

DOI: 10.19783/j.cnki.pspc.181533

基于 NARX 及混沌支持向量机的短期风速预测

李应求¹, 安勃¹, 李恒通²

(1. 长沙理工大学 数学与统计学院, 湖南 长沙 410114; 2. 新加坡国立大学 理学院, 新加坡 肯特岗 119077)

摘要: 风速预测精度的提高, 对降低风力发电成本、合理安排风场选址等方面有着积极作用。使用 DBSCAN 聚类对所有数据进行去噪处理, 选择最合适的风速数据序列进行实证研究。首先, 针对风速数据序列具有混沌性而对预测结果产生影响的问题, 采用 C-C 法确定相空间重构中所需参数。与此同时, 结合混沌理论建立混沌支持向量机模型, 用以预测未来 24 h 的风速值。之后, 将该模型与 EGARCH 模型以及具有外生输入的非线性自回归网络(NARX)模型的预测结果进行对比。最后, 根据各预测模型的 RMSE 和 MAPE 精度对模型预测效果进行评估。结果表明: 基于混沌时间序列的支持向量机模型对 NWTC m2 气象站所在地风速具有最佳预测效果。

关键词: 风速短期预测; 混沌特性; 时间序列; EGARCH; NARX; 支持向量机

Short-term wind speed prediction based on NARX and chaos-support vector machine

LI Yingqiu¹, AN Bo¹, LI Hengtong²

(1. School of Mathematics and Statistics, Changsha University of Science and Technology, Changsha 410114, China;
2. Faculty of Science, National University of Singapore, Kent Ridge 119077, Singapore)

Abstract: The improvement of wind speed prediction accuracy plays a positive role in reducing the cost of wind power generation and arranging the location of wind farms. The DBSCAN clustering method is used to denoise all data and select the most appropriate sequences for empirical research. Firstly, since the chaotic wind speed data sequences would affect the prediction results, C-C method is made to determine the needed parameters in phase space reconstruction. Meanwhile, a model of chaos support vector machine combined chaos theory is established to predict the wind speed value in the coming 24h. Then the model is compared with the EGARCH model and the nonlinear self-regression network with exogenous input (NARX) model in terms of prediction results. Finally, the prediction effect of models is evaluated by the RMSE and MAPE of each prediction model. The results show that the support vector machine model based on the chaotic time series has the best prediction effect on the wind speed of the NWTC m2 weather station.

This work is supported by National Natural Science Foundation of China (No. 11731012 and No. 11571052) and Natural Science Foundation of Hunan Province (No. 2018JJ2417).

Key words: short-term wind speed prediction; chaotic characteristics; time series; EGARCH; NARX; support vector machine

0 引言

风力发电产业发展前景光明, 据国际能源署 (IEA)保守预测, 2050 年风电将占全球电力生产量的 20%左右^[1]。风电价格的降低也是促进产业发展的关键因素, 而提高风速预测的精度有助于降低风

力发电的成本^[2]。经验表明, 预测误差所导致的额外费用高达风力发电总收入的 10%^[3]。

短期风速预测可以指时间尺度从数分钟到数天不等的预测, 定义为未来 1 h 甚至更短时间内的平均风速预测^[4], 主要用于电网有效合理调度和风力发电机组控制。依据风电公司的真实需求选择预测时间尺度是一种合理的方法。选取 7 个有代表性的电力公司, 对风电预测的用途及要求进行了问卷调查, 结果表明 7 个电力公司都认为次日风电预测(从

基金项目: 国家自然科学基金项目资助 (11731012, 11571052); 湖南省自然科学基金项目资助 (2018JJ2417)

当天中午到次日午夜)是一个重要的预测期,大多数的电力公司希望次日风电预测能够以每小时预测值的方式给出,也有部分电力公司可以接受每 3 h 或 6 h 的预测;部分电力公司对提前 3~7 天的风电预测也感兴趣^[5]。对于电力系统而言,由风速预测值转化为风电预测值,是比直接进行风电输出预测更好的方法^[6]。中国许多电力公司认为以小时作为预测风速的时间尺度应用更广^[7]。因此可以认为,大部分电力公司希望得到的是以小时预测值的方式给出的次日风电预测,那么选择以小时方式给出的短期风速预测是合理的。

国内外学者使用了大量的模型来提高风速短期预测的精度,常用的风速预测方法有时间序列法、小波分析法、神经网络法及最小二乘法等^[8]。其中时间序列模型已经被广泛应用,但其对非线性风速序列的预测效果要次于数据挖掘中的非线性模型。近几年,为提高风速预测准确性,组合模型概念被提出^[9-13]。也有研究者着手于研究风速时间序列本身的属性^[14],如周期性、非平稳性、非线性和混沌特性等,并以此为出发点建立预测模型^[15-17]。而在现实中,不同纬度和经度地点风的特征也不同,没有一个模型是通用且最佳的,因此想要提高预测精度,通常要根据风的特征尝试对比不同的模型得出最佳预测模型。

本文使用 DBSCAN 聚类去噪后的数据建立了混沌支持向量回归模型以及 NARX 模型,对 NWTC m2 气象站风速进行短期风速预测,将模型数据分类进行验证。为避免模型比较同质性问题,引入 EGARCH 模型进行预测,比较各模型的预测精度及误差,最后对模型的优缺点进行分析,给出合理建议。

1 风场数据来源

数据来源于美国各地不同气象站的气象数据,其中绝大多数气象站的数据以分钟为记录单位,且包含“风速”、“风向”、“温度”、“气压”、“相对湿度”等各类气象指标。选取 5 个气象站, NWTC m2、SolarTAC、UA OASIS、UO SRML 和 VTIF RSR 从 2016 年 1 月 1 日到 2017 年 12 月 31 日的全部指标数据,以分钟为单位记录数据,选择与本研究相关的气象指标,形成原始数据集,如表 1 所示。

