差在 25%~40% 水平<sup>[2-7]</sup>。由于 ARIMA 时间序列方法, 可以针对不平稳时间序列进行建模, 且只需知道风电场单一的风速或功率数据, 简单易行且预测效果较好, 已被广泛用于风速和发电功率的预测中。

从时间序列预测的原理来看, 它是基于历史规律的总结与归纳, 通过寻找其变化规律的数学表达式, 对未来的变化趋势进行外推<sup>[8]</sup>。因此历史数据对预测模型的性能有很大的影响。在这种情况下, 根据预测目标的特征, 合理选择训练样本, 成为提高预测模型精度的有效方法之一。如在负荷预测

中,运用该预测思想,考虑了相似日的预测模型不乏人在<sup>[9]</sup>。

基于以上思路,本文提出了运用聚类分析方法<sup>[10]</sup>,对历史风速数据进行自动分类。并借助于预测日已有的风速特征数据,按照相似度最大的原则,选择与预测日相似度最大的那组数据,作为预测建模用的训练样本。并通过实际算例的计算,说明采用所提出的数据预处理方法后,预测模型的预测精度得到了提高。

最后借助于风力机组的风速—功率曲线,非常方便地得到了制定电力系统运行计划所需要的风电机组的预测出力。

## 1 基于时间序列法的风速预测

### 1.1 时间序列法建模

时间序列模型<sup>[11]</sup>可分为自回归(AR)模型、滑动平均(MA)模型、自回归滑动平均(ARMA)模型和自回归积分滑动(ARIMA)模型等。ARMA模型是AR模型和MA模型的结合,ARIMA模型可以通过差分后转化为ARMA模型来处理。因此,可以说AR、MA和ARMA模型都是ARIMA模型的特例。

#### 1) 模型表述

若时间序列 $\{Y_t\}$ 是它的当期和前期的随机干扰项以及前期值的线性函数,可表示为:

$$Y_t = \sum_{i=1}^p \phi_i Y_{t-i} - \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t \quad (1)$$

则称式(1)为时间序列 $\{Y_t\}$ 服从 $(p, q)$ 阶的自回归滑动平均模型,简记为ARMA $(p, q)$ 。式中, $\phi_1, \phi_2, \dots, \phi_p$ 为自回归系数; $\theta_1, \theta_2, \dots, \theta_q$ 为滑动平均系数; $\varepsilon_t$ 为随机误差项,是一相互独立的白噪声序列。

若 $\{Y_t\}$ 是一非平稳序列,经过 $d$ 阶差分后, $\{\nabla^d Y_t\}$ 是一平稳的ARMA序列,记为式(2)。

$$\phi_p(B) \nabla^d Y_t = \theta_q(B) \varepsilon_t \quad (2)$$

则式(2)为 $(p, d, q)$ 阶的自回归积分滑动平均模型,简记为ARIMA $(p, d, q)$ 。

式中:

$$\phi_p(B) = 1 - \phi_1 B^1 - \phi_2 B^2 - \dots - \phi_p B^p$$

$$\theta_q(B) = 1 - \theta_1 B^1 - \theta_2 B^2 - \dots - \theta_q B^q$$

其中, $B$ 为滞后算子。

对于ARMA $(p, q)$ ,如果阶数 $q=0$ ,则是自回归模型AR $(p)$ ;如果阶数 $p=0$ ,则成为移动平均模型MA $(q)$ 。

### 2) 风速序列的平稳性检验及其平稳化

上述模型的应用都是以时间序列为平稳随机序列为前提的。因此在构建模型之前,需要对序列进行平稳性检验,常用的检验方法是扩展的Dickey-fuller(简称ADF检验)。ADF检验法需要进行如式(3)回归。

$$\nabla Y_t = \beta_0 + \lambda T + (\rho - 1)Y_{t-1} + \sum_{i=1}^p \gamma_i Y_{t-i} + \varepsilon_t \quad (3)$$

式中: $\beta_0$ 为截距项; $T$ 为趋势项; $\nabla$ 为一阶差分; $\varepsilon_t$ 为随机误差项;滞后项 $p$ 可按照赤池准则(AIC)来确定。对于ADF检验而言,

原假设为 $H_0: \rho = 1$ ;

备择假设为 $H_1: \rho < 1$ 。

若ADF值大于临界值,则接受 $H_0$ ,意味着变量时间序列 $Y_t$ 含有一个单位根,即该时间序列不是平稳的;反之,若ADF小于临界值,则拒绝 $H_0$ ,接受 $H_1$ ,表明该时间序列是平稳的。

对于不平稳的风速序列,可经过适当次差分将非平稳的数据序列转化为均值为零的平稳随机序列,同时经差分平稳化处理后可将ARIMA模型简化为相应的AR、MA和ARMA模型。

