

基于 LS-SVM 和核密度估计的概率性风电功率预测

孟升卫, 冯晓晓, 庞景月, 崔秀海

(哈尔滨工业大学 自动化测试与控制研究所, 哈尔滨 150080)

摘要: 近年来风电在我国发展迅猛, 但风速的不稳定性和间歇性, 使风电功率也具有同样的性质, 这样的电功率注入会带来电力系统运行的不稳定, 因此, 风电功率的预测对风电并网及使用具有重要意义; 鉴于此, 开展风电功率的短期预测研究, 利用 LS-SVM 对风电功率进行建模并实现确定性的短期预测, 在此基础上使用非参数统计法对确定性预测模型的预测误差进行拟合获得其密度函数, 计算各功率段的置信区间以得到概率性预测结果, 从而提高风电功率预测结果的实用性和可靠性; 与常用的自回归滑动平均模型和 BP 神经网络模型进行对比实验, 证明该方法的性能及优势。

关键词: 风电功率预测; 概率性预测; LS-SVM; 核密度估计

Probabilistic Wind Power Prediction Based on LS-SVM and Kernel Density Estimation

Meng Shengwei, Feng Xiaoxiao, Pang Jingyue, Cui Xiuhai

(Institute of Automatic Test and Control, Harbin Institute of Technology, Harbin 150080, China)

Abstract: In recent years, wind power has developed rapidly in China, but the instability and intermittence of wind speed make the wind power have the same nature. Such power injection will bring instability to the operation of the power system. Therefore, prediction of wind power is of great significance to wind power network and utilization. In view of this, short-term prediction of wind power is focused based on LS-SVM, and thus the deterministic short-term prediction is achieved. In addition, the non-parametric statistical method is used to fit the prediction error of the deterministic prediction model to estimate its density function, and calculate the confidence interval of each power segment to obtain probabilistic prediction results. As a result, the practicability and reliability of wind power prediction results are improved. Compared with the commonly used autoregressive moving average model and BP neural network model, the performance and advantages of the method are proved.

Keywords: wind power output prediction; probabilistic prediction; LS-SVM; kernel density estimation

0 引言

目前, 我国已成为世界上第一大能源消费国, 在能源消费结构中煤炭占比最大, 这使我国面临着巨大的环境压力, 发展绿色能源显得尤为重要。我国的风能资源非常丰富, 风电建设在我国的能源发展格局中有着举足轻重的地位。但风速的不稳定性和间歇性使风电功率也具有同样的性质, 这给风电功率注入电网带来了巨大的挑战, 直接影响我国风电发展的进程。因此, 风电功率预测一直都是研究的热门领域^[1]。

风电功率预测方法按照预测结果的形式可以分为确定性预测和概率性预测。确定性预测仅给出预测点的信息, 不能描绘风电功率的不稳定性。而概率性预测可以提供预测点的误差信息, 在应用层面, 这对电网运行调度、规划

和稳定性分析有重要作用。

风电功率概率性预测主要有两类方法, 一是对现有确定性预测模型的误差进行拟合获得其密度函数, 二是直接对变量进行概率性预测^[2]。文献 [3] 中在 ARMA (Autoregressive Moving Average Model) 和 SVM (Support Vector Machine) 两种确定性预测模型的基础上, 实现了基于分位数的概率性预测方法^[3]。文献 [4] 中使用了小波阈值去噪和递归神经网络作为风速预测的子模型, 从子模型中获得方差, 然后计算概率性预测的预测区间, 其中方差包括了建模和预测的不确定性^[4]。文献 [5] 中提出了一种新的多模型组合概率风电预测方法, 以利用不同预测模型的优势, 这些模型提供不同类型的概率密度函数以提高概率预测的性能。基于稀疏贝叶斯学习、核密度估计和 β 分布拟合 3 个概率预测模型, 用于形成组合模型^[5]。本文采用第一种概率性预测方法, 在 LS-SVM (Least Squares Support Vector Machines) 确定性预测模型的基础上, 应用核密度估计法对模型的误差进行密度估计, 实现风电功率概率性预测。

收稿日期: 2019-04-10; 修回日期: 2019-05-08。

作者简介: 孟升卫(1971-), 男, 博士, 副教授, 主要从事仪器总线技术、精密时间间隔测量、穿墙成像雷达等方向的研究。

崔秀海(1971-), 男, 博士, 副研究员, 主要从事电子系统可靠性及高性能计算方向的研究。

1 研究方法的基本原理

1.1 LS-SVM 的基本原理

支持向量机 (SVM) 是一个十分出色的机器学习算法, 它成功的解决了一系列难题, 例如高维和局部极小值问题等^[6]。

为了解决支持向量机模型在处理大数据集时优化困难的问题, LS-SVM 算法在 SVM 算法的基础上稍作改进, 在优化时引入了误差的平方项, 并用正则化参数 γ 来调节对误差项的惩罚力度, 使不等式约束变成了等式约束, 降低了 LS-SVM 模型计算的复杂性。