将收集到的所有分钟数据进行数据预处理,包括删除重复数据、填补缺失时刻、进行时间单位转换、剔除异常值。处理结果如表 2 所示。

表 1 各气象站风速影响变量

Table 1 Variables of wind speed impact for weather stations

气象站	变量
NWTC m2	年、日、分钟、风速、风向、温度、相对湿度、气象站气压、海平面气压
SolarTAC	年、日、分钟、风速、风向、温度、相对湿度
UA OASIS	年、日、分钟、风速、风向、温度、相对湿度、气象站气压
UO SRML	年、日、分钟、风速、风向、温度、相对湿度、气象站气压
VTIF RSR	年、日、分钟、风速、风向、温度、相对湿度、气象站气压

表 2 数据预处理结果

Table 2 Results of data preprocessing

变量名	单位	取值范围
年	年	2016,2017
日	个	[1,366]
小时	时	[1,24]
风速	m/s	非负数
风向	自正北°	[0,360]
温度	°C	有理数(正常范围)
相对湿度	%	[0,100]
气象站气压	mBar	正数(正常范围)
海平面气压	mBar	正数(正常范围)

2 DBSCAN 聚类分析

DBSCAN 算法将观测点分为四类:核心点、核心点的直接密度可达点、核心点的密度可达点以及噪声点。其中噪声点是那些在邻域半径范围内有足够邻居,且无法通过其他观测点实现直接密度可达、密度可达和密度相连的点^[18],其思想展现图如图 1 所示。

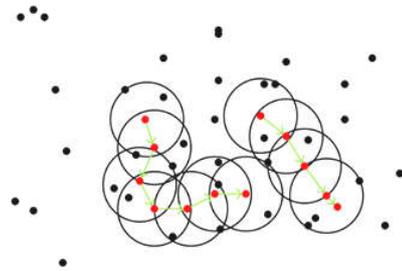


图 1 DBSCAN 聚类思想展现图

Fig. 1 DBSCAN cluster idea illustration

现使用 NWTC m2 气象站的小时数据集进行 DBSCAN 聚类分析,从中随机抽取 10 个子数据集,每个子集包含 3 000 行数据,每行数据均包含风速、风向、温度等 10 个气象变量。对未处理的小时数据集和经 DBSCAN 算法去除噪声点的小时数据集,

分别建立预测模型,使用 10 个子数据集进行交叉验证,以 RMSE 和 MAPE 为评价指标,衡量模型的预测效果。结果表明,去除噪声点后,所建立的模型具有更好的预测效果,对比结果如表 3 所示。

表 3 数据去噪前后模型预测结果对比

Table 3 Comparison of predicted results of pre-and post-noising

数据集	RMSE 平均值	MAPE 平均值
原始小时数据集	1.738 7	0.404 4
DBSCAN 算法 去噪后小时数据集	1.578 2	0.347 2

使用 R 语言编程,对 NWTC m2 气象站的小时数据集执行 DBSCAN 算法^[19],程序运行结果表明,算法共找到 523 个噪声点。DBSCAN 聚类结果如图 2 所示。

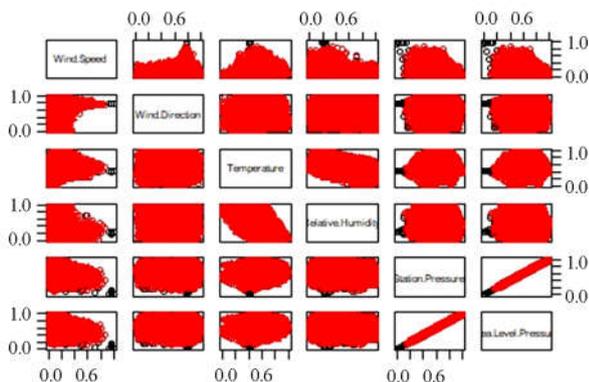


图 2 DBSCAN 聚类结果图示

(邻域半径为 0.3, 邻域半径范围内最少观测点个数为 200)

Fig. 2 Icon of DBSCAN clustering results

(neighborhood radius 0.3, minimum number of observation points within neighborhood radius 200)

3 风速混沌性的判定

考虑各气象站数据的完整性以及准确性,选择 NWTC m2 气象站从 2016 年 1 月 1 日到 2017 年 12 月 31 日,共 731 天的 17 544 组小时数据。

利用 0-1 混沌测试法对 NWTC m2 气象站 731 天共 17 544 个小时风速数据进行分析。为了验证风速数据的混沌特性与时间无关,以风速序列的 10 个不同时刻为起点,各取 8 000 个连续风速数据,对 10 个子序列分别进行 0-1 混沌测试。给出其中一个子序列的检验结果,运行程序后,得到 $p(n)$ 与 $q(n)$ 的相图、均方位移 $M(n)$ 变化图和渐进增长率 K_c 变化图分别如图 3、图 4 和图 5 所示。