### 3) 模型的识别与定阶

对平稳的风速时间序列数据,可借助序列的自相关函数AC和偏相关函数PAC来识别模型。

$$AC: \hat{\rho}_k = \frac{\sum_{i=1}^{n-k} Y_i \cdot Y_{i+k}}{\sum_{i=1}^n Y_i^2}, k=1, 2, \dots \quad (4)$$

$$PAC: \hat{\phi}_{kk} = \begin{cases} \hat{\rho}_1, & k=1 \\ \hat{\rho}_k - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} \cdot \hat{\rho}_{k-j}, & k=2, 3, \dots \\ \frac{\hat{\rho}_k - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} \cdot \hat{\rho}_{k-j}}{1 - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} \cdot \hat{\rho}_j}, & k=2, 3, \dots \end{cases} \quad (5)$$

可通过判断AC和PAC的截尾性来初步确定模型的阶数,如果偏相关函数在 $p$ 步截尾,则模型为AR $(p)$ ;如果自相关函数在 $q$ 步截尾,则模型为MA $(q)$ ;如果自相关函数和偏相关函数均拖尾,则认为模型为ARMA $(p, q)$ 。

确定模型后,可借助Eviews、SPSS、SAS等软件,对模型进行定阶。使用软件工具可避免繁杂的数据计算,由软件计算出不同 $(p, q)$ 组合(要求 $0 \leq p \leq \sqrt{n}$ ,  $0 \leq q \leq \sqrt{n}$ )的SC(Schwarz criterion)、AIC(Akaike information criterion)值和调整 $R^2$ 值,选择能使得SC或AIC值最小及调整 $R^2$ 值最大的阶数,将其定为模型的阶数。

## 4) 模型的参数估计和检验

确定模型阶数后,可采用最小二乘法来估计模型中的参数,其目标是使得式(6)残差的平方和达到最小值。

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (\theta_q^{-1}(B)\phi_p(B)\nabla^d Y_i)^2 \quad (6)$$

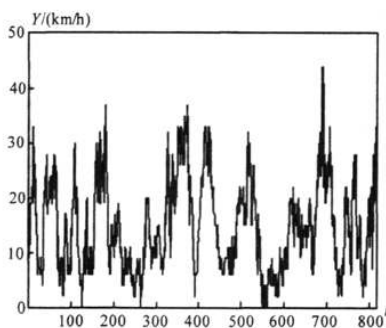
最后,检验残差序列  $\hat{\varepsilon}_i = Y_i - \hat{Y}_i$  ( $\hat{Y}_i$  为预测值) 是否为白噪声的一个样本序列。若是,则所建模型合适;否则需对所建模型进行修改,直到通过检验为止。

## 1.2 算例分析

选取加拿大安大略省伦敦市的一组从 2004 年 10 月 2 日至 11 月 4 日共 34 天的历史风速数据进行分析<sup>[12]</sup>,其中 11 月 4 日作为预测日。数据源的序列为每小时采样一个点,因此共有 792 个风速值作为建模用。然后用所建的模型对 11 月 4 日的风速进行预测,之后再与 11 月 4 日的实际值进行对比,分析预测的误差。分析所采用的建模分析软件是 Eviews5.0<sup>[13]</sup>。

## 1) 数据分析

如前所述,建模前需对风速数据序列的平稳性进行检验,首先计算自相关函数和偏相关函数。图 1 为风速时间序列  $\{Y_t\}$ ,图 2 为  $\{Y_t\}$  的相关函数计算结果,其中 AC 表示自相关函数, PAC 表示偏相关函数。

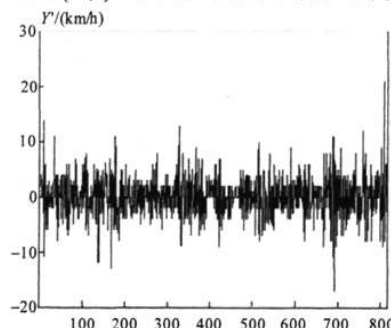
图 1 风速时间序列  $\{Y_t\}$ Fig.1 Wind speed sequences  $\{Y_t\}$ 

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.895	0.895	656.24	0.000	
2	0.823	0.110	1211.9	0.000	
3	0.775	0.106	1705.2	0.000	
4	0.718	-0.035	2128.4	0.000	
5	0.661	-0.023	2487.6	0.000	
6	0.608	-0.017	2792.2	0.000	
7	0.552	-0.040	3043.9	0.000	
8	0.506	0.013	3255.6	0.000	
9	0.459	-0.027	3429.5	0.000	
10	0.421	0.030	3576.1	0.000	