在 LS-SVM 中, 回归问题对应的优化式为:

$$\min_{w, \xi, e} Q(w, \xi, e) = \frac{1}{2} \|w\|^2 + \frac{\gamma}{2} \sum_{i=1}^l e_i^2 \quad (1)$$

其中: e_i 为估计误差, γ 为正则化参数。

约束条件为:

$$|w^T \varphi(x_i) + \xi + e_i| < \epsilon \quad (2)$$

其中: $\varphi(x)$ 是非线性函数。

其拉格朗日函数如下:

$$L(w, b, e, \alpha) = J(w, b, e) - \sum_{i=1}^N \alpha_i [w^T \varphi(x_i) + b + e_i - y_i] \quad (3)$$

对拉格朗日函数各参数求偏导数有:

$$\begin{cases} \frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^N \alpha_i \varphi(x_i) \\ \frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^N \alpha_i = 0 \\ \frac{\partial L}{\partial e_i} = 0 \Rightarrow \alpha_i = C e_i \\ \frac{\partial L}{\partial \alpha_i} = 0 \Rightarrow w^T \varphi(x_i) + b + e_i - y_i = 0 \end{cases} \quad (4)$$

整理成矩阵形式可得:

$$\begin{bmatrix} 0 & \vec{1}^T \\ \vec{1} & \Omega + C^{-1} I_N \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ Y \end{bmatrix} \quad (5)$$

其中: $\vec{1} = [1; \dots; 1]$; $\Omega_{ij} = \varphi(x_i) \varphi(x_j) = K(x_i, x_j)$, $i, j = 1, \dots, l$; I_N 为 N 阶单位矩阵。令 $A = \Omega + C^{-1} I_N$ 可得:

$$\begin{cases} b = \frac{\vec{1}^T A^{-1} y}{\vec{1}^T A^{-1} \vec{1}} \\ \alpha = A^{-1} (Y - b \vec{1}) \end{cases} \quad (6)$$

所得 LS-SVM 模型预测函数为:

$$f(x) = \sum_{i=1}^l \alpha_i K(x, x_i) + b \quad (7)$$

1.2 核密度估计法

对样本数据分析时, 一般用密度函数和分布函数来描述样本数据分布的特点。一些特点类型的样本数据, 有足够的经验可以说明其数据分布的特征, 若密度函数形式已知, 可以利用样本数据估计出密度函数的具体参数, 这种方法称为参数估计法。但当密度函数形式未知, 用参数估计法就无从下手了, 这种情况下就可以用非参数估计方法

对样本数据的密度函数进行估计。

非参数估计方法不依赖假定的密度函数形式, 可以用于估计多种类型数据的密度函数, 例如非正态数据、重尾数据等^[7]。基本的非参数估计方法有直方图估计法、核密度估计法、近邻估计法、序列估计法等。下面简要介绍核密度估计法的基本原理:

假设有一组一维离散随机样本, 其密度函数 $f(x)$ 和分布函数 $F(x)$ 形式未知, 则有:

$$f(x) = F'(x) = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x-h)}{2h} \quad (8)$$

其中: h 是核密度估计的带宽 (窗口)。

一个常用的经验布函数为:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x) \quad (9)$$

其中: $I(x)$ 为核函数。

将 $F_n(x)$ 代入式 (8), 则密度函数的估计为:

$$f_n(x) = \frac{F_n(x+h) - F_n(x-h)}{2h} \quad (10)$$

计算 $f_n(x)$ 的方差和均值, 当 $h \rightarrow 0$ 且 $nh \rightarrow \infty$ 时有: 方差趋于 0, 而均值趋于 $f(x)$, 说明 $f_n(x)$ 对 $f(x)$ 的估计是准确的。

取 $k(u) = \frac{1}{2} I(|u| \leq 1)$ 为核函数, 得:

$$f_n(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} k\left(\frac{X_i - x}{h}\right) \quad (11)$$

其中: u 为核函数宽度。

由公式 (9) 可知, 核密度函数估计是一个加权和, 在进行加权求和时, 对处于区间 $[x-h, x+h]$ 中的样本点赋予同样的权重。这样的处理实际上是不合理的, 因为在估计 X 在 x 点处的密度 $f(x)$ 时, 离 x 越近的样本点提供的关于 $f(x)$ 的信息越多。因此, 在定义 $f(x)$ 的估计时, 就要对距离 x 较近的点赋予较大的权值, 对距离 x 较远的点赋予较小的权值, 可以选取不同的核函数来决定各点的权值。