从图 3、图 4 和图 5 中可以看出, $p(n)$ 与 $q(n)$ 的相图呈现布朗特性,均方位移 $M(n)$ 随时间线性

增长,渐进增长率 K_c 的平均值为 0.998 0,趋近于 1,因此可以证明该风速时间序列具有混沌特性。

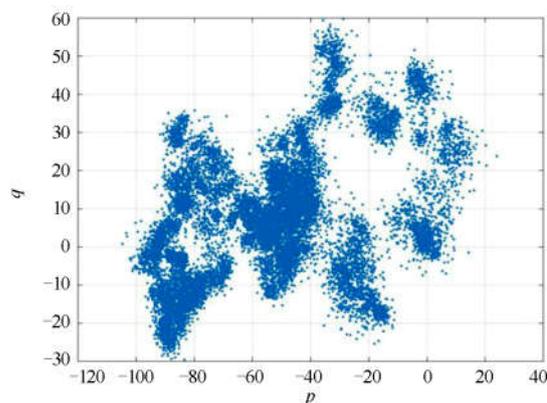


图 3 $p(n)$ 与 $q(n)$ 的相图

Fig. 3 Phase diagram of $p(n)$ and $q(n)$

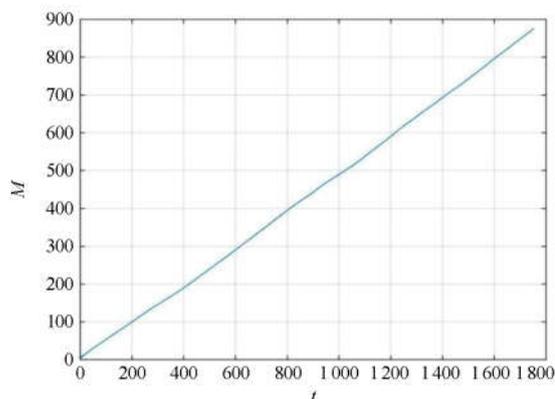


图 4 均方位移 $M(n)$ 变化图

Fig. 4 Changes displacement of mean square $M(n)$

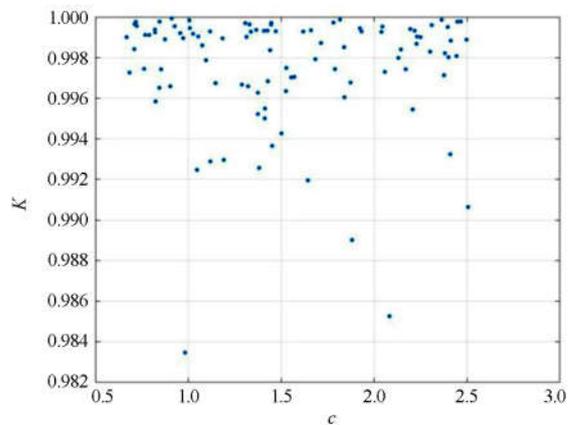


图 5 渐进增长率 K_c 变化图

Fig. 5 Evolution of incremental growth rates K_c

4 风速短期预测

取用 NWTC m2 气象站的 17 544 组小时数据

中, 每组数据包含 9 个变量, 分别为“年”、“日”、“小时”、“风速”、“风向”、“温度”、“相对湿度”、“气象站气压”和“海平面气压”。各气象变量统计摘要如表 4 所示。

表 4 气象变量统计摘要

Table 4 Statistical summary of meteorological variables

变量	最小值	最大值	平均值
风速/(m/s)	0.305 9	24.192 8	3.546 7
风向/(自正北°)	7.535	359.217	211.791
温度/°C	-20.581	34.171	10.831
相对湿度/%	10.05	100.00	50.12
气象站气压/mBar	787.1	828.5	814.0
海平面气压/mBar	986.6	1 036.3	1 018.8

为了验证预测模型在不同时间段内的预测效果, 采用交叉验证的方式, 在风速时间序列(共 17 544 个数据)的不同位置, 分别提取 10 个子序列, 每个子序列包含 3 000 个连续风速数据。10 个子序列分别为: 第 1 个数据到第 3 000 个数据(简记为 1-3000), 440-3439, 2880-5879, 4320-7319, 5760-8759, 7200-10199, 8785-11784, 10224-13223, 11664-14663, 13104-16103。

使用 10 个子序列作为训练样本建立模型, 并分别预测各子序列未来连续 24 h 的风速值。预测方式分为多步预测和单步预测两种。单步预测时进行下一步预测前, 需要更新前一时刻的实际值, 用于下一步的预测, 而多步预测则利用前一步的预测值进行下一步预测。

单步预测效果要好于多步预测, 但实际生产中, 多步预测更被决策者所青睐^[20]。另外多步预测可以预测未来一段时期内的风速变动情况, 以提前做好应对措施, 多步预测虽然精度欠佳, 但可以提供更丰富的信息辅助决策。

4.1 建立EGARCH模型

EGARCH模型是一种广义的统计模型^[21], 可以提取序列中的线性信息, 使残差序列为白噪声序列。在Eviews软件中, EGARCH模型的均值方程为

$$x_t = f(x_o, x_{t-1}, x_{t-2}, \dots) + h_t \quad (1)$$

条件方差方程为

$$\ln(\sigma_t^2) = \omega + \beta \ln(\sigma_{t-1}^2) + \alpha \left| \frac{u_{t-1}}{\sigma_{t-1}} \right| + \gamma \frac{u_{t-1}}{\sigma_{t-1}} \quad (2)$$

式中, 若 $\gamma \neq 0$, 则模型中存在非对称项。

提取风速序列中的确定性信息, 建立风速关于时间的一元线性回归方程, 其中时间变量为 1 到 8784 的计数序列。得到回归方程如下:

$$Wind\ Speed_t = -6.7733 \times 10^{-5} Hour_t + 3.9259 + \varepsilon_t \quad (3)$$

根据回归残差序列异方差性检验, 得出方程的残差序列 ε_t 的自相关系数高阶拖尾, 具有长期自相关性, 偏自相关系数直至 4 阶截尾, 因此建立 4 阶自回归模型 AR(4), 自回归系数多项式为

$$\varepsilon_t = 0.8923 \times \varepsilon_{t-1} - 0.0963 \times \varepsilon_{t-2} + 0.0757 \times \varepsilon_{t-4} + v_t \quad (4)$$