图 2 序列  $\{Y_t\}$  的相关图Fig.2 Correlation function of  $\{Y_t\}$ 

由图 2 可以看出,自相关系数不能很快地落入随机区内(即趋于 0),因此该时间序列是非平稳的。

通过差分的方式,将非平稳的时间序列转化为平稳的时间序列。图 3 为对原序列进行一阶差分后得到的序列  $\{Y_t'\}$ ,其对应的相关计算结果见图 4。

图 3 一阶差分后的风速时间序列  $\{Y_t'\}$ Fig.3 First-order difference sequences of  $\{Y_t'\}$ 

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	-0.167	-0.167	22.694	0.000	
2	-0.088	-0.119	29.035	0.000	
3	0.044	0.008	30.602	0.000	
4	0.006	0.005	30.630	0.000	
5	-0.019	-0.011	30.918	0.000	
6	0.010	0.006	31.001	0.000	
7	-0.038	-0.041	32.217	0.000	
8	-0.004	-0.016	32.228	0.000	
9	-0.037	-0.051	33.359	0.000	
10	-0.037	-0.055	34.465	0.000	

图 4 序列  $\{Y_t'\}$  的相关图Fig.4 Correlation function of  $\{Y_t'\}$ 

由图 4 可以看出,序列的自相关系数能较快地落入到随机区域内,由此初步判断该序列是平稳的。

为了增加判断的可靠性,对一阶差分序列作 ADF 检验,其结果见表 1。

表 1 一阶差分序列 ADF 检验结果

Tab.1 ADF test referred to first-order difference series

检验统计量	1%临界值	5%临界值	10%临界值
-24.589 6	-3.438 2	-2.864 9	-2.568 6

由表 1 可以看出 ADF 的检验统计量等于 -24.589 6,小于  $\alpha=1\%$ , 5%和 10%的临界值,说明在 95%置信水平下完全有理由拒绝原假设,即一阶差分后的风速数据序列是平稳的。

## 2) 模型建立

对时间序列进行以上平稳处理后,即可在平稳差分序列的基础上进行时间序列的建模。由图 2 可知,该序列的自相关函数是拖尾的,而偏相关函数具有明显的截尾特点,因此可对一阶差分后的时间序列建立 AR(p)模型。

AR(p)模型的阶数 p 是由显著不为 0 的偏相关

系数的数目决定的。观察图 2，偏相关函数在  $k=1$  显著不为 0， $k=2$  和  $k=3$  处的值处于显著和不显著之间， $k>3$  时显著水平很低，因此初步选择模型为 AR (2) 或 AR (3)。

通过 Eviews 软件计算出 AR (2) 和 AR (3) 的 AIC 值、SC 值和调整  $R^2$  值。为增加对比性，把 AR (1) 和 AR (4) 也加入到比较中，见表 2。

表 2 AR 模型的参数值比较

Tab.2 Parameters comparison referred to AR models			
模型	AIC 值	SC 值	调整 $R^2$ 值
AR (1)	5.369 0	5.374 9	0.032 0
AR (2)	5.361 7	5.373 6	0.040 0
AR (3)	5.361 9	5.379 7	0.040 7
AR (4)	5.365 5	5.389 2	0.039 7

由表 2 的比较结果可看出，模型 AR (2) 的 AIC 值和 SC 值最小，其调整  $R^2$  值较大。故选择 AR (2) 模型作为该风速数据的时间序列模型。

由 Eviews 软件对模型 AR (2) 进行参数估计和检验，其结果如图 5。

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.196861	0.035454	-5.552564	0.0000
AR(2)	-0.095401	0.035454	-2.690831	0.0073
R-squared	0.041184	Mean dependent var	-0.011407	
Adjusted R-squared	0.039966	S.D. dependent var	3.600426	
S.E. of regression	3.527745	Akaike info criterion	5.361727	
Sum squared resid	9794.206	Schwarz criterion	5.373566	
Log likelihood	-2113.201	Durbin-Watson stat	1.995309	

图 5 AR (2) 模型参数估计与检验结果

Fig.5 Estimated parameters and its test for model AR (2)

由于  $t$  检验的相伴概率值均小于 0.05，表明在 95% 的置信区间内各项系数值均可接受。故 AR (2) 模型方程为式 (7)。

$$\hat{Y}_t' = -0.1969Y_{t-1}' - 0.0954Y_{t-2}' \quad (7)$$

经反差分还原得预测值， $\hat{Y}_t = Y_{t-1} + \hat{Y}_t'$ 。

3) 模型检验和预测

对 AR (2) 模型的合适性进行检验，对模型残差序列  $\varepsilon_t$  进行白噪声检验如图 6。

由图 6 可以看出残差序列的样本自相关系数基本都落入随机区间，近似都等于零，表明残差序列是纯随机的，模型通过检验，可以用来预测。

衡量预测结果时，使用平均相对误差  $\delta$ ，其定义如式 (8)。

$$\delta = \frac{1}{T} \sum_{i=1}^T \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \times 100\% \quad (8)$$