2 基于 LS-SVM 和核密度估计的预测模型

风电功率预测按照预测方法的结果可以分为确定性预测和概率性预测。确定性预测的主要不足是预测结果中没有提供任何关于预测值预测误差的信息, 限制了这类方法在实际中的应用。而概率性预测则可以为每个预测量提供一个预测区间, 弥补了确定性预测的不足。

本文在 LS-SVM 确定性预测模型基础上, 利用核密度估计方法计算确定性预测模型误差的分布函数, 实现了风电功率的概率性预测, 概率性预测模型的算法框图如图 1 所示。

2.1 基于 LS-SVM 的确定性预测模型

在 LS-SVM 预测模型进行时间序列预测之前, 需要对输入时间序列进行相空间重构得到输入向量。构造的输入向量不仅关系到模型的复杂度, 而且会影响到模型的预测精度。

重构得到的输入向量并不能直接作为模型的输入, 因为原始数据的范围变化较大, 有些数据可能会使模型内部

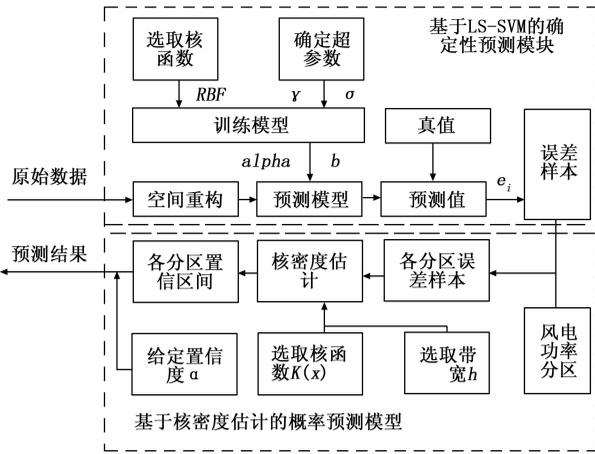


图 1 整体算法框架

函数饱和从而影响模型的预测精度，所以输入向量需要归一化处理后才能输入模型。预测得到的预测值经过反归一化后，才能得到实际的预测值。

在模型构建的过程中需要选择核函数，常见的核函数有 Sigmoid 核函数、多项式核函数、径向基 RBF (Radial Basis Function) 核函数等^[8]。本文选取 RBF 核函数。另外，超参数的选择也是构建模型的重点，需要设定的超参数有正则化参数 γ 和核函数参数 σ 。 γ 称作惩罚因子，决定对误差平方项的惩罚力度，关系到模型的训练误差和泛化能力。在实际应用中 σ 和 γ 的偏大或者偏小都会引起模型的过学习或欠学习现象，因此选择合理的超参数非常重要。

本文采用网格搜索法来确定超参数，利用此方法可以得到全局范围内的最优解^[9]。具体步骤如下：

- 1) 设定两个参数的选择范围和步长。
- 2) 设定初始化参数，移动网格交点在每一处交点建立模型，计算得模型的平均绝对误差 MAE。
- 3) 搜索完毕后选取计算指标最优的一组参数作为模型的超参数。

在模型建立完成之后，需要在训练集上训练模型，得到支持向量 $alpha$ 和偏移参数 b ，构造决策函数，确定风电功率确定性预测模型。

2.2 基于核密度估计的概率性预测模型

在确定性预测的基础上，通过对确定性预测模型误差密度函数的估计，应用概率论理论可以计算任意置信度下的置信区间。

对不同的风电功率区段，风电功率确定性预测模型的误差分布的范围有明显区别。因此，需要将风电功率进行分区，对不同的分区分别计算误差概率密度曲线和各功率区间的置信区间^[10]，实现全功率段的概率性预测。

假设风电功率值的最大值为 P_{max} 、最小值为 P_{min} 、区间长度为 ΔP ，分区数 n 为：

$$n = (P_{max} - P_{min}) / \Delta P + 1 \quad (12)$$

初步区间划分后，有时会出现某些分区样本点不足的情况，这样所得到的误差密度分布不能准确地反映误差的

实际分布情况。这时需要将相邻的样本点较少的区间合并，直到合并区间的误差样本点达到数量要求。

应用核密度估计法计算各区间的误差概率密度函数：

$$f(e) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{e - e_i}{h}\right) \quad (13)$$