经验证, EGARCH模型的效果最好, 比同类模型, 如 GARCH、TGARCH 以及含有非对称项的 GARCH 衍生模型相比, 具有最小的 AIC 值和 SC 值。建立模型 EGARCH(1,1), 模型含有 1 个非对称项和一阶滞后项 v_{t-1} , 模型如下:

均值方程

$$v_t = -0.0263 \times v_{t-1} + h_t \quad (5)$$

$$h_t = \sigma_t e_t \quad e_t \sim N(0, \text{var}(v_t)) \quad (6)$$

条件方差方程

$$\ln(\sigma_t^2) = -0.0413 + 0.0947 \times |h_{t-1} / \sigma_{t-1}| + 0.2844 \times h_{t-1} / \sigma_{t-1} + 0.9011 \times \ln(\sigma_{t-1}^2) \quad (7)$$

对 EGARCH(1,1) 模型进行估计, 得到结果如表 5 所示。

表 5 EGARCH(1,1)模型的估计结果

Table 5 Estimation results of the EGARCH (1,1) model

Variable	Coefficient	Std. Error	z-Statistic	Prob.
AR_RESID(-1)	-0.026 269	0.009 899	-2.653 620	0.008 0
Variance Equation				
C(2)	-0.041 348	0.004 744	-8.716 388	0.000 0
C(3)	0.094 660	0.006 662	14.207 96	0.000 0
C(4)	0.284 357	0.006 843	41.552 46	0.000 0
C(5)	0.901 076	0.002 483	362.921 3	0.000 0
R-squared	-0.000 744	Mean dependent var	-0.000 397	
Adjusted R-squared	-0.000 744	S.D. dependent var	1.329 201	
S.E. of regression	1.329 695	Akaike info criterion	3.105 984	
Sum squared resid	15 527.36	Schwarz criterion	3.110 015	
Log likelihood	-13 634.93	Hannan-Quinn criter	3.107 357	
Durbin-Watson stat	1.946 313			

由表 5 可知, 模型各项系数及常数项在 0.05 的显著性水平下, 均显著不为 0。DW 统计量为 1.95, 说明残差序列没有一阶自相关性。经检验, 在各种滞后阶数下, F 统计量均不显著, 说明 EGARCH 模型的残差序列已经不存在异方差现象。

根据上述信息, 建立模型如下:

$$\begin{cases} Wind\ Speed_t = -6.7733 \times 10^{-5} Hour_t + 3.9259 + \varepsilon_t \\ \varepsilon_t = 0.8923 \times \varepsilon_{t-1} - 0.0963 \times \varepsilon_{t-2} + 0.0757 \times \varepsilon_{t-4} + v_t \\ v_t = -0.0263 \times v_{t-1} + h_t \\ h_t = \sigma_t e_t \quad e_t \sim N(0, \text{var}(v_t)) \\ \ln(\sigma_t^2) = -0.0413 + 0.0947 \times |h_{t-1} / \sigma_{t-1}| + 0.2844 \times h_{t-1} / \sigma_{t-1} + 0.9011 \times \ln(\sigma_{t-1}^2) \end{cases} \quad (8)$$

式中: $Wind\ Speed_t$ 为 t 时刻的风速值; $Hour_t$ 表示 t 时刻对应的小时在当年的计数位置, 比如 2016 年第二天第一个小时的 $Hour_t$ 值为 25; ε_t 为回归模型在 t 时刻的残差; v_t 为 AR 模型在 t 时刻的残差; h_t 为 EGARCH 模型在 t 时刻的残差; σ_t^2 为 EGARCH 模型在 t 时刻的条件方差。

4.2 建立 NARX 预测模型

NARX 模型是线性外生输入的自回归(ARX)模型在非线性情况下的拓展。NARX 模型的简化模型可以表述为

$$y(t) = f[y(t-1), \dots, y(t-n_a), u(t-n_k), \dots, u(t-n_k-n_b+1)] \quad (9)$$

式中: $y(t)$ 为模型的输出值; $u(t)$ 为模型的输入值; $f(\cdot)$ 是一个非线性函数, 它取决于有限个先前的输入输出值, 即 $f(\cdot)$ 的确定取决于训练数据; n_a 是 NARX 模型的输出阶次; n_b 是 NARX 模型的输入阶次; n_k 是 NARX 模型输出的延时段。

NARX 模型主要由回归分析器和非线性估计器两部分组成, 其模型结构如图 6 所示。

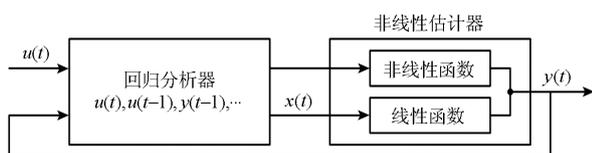


图 6 NARX 模型结构

Fig. 6 NARX model structure

NARX 模型的回归分析器可以视为一个“黑箱”, 操作者无法看到“黑箱”内部的具体运算过程, 也无法获得“黑箱”中回归分析各节点的权重, 但可以通过“黑箱”运算后的返回值评价回归分析器的性能。

NARX 模型的回归分析器利用“黑箱” t 时刻以及之前的输入数据 $u(t), u(t-1), u(t-2), \dots$ 和 t 时刻以前的输出数据 $y(t-1), y(t-2), \dots$, 计算确定各层各节点的权重, 然后将结果 $x(t)$ 传递给模型的非线性估计器。