式中： $T$  表示预测值个数， $Y_t$  表示风速实测值； $\hat{Y}_t$  表示风速预测值。

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1	0.007	0.007	0.0418 0.838
		2	-0.029	-0.029	0.7368 0.692
		3	0.014	0.015	0.8997 0.825
		4	0.005	0.004	0.9231 0.921
		5	-0.019	-0.018	1.2121 0.944
		6	-0.003	-0.002	1.2172 0.976
		7	-0.051	-0.052	3.3438 0.851
		8	-0.020	-0.019	3.6605 0.886
		9	-0.048	-0.051	5.5855 0.781
		10	-0.047	-0.047	7.4018 0.687
		11	-0.008	-0.010	7.4561 0.761
		12	0.001	-0.002	7.4577 0.826

图 6 残差检验图

Fig.6 Plot used in residual test

采用 AR (2) 模型，代入  $t=793$  (表示 11 月 4 号 1 点) 于式 (7) 后经反差分，即可得到对应时刻的风速预测值  $\hat{Y}_t$ 。运用递归方法，便可得到待测日逐时的风速预测值，见表 3 (单位：km/h)。

表 3 AR (2) 模型的预测结果

Tab.3 Forecasted results based on AR（2）model					
时序	实际值	预测值	时序	实际值	预测值
0	11.000 0	5.797 1	12	20.000 0	21.606 3
1	11.000 0	9.824 9	13	11.000 0	20.202 9
2	13.000 0	10.523 0	14	7.000 0	12.962 6
3	13.000 0	12.606 3	15	28.000 0	8.646 1
4	17.000 0	12.809 2	16	28.000 0	24.247 5
5	19.000 0	16.212 6	17	20.000 0	25.996 6
6	20.000 0	18.224 7	18	19.000 0	21.574 9
7	19.000 0	19.612 3	19	20.000 0	19.960 1
8	11.000 0	19.101 5	20	26.000 0	19.898 5
9	20.000 0	12.670 3	21	32.000 0	24.723 4
10	20.000 0	18.991 5	22	33.000 0	30.246 4
11	22.000 0	19.141 4	23	28.000 0	32.230 7

由表 3 计算得 11 月 4 日预测风速值的平均相对误差为 26.36%。

2 基于统计聚类分析的时间序列风速预测

2.1 风速数据的聚类分析

1) 聚类的基本原理

聚类分析的方法有很多，其中统计聚类是较为常用的一种<sup>[14]</sup>。它是研究样本之间存在着程度不同的相似性时，根据样本的多个观察指标 (这些指标反映了样本的属性)，找出一些能够度量样本之间相似程度的统计量，以这些统计量作为划分类型的依据，把相似程度较大的样本聚为一类。

某些样本间存在着相似性是聚类的基础，而相

似性度量是对样本分类的依据。相似性度量的方法主要有距离和相似系数两类,距离判据通常用于数值型数据,当距离越接近 0,相似性就越大。

假设每个对象有  $m$  个指标,则欧氏距离可定义如式 (9)。

$$d_{ij} = \left( \sum_{k=1}^m |x_{ik} - x_{jk}|^2 \right)^{\frac{1}{2}} \quad (9)$$

式中,  $x_{ik}$  和  $x_{jk}$  分别表示第  $i$  个样本和第  $j$  个样本的第  $k$  个指标下的值。欧氏距离表示各指标之差的平方和的平方根。

将聚类分析运用于时间序列法风速预测,就是要通过对原数据进行筛选,选取与预测日的风速特征参数具有相似性的历史日的风速数据,将它们作为建模中的样本。

预测日的风速是未知的,但气象部门能够提供次日(即预测日)的平均风速和最大风速的信息。因此,将这两个参数作为预测日的特征参数,用欧氏距离作为相似性度量的方法。

## 2.2 算例分析

### 1) 样本的选取

同样采用 1.2 算例分析中的数据,采用日平均风速和日最大风速作为预测日的风速特征参数,计算出各个历史日和预测日的风速属性值之间的欧氏距离,将其中相似性最高的 7 个历史日及其对应的欧氏距离列在表 4 中。

表 4 欧氏距离计算

Tab.4 Euclidean distance and the corresponding dates							
日期	10月 2日	10月 4日	10月 8日	10月 9日	10月 15日	10月 23日	10月 31日
欧氏距离	2.79	5.02	1.77	4.17	3.44	2.54	1.83

### 2) 数据平稳性检验

用这 7 个历史日共 168 个风速值按时间顺序组成的新样本进行时间序列建模,同样需要进行平稳性检验。记新样本的风速时间序列为  $\{y_{ct}\}$ ,图 7 的序列  $\{y_{ct}\}$  相关系数计算表明,由于  $AC$  不能很快地落入随机区内(即趋于 0),故该时间序列是非平稳的。对  $\{y_{ct}\}$  进行一阶差分后,序列  $\{y'_{ct}\}$  的  $AC$  能较快地落入到随机区域内,见图 8,可初步断定是平稳的。再对  $\{y'_{ct}\}$  进行 ADF 检验,见表 5。根据同样的判据,可断定此刻的  $\{y'_{ct}\}$  是平稳的。

表 5 一阶差分序列 ADF 检验结果

Tab.5 ADF test referred to first-order difference series

检验统计量	1%临界值	5%临界值	10%临界值
-12.406 3	-3.465 0	-2.876 7	-2.574 9

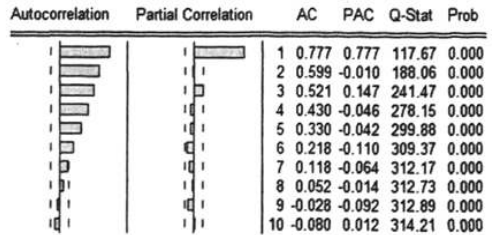


图 7 序列  $\{y_{ct}\}$  的相关图

Fig.7 Correlation function of  $\{y_{ct}\}$

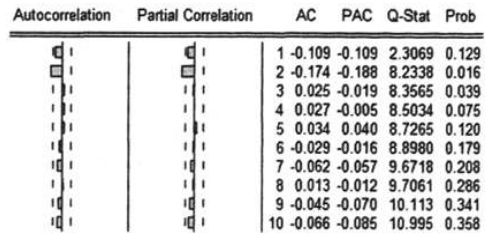


图 8 序列  $\{y'_{ct}\}$  的相关图

Fig.8 Correlation function of  $\{y'_{ct}\}$

### 3) 模型建立

从图 7 看  $AC$  是拖尾的,且  $PAC$  也有一定的拖尾特点。故  $\{y_{ct}\}$  适合  $ARMA(p, q)$  模型。 $ARMA$  模型的精确定阶较之  $AR$  和  $MA$  模型要困难一些,通常采用由低到高的试探法进行,见表 6。根据 1.1 介绍的模型定阶原则,  $ARMA(1, 1)$  为最好的模型。随后由 Eviews 软件对模型  $ARMA(1, 1)$  进行参数估计和检验的结果如图 9 所示。

表 6  $ARMA$  模型的参数值比较

Tab.6 Parameters comparison referred to  $ARMA$  models

模型	AIC 值	SC 值	调整 $R^2$ 值
$ARMA(1,1)$	5.705 9	5.743 4	0.153 9
$ARMA(1,2)$	5.845 5	5.901 8	0.032 9
$ARMA(2,1)$	5.859 0	5.915 5	0.021 1
$ARMA(2,2)$	5.799 2	5.874 5	0.083 5
$ARMA(2,3)$	5.806 0	5.900 2	0.082 6
$ARMA(3,1)$	5.868 1	5.943 7	0.019 4
$ARMA(3,2)$	5.879 1	5.973 6	0.014 5
$ARMA(3,3)$	5.784 8	5.898 2	0.108 4

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.836533	0.037979	22.02632	0.0000
MA(1)	-1.062078	0.042922	-24.74452	0.0000
R-squared	0.159076	Mean dependent var	-0.012048	
Adjusted R-squared	0.153948	S.D. dependent var	4.534030	
S.E. of regression	4.170449	Akaike info criterion	5.705899	
Sum squared resid	2852.394	Schwarz criterion	5.743393	
Log likelihood	-471.5897	Durbin-Watson stat	2.121882	

图 9  $ARMA(1, 1)$  模型参数估计与检验结果

Fig.9 Estimated parameters and its test for  $ARMA(1, 1)$

经  $t$  检验, 模型的各项系数值均通过检验。故 ARMA (1, 1) 模型方程如式 (10) 示。

$$\hat{y}_{ct} = 0.8365 y_{c(t-1)} + \varepsilon_t - 1.0620 \varepsilon_{(t-1)} \quad (10)$$

经反差分还原后得预测值,  $\hat{y}_{ct} = y_{c(t-1)} + \hat{y}_{ct}$ 。

4) 模型检验和预测

对 ARMA (1, 1) 模型应用之前的残差检验表明残差序列是纯随机的, 故模型通过检验, 可以用来预测。残差检验图如图 10 所示。同样地, 依次代入  $t=169, 170\cdots$ , 于式 (10) 中反差分后, 可以得到在基于统计聚类的基础上, 运用 ARMA (1,1) 模型待测日的逐时风速预测值, 见表 7 (单位: km/h)。