NARX 模型的非线性估计器由两部分组成, 非线性函数模块和线性函数模块, 这两个模块可以在回归分析器与模型输出之间建立起映射关系。NARX 模型的非线性估计器的函数表达式为

$$y = F(x) = L^T(x-r) + d + g[Q(x-r)] \quad (10)$$

式中: y 是非线性估计器的输出结果, 即 NARX 模型的最最终输出; x 是非线性估计器的输入; $L^T(x-r) + d$

是线性函数模块的输出; d 是输出偏置量; $g[Q(x-r)]$ 是非线性函数模块的输出; r 是回归矢量的平均值; Q 是非线性子空间^[22]。

4.3 建立混沌支持向量机预测模型

混沌时间序列被认为是一种存在于确定性系统中的不规则运动, 具有复杂性、伪随机性和初值敏感性等特征。因此识别时间序列的混沌特性, 并建立模型提取不确定性信息, 可以更充分地挖掘序列中的信息, 从而提高对时间序列的预测能力。基于风速时间序列具有混沌特性的前提下, 结合混沌时间序列的相空间重构和支持向量机回归理论^[23], 建立一种基于风速混沌特性和最小二乘向量机的短期风速预测模型。

取风速子序列进行说明, 子序列包含全部数据的第 1 个到第 3 000 个数据。 $S_{cor}(t)$ 的最小值为 33, 所以延迟时间窗口 $\tau_w = 33$, 从图 9 中可以看出, $\bar{S}(t)$ 的第一个零点为 25, $\Delta\bar{S}(t)$ 的第一个局部极小值点也为 25, 因此最佳延迟时间 $\tau_d = 25$, 进一步最佳嵌入维数 $m = \text{fix}(\tau_w / \tau_d) + 2 = 3$ 。由 C-C 法得出各统计量曲线如图 7 所示。

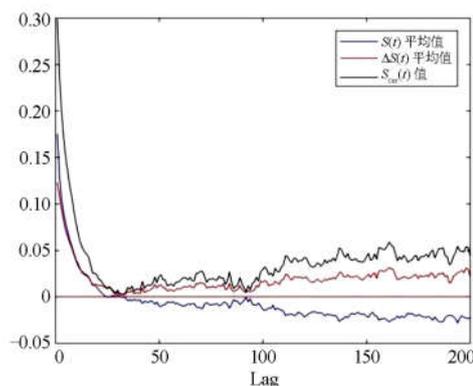


图 7 C-C 法各统计量曲线

Fig. 7 Statistical curves of C-C method

对 10 个子序列进行模型适应性分析, 10 次检验中, 由 C-C 法确定的延迟时间和嵌入维数如表 6 所示。

使用 10 个风速子序列, 每个子序列包含 3 000 个数据量。以 C-C 法得到的延迟时间 τ_d 和嵌入维数 m 为中心, 在其适当的参数范围内, 循环对各个子序列进行相空间重构, 然后使用支持向量回归模型进行风速短期预测, 参考预测评价指标 RMSE 以及 MAPE, 选择使得预测效果最好的延迟时间 τ_d 和嵌入维数 m 为最佳相空间重构参数^[24]。10 个子序列所确定的最佳延迟时间和嵌入维数值如表 7 所示。

表 6 C-C 法计算延迟时间和嵌入维数

Table 6 Delay time and embedded dimension calculated by C-C method

序列	延迟时间/s	时间窗口/s	嵌入维数
1-3000	25	33	3
440-3439	29	43	3
2880-5879	9	18	4
4320-7319	26	132	7
5760-8759	30	159	7
7200-10199	19	39	3
8785-11784	18	35	3
10224-13223	29	84	4
11664-14663	10	40	6
13104-16103	19	45	4

注:完整序列共有 17 544 个数据,1-3000 表示从第 1 个数据到第 3 000 个数据。

表 7 遍历后最佳延迟时间和嵌入维数

Table 7 Optimal delay time and embedding dimension after traversal

序列	延迟时间/s	嵌入维数
1-3000	10	5
440-3439	18	5
2880-5879	11	6
4320-7319	67	7
5760-8759	47	10
7200-10199	42	9
8785-11784	89	5
10224-13223	90	5
11664-14663	24	4
13104-16103	23	6

5 模型预测结果对比

以均方根误差(RMSE)和平均绝对百分误差(MAPE)作为预测评价指标^[25]。

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2} \quad (11)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_i - \hat{x}_i}{x_i} \right| \quad (12)$$

式中: x_i 为实际值; \hat{x}_i 为预测值; n 为预测数据的个数,本次预测 n 为 24。

对于混沌支持向量机中子序列之后的预测,均使用得到的最佳参数重构相空间,并利用支持向量回归模型进行预测。当预测精度明显下降时,再重新选择近期的 3 000 个数据,采用同样的方式寻找最佳延迟时间和嵌入维数,重构相空间并进行预测。预测结果拟合图如图 8 所示。

多步预测 10 次验证中,预测评价指标 RMSE 和 MAPE 的值如表 8 所示。

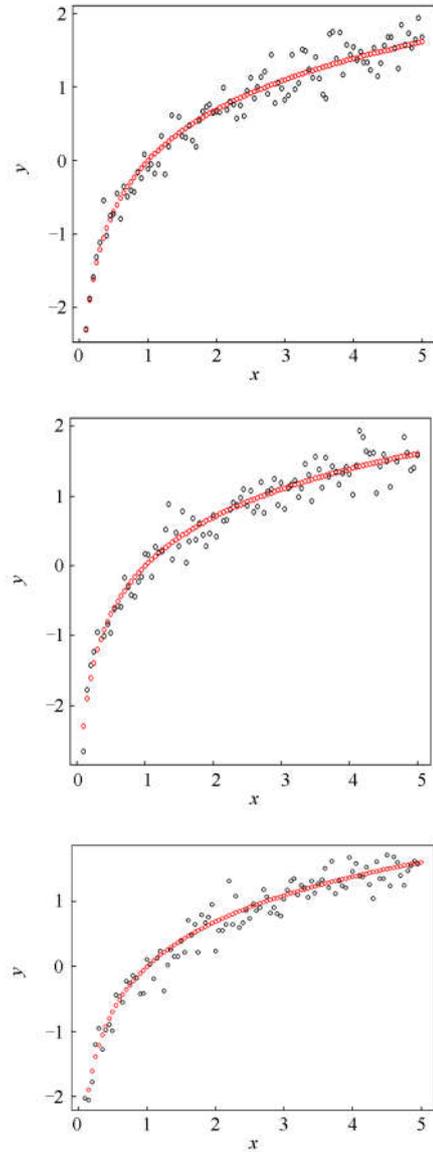


图 8 混沌支持回归向量机预测观察拟合图

Fig. 8 Predictive observation fit graph of chaotic support regression vector machine