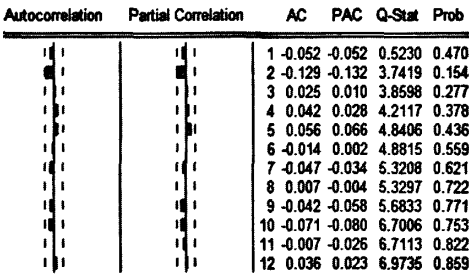


图 10 残差检验图

Fig.10 Plot used in residual test

表 7 基于统计聚类分析后的 ARMA (1, 1) 模型的预测结果

Tab.7 Forecasted results of ARMA (1,1) model based on statistic clustering

时序	实际值	预测值	时序	实际值	预测值
0	11.000 0	10.229 0	12	20.000 0	20.261 4
1	11.000 0	11.854 2	13	11.000 0	18.604 6
2	13.000 0	11.907 2	14	7.000 0	11.547 9
3	13.000 0	13.512 4	15	28.000 0	8.484 1
4	17.000 0	13.544 2	16	28.000 0	24.839 8
5	19.000 0	16.675 8	17	20.000 0	24.643 6
6	20.000 0	18.204 6	18	19.000 0	18.239 6
7	19.000 0	18.929 7	19	20.000 0	17.355 8
8	11.000 0	18.088 8	20	26.000 0	18.028 2
9	20.000 0	11.836 6	21	32.000 0	22.552 5
10	20.000 0	18.858 6	22	33.000 0	26.985 3
11	22.000 0	18.787 8	23	28.000 0	27.448 4

此次预测风速的日平均相对误差为 22.16%, 较之先前的 26.36% 要小。

为了检验以上结论的可靠性, 对其他多处历史时间的风速值分别进行反复的建模预测分析。我们设已做算例为时间段 1, 2004 年的 9 月 1 日至 9 月

30 日为时间段 2, 7 月 1 日至 8 月 1 日为时间段 3, 分别预测最后一日的风速值。此外, 为了与其他方法比较, 选择 BP 神经网络方法预测相同时刻的风速值。设计一个三层 BP 网络, 以预测日前的历史风速值及其次日的风速、最大风速作为输入量, 预测日的风速值作为输出量, 隐层节点数由多次试验确定, 训练 BP 网络进行预测。中间层采用 S 型正切函数 “Tansig”, 输出层采用 S 型对数函数 “Logsig”。并对数据进行归一化处理使输入输出数据在区间 [0,1] 之间。预测结果如表 8 所示。

表 8 三种预测模型的日平均误差

Tab.8 Daily average error of three models

	ARIMA 模型	聚类后的模型	BP 模型
时间段 1	26.36%	22.16%	27.48%
时间段 2	28.12%	23.64%	24.85%
时间段 3	25.63%	22.08%	26.46%

由表 8 可以看出, BP 模型的预测误差和 ARIMA 模型相当, 聚类后的 ARIMA 模型的预测精度较原模型均有一定程度的提高, 实现了预期的目标。

3 风电机组出力的预测

依据风速预测值, 由风电机组功率特性曲线, 或者是风电机组的风速—功率的近似特性曲线  $p_E-v$ , 均可求得对应风速情况下的风力机组的出力 [15]。图 11 为风电机组功率特性曲线的示意图, 其中  $v_{in}$ 、 $v_{out}$  表示切入、切出风速。德国 Dewind 公司制造的单机容量为 1 MW 的变桨距风力发电机组, 其切入风速、额定风速、切出风速分别为 2.5、11.5、23 m/s, 用 km/h 表示为 9, 41.4, 82.8。

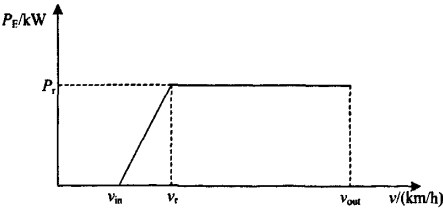


图 11 风电机组功率特性曲线

Fig.11 Power curve of a wind turbine

将表 7 基于统计聚类分析后得到的 24 点风速预测值, 先转换计量单位为 m/s, 根据功率曲线和参数得到相应的风电机组出力预测值, 见表 9。其中, 在第 15 时刻, 由于风速低于切入风速, 达不到开始并网发电所要求的最低风速, 机组不能启动, 输出功率为 0。

表 9 风电机组出力预测值

Tab.9 Forecasted power output of a wind turbine

时序	风速预测值 / (km/h)	功率值 / kW	时序	风速预测值 / (km/h)	功率值 / kW
0	13.194 7	37.931 1	12	20.261 4	347.575 1
1	12.772 2	88.091 7	13	18.604 6	296.438 1
2	12.272 5	89.728 3	14	11.547 9	78.638 2
3	13.275 3	139.272 3	15	8.484 1	0.000 0
4	12.999 0	140.254 1	16	24.839 8	488.881 2
5	9.787 5	236.908 6	17	24.643 6	482.826 1
6	9.214 4	284.093 3	18	18.239 6	285.171 8
7	11.881 8	306.472 3	19	17.355 8	257.895 9
8	9.766 6	280.518 7	20	18.028 2	278.648 6
9	4.280 9	87.549 5	21	22.552 5	418.288 2
10	8.282 2	304.278 7	22	26.985 3	555.100 4
11	11.302 3	302.091 9	23	27.448 4	569.395 0

## 4 结论

通过以上所做的工作,可以得出以下结论:

1) 时间序列预测法对数据要求不高,只需知道风电场的单一风速时间序列即可,简单易行。考虑到训练样本的选择,对预测模型精度的提高有影响,借鉴负荷预测的经验,提出了基于统计聚类分析的时间序列预测法,来进行风电功率的预测。

2) 通过对加拿大安大略省伦敦市的一组风速数据,按照时间序列法,建立基于 ARIMA 模型的预测风速,得出该情况下的日平均预测误差为 26.36%。

3) 利用同一组数据,选择气象部门能够提供的预测日平均风速和最大风速作为聚类分析指标,采用欧氏距离作为确定相似日的依据,筛选出新的训练样本。

4) 运用新样本建立 ARIMA 模型,得出相应的日平均预测误差为 22.16%。这较之常规的时间序列预测法,在预测精度上得到了一定程度的提高。

鉴于风电功率的预测误差还是比较大的,尤其是某些单点的预测值,偏差较大。依赖气象部门能提供更多的预测日的风速特征信息,将更容易掌握风速的变化范围。此外,运用数据挖掘技术,深入挖掘数据序列本身的相似性,将是今后要做的工作。

## 参考文献

[1] Brendan Fox, Damian Flynn, Leslie Bryans, et al. Wind power integration connection and system operational Aspects[M]. UK: The Institution of Engineering and Technology, 2007.

- [2] 杨秀媛,肖洋,陈树勇. 风电场风速和发电功率预测研究[J]. 中国电机工程学报, 2005, 25 (11): 1-5.  
YANG Xiu-yuan, XIAO Yang, CHEN Shu-yong. Wind speed and generated power forecasting in wind farm[J]. Proceedings of the CSEE, 2005, 25 (11): 1-5.
- [3] 丁明,张立军,吴义纯. 基于时间序列分析的风电场风速预测模型[J]. 电力自动化设备, 2005, 25 (8): 32-34.  
DING Ming, ZHANG Li-jun, WU Yi-chun. Wind speed forecast model for wind farms based on time series analysis[J]. Electric Power Automation Equipment, 2005, 25 (8): 32-34.
- [4] Ernst B, Oakleaf B, Ahlstrom M L, et al. Predicting the wind[J]. IEEE Power & Energy Magazine, 2007, 11: 79-89.
- [5] Khan A A, Shahidehpour M. One day ahead wind speed forecasting using wavelets[C]. //2009 IEEE Power Systems Conference and Exposition, 2009.
- [6] 吴兴华,周晖,黄梅. 基于模式识别的风电场风速和发电功率预测[J]. 继电器, 2008, 36 (1): 27-32.  
WU Xing-hua, ZHOU Hui, HUANG Mei. Wind speed and generated power forecasting based on pattern recognition in wind farm[J]. Relay, 2008, 36 (1): 27-32.
- [7] 王丽婕,廖晓钟,高阳,等. 风电场发电功率的建模和预测研究综述[J]. 电力系统保护与控制, 2009, 37 (13): 118-121.  
WANG Li-jie, LIAO Xiao-zhong, GAO Yang, et al. Summarization of modeling and prediction of wind power generation[J]. Power System Protection and Control, 2009, 37 (13): 118-121.
- [8] 牛东晓. 电力负荷预测技术及应用[M]. 北京: 中国电力出版社, 1998.  
NIU Dong-xiao. Electric power forecasting techniques and its application[M]. Beijing: China Electric Power Press, 1998.
- [9] 金海峰,熊信良,吴耀武. 基于相似性原理的短期负荷预测方法[J]. 电力系统自动化, 2001, 25 (23): 45-48.  
JIN Hai-feng, XIONG Xin-gen, WU Yao-wu. Short-term load forecasting based on analogous theory[J]. Automation of Electric Power Systems, 2001, 25 (23): 45-48.
- [10] 陈安,陈宁,周龙骧,等. 数据挖掘技术及应用[M]. 北京: 科学出版社, 2006.  
CHEN An, CHEN Ning, ZHOU Long-rang, et al. Data mining and its application[M]. Beijing: Science Press, 2006.
- [11] 张思成. 金融计量学——时间序列分析视角[M]. 大连: 东北财经大学出版社, 2008.