表 8 混沌支持向量机模型多步预测 RMSE 值和 MAPE 值

Table 8 Multi-step prediction of RAMSE and MAPE values in chaotic support vector machine model

次数	1	2	3	4	5	6
RMSE	0.717 2	0.738 2	1.140 8	0.703 8	0.869 4	1.653 9
MAPE	0.292 0	0.231 0	0.338 1	0.223 0	0.394 2	0.216 3
次数	7	8	9	10	平均值	标准差
RMSE	1.354 8	0.396 9	0.782 9	0.829 8	0.918 8	0.365 2
MAPE	0.292 5	0.190 1	0.223 9	0.600 8	0.300 2	0.123 2

针对 NARX 预测模型,通过调用神经网络时间序列工具箱“ntstool”中的“narxnet”函数来构建^[26],

隐含层个数为 20, 输入延迟和反馈延迟均为 4(实验发现此时收敛最快)。

使用 10 个子序列对 NARX 模型进行交叉验证, 分别进行多步预测和单步预测, 均预测序列未来 24 h 的风速值。预测结果如图 9 所示, 10 次检验中 RMSE 和 MAPE 的值如表 9 所示。

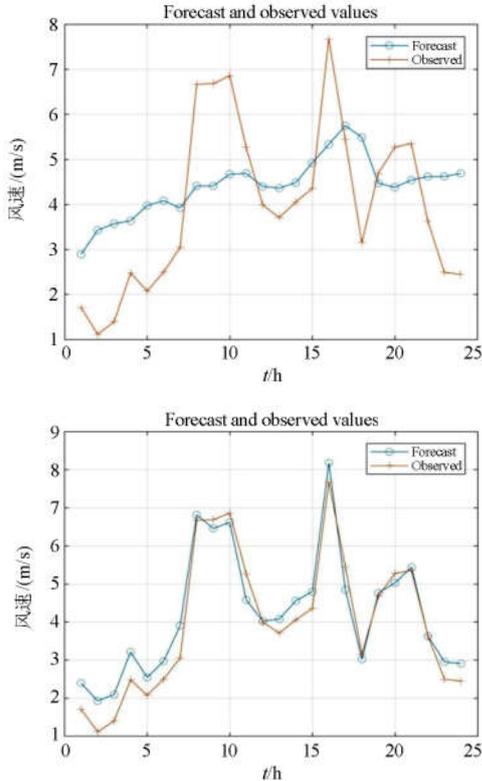


图 9 NARX 模型多步预测(上)、单步预测(下)结果图
Fig. 9 Multi-step prediction (above) and single-step prediction (below) of NARX model

表 9 NARX 模型多步预测 RMSE 值和 MAPE 值
Table 9 Multi-step prediction of RMSE and MAPE values in NARX model

次数	1	2	3	4	5	6
RMSE	0.648 3	0.658 6	1.172 8	0.765 5	1.186 1	2.016 7
MAPE	0.180 5	0.170 2	0.297 3	0.193 6	0.191 6	0.198 4
次数	7	8	9	10	平均值	标准差
RMSE	1.542 1	0.685 7	0.825 6	1.015 8	1.051 7	0.446 0
MAPE	0.256 2	0.196 5	0.207 1	0.295 9	0.218 7	0.046 8

E 的平均值为 1.051 7, $MAPE$ 的平均值为 0.218 7, $RMSE$ 的标准差为 0.446 0, $MAPE$ 的标准差为 0.046 8。

针对 EGARCH 模型, 用 2016 年全年的风速时间序列建模, 预测 2017 年前 24 h 的风速值。使用 Eviews 8 建立模型, 并解出模型参数, 使用 R 语言

编程进行多步预测, 得到预测值与观察值的时序图如图 10 所示, 并得出两者的误差统计表, 如表 10 所示。

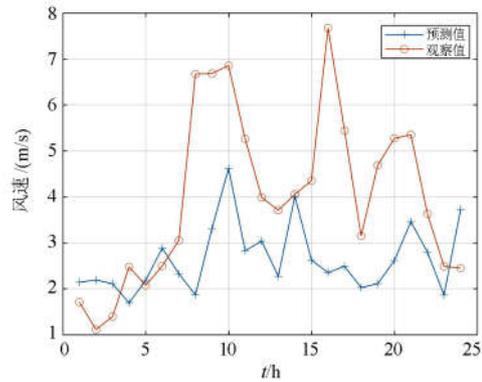


图 10 预测值与观察值时序图
Fig. 10 Timing diagram of predicted and observed values

表 10 预测值和观测值的相对误差统计表
(预测 2017 年前 24 h 风速)

Table 10 Relative error statistics of predicted and observed values (predicting wind speeds for the first 24 h of 2017)

时间/h	预测值/(m/s)	观测值/(m/s)	相对误差/%
1	2.138 8	1.707 2	25.28
2	2.190 8	1.111 5	97.10
3	2.109 0	1.394 3	51.27
4	1.689 8	2.475 6	31.74
5	2.186 6	2.071 9	5.54
6	2.885 8	2.495 6	15.63
7	2.325 6	3.050 8	23.77
8	1.866 7	6.669 0	72.01
9	3.309 4	6.681 7	50.47
10	4.618 3	6.855 7	32.64
11	2.828 7	5.262 3	46.25
12	3.034 0	3.986 8	23.90
13	2.270 5	3.707 7	38.76
14	4.009 4	4.053 4	1.08
15	2.622 1	4.351 2	39.74
16	2.352 7	7.671 5	69.33
17	2.486 7	5.437 2	54.27
18	2.025 9	3.148 2	35.65
19	2.107 4	4.685 3	55.02
20	2.613 0	5.271 5	50.43
21	3.459 7	5.351 8	35.35
22	2.794 3	3.628 2	22.98
23	1.861 1	2.486 9	25.16
24	3.724 4	2.446 6	52.23

使用 10 个子序列分别代入模型, 并预测接下来 24 h 的风速值, 10 次测试的预测评价指标值如表 11 所示。

表 11 EGARCH 模型多步预测的 RMSE 值与 MAPE 值
Table 11 RMSE values and MAPE values for multi-step prediction of EGARCH model

次数	1	2	3	4	5	6
RMSE	1.487 0	3.616 6	1.913 6	1.566 5	2.477 6	5.348 0
MAPE	0.626 9	1.151 3	0.840 4	0.593 4	1.384 7	0.556 2
次数	7	8	9	10	平均值	标准差
RMSE	3.009 3	1.184 9	1.500 0	1.790 2	2.389 4	1.289 7
MAPE	0.659 7	0.570 5	0.416 4	1.395 9	0.819 6	0.360 5

10 次验证中, EGARCH 模型预测结果的平均 RMSE 值为 2.389 4, 平均 MAPE 值为 0.819 6。10 次预测中, 所得 RMSE 的标准偏差为 1.289 7, 所得 MAPE 的标准偏差为 0.360 5。

对比 3 个模型的多步预测效果, 结果如表 12 所示。

表 12 模型多步预测效果对比

Table 12 Comparison of multi-step predictive effects of model

模型	RMSE		MAPE	
	平均值	标准差	平均值	标准差
EGARCH	2.389 4	1.289 7	0.819 6	0.360 5
NARX	1.051 7	0.446 0	0.218 7	0.046 8
混沌支持向量机	0.918 8	0.365 2	0.300 2	0.123 2

从 RMSE 均值可以看出, 重构相空间后, 混沌支持向量机预测精度更高, 且 10 次验证中 RMSE 的波动范围更小。从 MAPE 均值可以看出, 同样机器学习模型的预测效果要好于 EGARCH 模型, 相空间重构后, 混沌支持向量机的预测效果要强于 NARX 模型。且 NARX 模型 10 组 MAPE 的标准差仅为 0.046 8, 表明模型具有较强的鲁棒性, 对一年中不同时刻风速数据有较好的预测能力。

6 结论

本文将 NARX 模型和基于混沌时间序列的支持向量机模型应用于短期风速预测, 为避免模型比较同质性问题, 引入 EGARCH 模型进行预测。对三种模型的预测结果比较发现: 基于混沌时间序列的支持向量机模型在预测具有混沌特性的数据时具有一定的优越性。

究其原因, 主要是由于混沌支持向量机模型考虑了风速序列具有混沌性这一特点。据此, 结合混沌时间序列的相空间重构和支持向量机回归理论, 建立了基于风速混沌特性和最小二乘支持向量机的混沌支持向量机短期风速预测模型。

此模型在整个构建过程中结合混沌原理, 使用 C-C 法计算出延迟时间和嵌入维数, 并利用这 2 个

参数重构相空间, 减弱了混沌特性对模型的影响, 使得模型可以最大限度地提取序列中的信息, 得到了更好的预测结果。

参考文献

- [1] SAWYER S. Global wind energy outlook 2016[EB/OL]. <https://gwec.net/publications/global-wind-energy-outlook/global-wind-energy-outlook-2016/#>.
- [2] 孙春顺, 王耀南, 李欣然. 小时风速的向量自回归模型及应用[J]. 中国电机工程学报, 2008, 28(14): 112-117. SUN Chunshun, WANG Yaonan, LI Xinran. A vector autoregression model of hourly wind speed and its applications in hourly wind speed forecasting[J]. Proceedings of the CSEE, 2008, 28(14): 112-117.
- [3] WAN Y H, BUCANE D. Short-term power fluctuations of large wind power plants[J]. Journal of Solar Energy Engineering, 2002, 124(4): 353-360.
- [4] 张博. 混沌支持向量机风速短期预测及其在 pcDuino 平台实现[D]. 太原: 中北大学, 2014. ZHANG Bo. Short-term prediction of wind speed based on chaotic support vector machine and implementation on pcDuino platform[D]. Taiyuan: North University of China, 2014.
- [5] SCHWARTZ M, BAILEY N, et al. Wind forecasting objectives for utility schedulers and energy traders[J]. Office of Scientific & Technical Information Technical Reports, 1998.
- [6] ZHU X, GENTON M G. Short-term wind speed forecasting for power system operations[J]. International Statistical Review, 2012, 80(1): 2-23.
- [7] 孙斌, 姚海涛, 李田, 等. 基于 Fast ICA 和改进 LSSVM 的短期风速预测[J]. 电力系统及其自动化学报, 2014, 26(1): 22-27. SUN Bin, YAO Haitao, LI Tian, et al. Short-term wind speed forecasting based on Fast ICA algorithm and improved LSSVM Model[J]. Proceedings of the CSU-EPSA, 2014, 26(1): 22-27.
- [8] 王静, 李维德. 基于 CEEMD 和 GWO 的超短期风速预测[J]. 电力系统保护与控制, 2018, 46(9): 69-74. WANG Jing, LI Weide. Ultra-short-term forecasting of wind speed based on CEEMD and GWO[J]. Power System Protection and Control, 2018, 46(9): 69-74.
- [9] 张妍, 王东风, 韩璞. 一种风电场短期风速组合预测模型[J]. 太阳能学报, 2017, 38(6): 1510-1516. ZHANG Yan, WANG Dongfeng, HAN Pu. Combination forecasting model of short-time wind speed for wind farm[J]. Acta Energiæ Solaris Sinica, 2017, 38(6): 1510-1516.