(下转第78页 continued on page 78)

- [2] 李涛, 杜晓平, 刘焕光. 电容式电压互感器自激法的测试及误差分析[J]. 电力系统保护与控制, 2009, 37 (5): 31-33.  
LI Tao, DU Xiao-ping, LIU Huan-guang. Discussion about self-excited method error on capacitive voltage transformer [J]. Power System Protection and Control, 2009, 37 (5): 31-33.
- [3] 李涛, 杜晓平, 陈瑞林. 电容型电流互感器现场试验 $\tan \delta$ 值异常的分析及解决方法[J]. 电力系统保护与控制, 2009, 37 (20): 120-122.  
LI Tao, DU Xiao-ping, CHEN Rui-lin. Solution to abnormal  $\tan \delta$  value problem in field test of capacitive current transformer[J]. Power System Protection and Control, 2009, 37 (20): 120-122.
- [4] 王笑棠, 李笑倩, 金贵, 等. 一种变频测量电流互感器介损的方法[J]. 电力系统保护与控制, 2010, 38 (5): 68-72.  
WANG Xiao-tang, LI Xiao-qian, JIN Gui, et al. A method of measuring dielectric loss of current transformer by changing frequency [J]. Power System Protection and Control, 2010, 38 (5): 68-72.
- [5] 陈楷, 王卉, 张承学. 介损角测量中非同步采样算法的研究[J]. 继电器, 2004, 32 (14): 5-7.  
CHEN Kai, WANG Hui, ZHANG Cheng-xue. Research on dielectric loss angle measurement in the condition of nonsynchronous sampling[J]. Relay, 2004, 32 (14): 5-7.
- [6] 律方成, 晁红军, 徐志钮, 等. 介质损耗数字化测量方法综述[J]. 华北电力大学学报: 自然科学版, 2008, 35 (6): 21-26.  
Lü Fang-cheng, CHAO Hong-jun, XU Zhi-niu, et al. Overview of digital method for dielectric loss measurement[J]. Journal of North China Electric Power University: Natural Science Edition, 2008, 35 (6): 21-26.
- [7] 赵智大. 高电压技术[M]. 北京: 中国电力出版社, 1998.  
ZHAO Zhi-da. High voltage engineering[M]. Beijing: China Electric Power Press, 1998.
- [8] 陈宝林. 最优化理论与算法[M]. 北京: 清华大学出版社, 2005.  
CHEN Bao lin. Optimized theory and its algorithm[M]. Beijing: Press of Tsinghua University, 2005.
- [9] 谢明, 丁康. 频谱分析的校正方法[J]. 振动工程学报, 1994, 7 (2): 172-179.  
XIE Ming, DING Kang. Rectifying technique of spectrum analysis[J]. Journal of Vibration Engineering, 1994, 7 (2): 172-179.
- [10] Xi Jiangtao, Chicharo J F. A new algorithm for improving the accuracy of periodic signal analysis[J]. IEEE Trans on Instrum Meas, 1996, 45 (4): 827-830.

收稿日期: 2010-03-13; 修回日期: 2010-04-13

作者简介:

徐志钮 (1979-), 男, 博士生, 讲师, 研究方向为电力系统外绝缘、电气设备状态监测与故障诊断、电力系统谐波分析; E-mail: wzcjxx @sohu.com

律方成 (1963-), 男, 博士, 教授, 博导, 主要研究方向为电气设备状态监测与故障诊断和高电压与绝缘技术。

(上接第 73 页 continued from page 73)

- ZHANG Si-cheng. Financial econometrics—viewed from time series[M]. Dalian: Dongbei University of Finance & Economics Press, 2008.
- [12] <http://www.climate.weatheroffice.gc.ca>.
- [13] 于俊年. 计量经济学软件——Eviews使用 [M]. 北京: 对外经贸大学出版社, 2006.  
YU Jun-nian. Users' guideline of Eviews microsoft based on econometrics[M]. Beijing: University of International Business and Economics Press, 2006.
- [14] 胡峰, 彭力. 基于时间序列模型的电价预测方法[J]. 继电器, 2008, 36 (2): 41-46.  
HU Feng, PENG Li. Electricity price forecasting solution based on time series models[J]. Relay, 2008, 36 (2): 41-46.
- [15] 王承煦, 张源. 风力发电[M]. 北京: 中国电力出版社, 2003.  
WANG Cheng-xu, ZHANG Yuan. Wind power generation[M]. Beijing: China Electric Power Press, 2006.

收稿日期: 2010-06-18; 修回日期: 2010-08-13

作者简介:

方江晓 (1986-), 男, 硕士, 主要从事风电预测及风电接入的研究; E-mail: 09122123@bjtu.edu.cn

周 晖 (1964-), 女, 博士, 副教授, 硕士生导师, 主要从事电力负荷预测、电力系统调度与运行, 电力市场, 风电接入等方面的研究;

黄 梅 (1959-), 女, 研究员, 硕士生导师, 主要从事电力系统负荷建模, 继电保护与新能源研究。