- [10] MENG A, GE J, YIN H, et al. Wind speed forecasting based on wavelet packet decomposition and artificial neural networks trained by crisscross optimization algorithm[J]. *Energy Conversion and Management*, 2016, 114: 75-88.
- [11] TASICARA OGLU A, UZUNOGLU M. A review of combined approaches for prediction of short-term wind speed and power[J]. *Renewable & Sustainable Energy Reviews*, 2014, 34(6): 243-254.
- [12] POUSINHO H M I, MENDES V M F, CATALAO J P S. A hybrid PSO-ANFIS approach for short-term wind power prediction in Portugal[J]. *Energy Conversion & Management*, 2011, 52(1): 397-402.
- [13] LIANG Z, LIANG J, WANG C, et al. Short-term wind power combined forecasting based on error forecast correction[J]. *Energy Conversion & Management*, 2016, 119: 215-226.
- [14] LIU H, ERDEM E, SHI J. Comprehensive evaluation of ARMA-GARCH (-M) approaches for modeling the mean and volatility of wind speed[J]. *Applied Energy*, 2011, 88(3): 724-732.
- [15] LIU H, TIAN H Q, LI Y F. An EMD-recursive ARIMA method to predict wind speed for railway strong wind warning system[J]. *Journal of Wind Engineering & Industrial Aerodynamics*, 2015, 141: 27-38.
- [16] 许刘锋, 杨玉良, 张磊, 等. 基于混沌—LSSVM 神经网络风电场风速短期预测[J]. *电气自动化*, 2013, 35(4): 25-27.
XU Liufeng, YANG Yuliang, ZHANG Lei, et al. Wind speed forecasting in wind farm based on chaotic-lssvm neural networks[J]. *Electrical Automation*, 2013, 35(4): 25-27.
- [17] 鲁迪, 王星华, 贺小平. 基于混合粒子群算法和多分位鲁棒极限学习机的短期风速预测方法[J]. *电力系统保护与控制*, 2019, 47(5): 115-122.
LU Di, WANG Xinghua, HE Xiaoping. Hybrid population particle algorithm and multi-quantile robust extreme learning machine based short-term wind speed forecasting[J]. *Power System Protection and Control*, 2019, 47(5): 115-122.
- [18] 荣秋生, 颜君彪, 郭国强. 基于 DBSCAN 聚类算法的研究与实现[J]. *计算机应用*, 2004, 24(4): 45-46.
RONG Qiusheng, YAN Junbiao, GUO Guoqiang. Research and implementation of clustering algorithm based on DBSCAN[J]. *Computer Applications*, 2004, 24(4): 45-46.
- [19] RESHEF D N, RESHEF Y A, FINUCANE H K, et al. Detecting novel associations in large datasets[J]. *Science*, 2011, 334(6062): 1518.
- [20] 杨杰, 霍志红, 何永生, 等. 基于小波与最小资源分配网络的超短期风电功率预测研究[J]. *电力系统保护与控制*, 2018, 46(9): 55-61.
YANG Jie, HUO Zhihong, HE Yongsheng, et al. Ultra-short-term wind power prediction based on wavelet and minimum resource allocation network[J]. *Power System Protection and Control*, 2018, 46(9): 55-61.
- [21] 崔海蓉, 张京波, 何建敏. 基于 AR-EGARCH 的空气温度预测模型[J]. *统计与信息论坛*, 2013, 28(10): 36-41.
CUI Hairong, ZHANG Jingbo, HE Jianmin. The air temperature forecasting model based on AR-EGARCH[J]. *Statistics & Information Forum*, 2013, 28(10): 36-41.
- [22] 郑伟. 基于 NARX 模型的光伏并网逆变器非线性模型辨识方法[J]. *智能电网*, 2015, 5(5): 228-241.
ZHENG Wei. A nonlinear model identification method of photovoltaic grid-connected inverters based on the NARX model[J]. *Smart Grid*, 2015, 5(5): 228-241.
- [23] KONG X, LIU X, SHI R, et al. Wind speed prediction using reduced support vector machines with feature selection[J]. *Neurocomputing*, 2015, 169: 449-456.
- [24] HOU K Y, SHAO G H, WANG H M, et al. Research on practical power system stability analysis algorithm based on modified SVM[J]. *Protection and Control of Modern Power Systems*, 2018, 3(3): 119-125. DOI: 10.1186/s41601-018-0086-0.
- [25] LIU H, MI X, LI Y. An experimental investigation of three new hybrid wind speed forecasting models using multi-decomposing strategy and ELM algorithm[J]. *Renewable Energy*, 2018, 123: 694-705.
- [26] NAIK J, SATAPARTHY P, DASH P K. Short-term wind speed and wind power prediction using hybrid empirical mode decomposition and kernel ridge regression[J]. *Applied Soft Computing*, 2018, 70(9): 1167-1188.

收稿日期: 2018-12-07; 修回日期: 2019-04-14

作者简介:

李应求(1965—), 男, 通信作者, 博士, 教授, 博士生导师, 研究方向为应用统计。E-mail: liyq-2001@163.com

(编辑 魏小丽)