其中： e_i 为 LS-SVM 模型预测误差样本， N 为样本总数， $K(x)$ 为核函数， h 为带宽。

对密度函数积分可得分布函数，设定置信度 α ，使下式成立：

$$P(x_{min} < x < x_{max}) = 1 - \alpha \quad (14)$$

则 $[x_{min}, x_{max}]$ 为该功率区间的置信区间。依次对每个功率区间进行以上步骤可得全功率段的置信区间。最后，将每个预测点的上下限分别于相邻的点上上下限相连即可得到整个时段的概率性预测结果^[11]。

3 实验结果与分析

3.1 实验数据

本文的实验数据来自比利时 Elia 公司 AggregateBelgian 风电场，其中记录了该风电场 2018 年 2 月 28 日至 2018 年 3 月 29 日期间的共 2974 个数据点，时间间隔为 15 分钟。

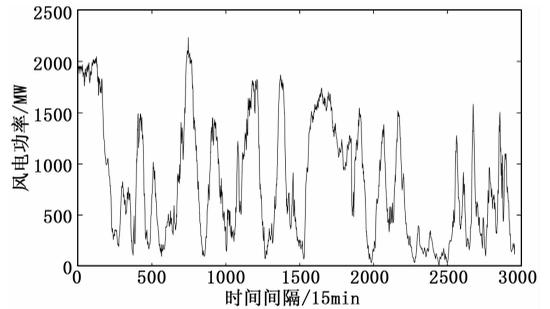


图 2 风电功率实验数据

从图 2 可以看出，实验数据存在数据缺失的情况，表现为数据中出现连续的 0 值，这些值将会对预测算法产生影响；因此，在预测实现之前，必须对这些缺失值进行插值处理。将预处理后的实验数据中的前 2000 个作为模型训练数据，后 974 个作为测试数据。

3.2 基于 LS-SVM 和核密度的概率性预测验证

使用该预测模型，对测试数据中的 974 个数据点进行预测，得到 974 个误差样本。将风电功率的范围等分为 4 个区间，分别为 $[0, 500]$ 、 $[500, 1000]$ 、 $[1000, 1500]$ 、 $[1500, 2000]$ ，对各区段的误差样本分别求取概率密度函数，如图 3 所示。

给定置信度为 0.9，则所得各分区的置信区间为：

- 1) 功率区间 $[0, 500]$ 的置信区间为 $[-75, 75]$ ；
- 2) 功率区间 $[500, 1000]$ 的置信区间为 $[-150, 150]$ ；
- 3) 功率区间 $[1000, 1500]$ 的置信区间 $[-125, 125]$ ；
- 4) 功率区间 $[1500, 2000]$ 的置信区间为 $[-100, 100]$ 。

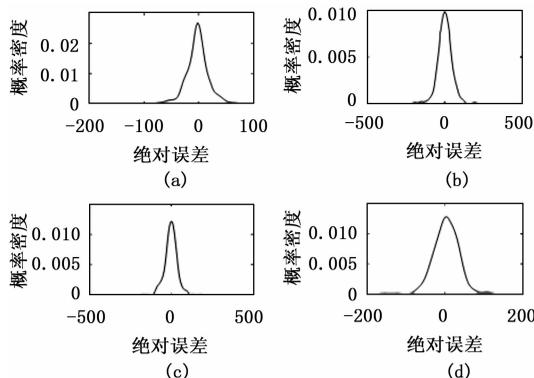


图 3 各功率区间误差样本概率密度图

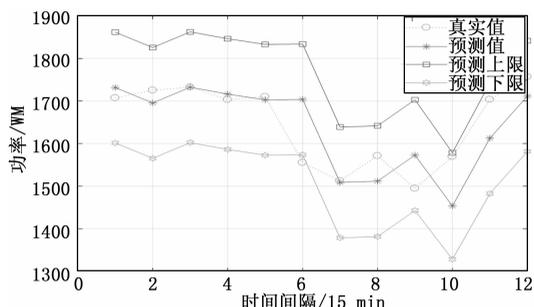


图 4 概率性预测结果 (置信度为 0.9)

给定置信度为 0.7, 则所得各分区的置信区间为:

- 1) 功率区间 $[0, 500]$ 的置信区间为 $[-52.5, 52.5]$;
- 2) 功率区间 $[500, 1000]$ 的置信区间为 $[-105, 105]$;
- 3) 功率区间 $[1000, 1500]$ 的置信区间 $[-87.5, 87.5]$;
- 4) 功率区间 $[1500, 2000]$ 的置信区间为 $[-70, 70]$ 。

由图 5 可以看出, 区间预测中置信度为 0.9 的结果中两条包络线完全将实际功率曲线包在中间, 而置信度取 0.7 时, 则有一点越过了下包络线, 这就说明合理的选取置信度十分关键。置信度的选择直接决定预测区间包括真值概率的大小。另外, 由于在不同功率区段预测的精度不同, 因此可以在不同的功率区段内可以采用不同的置信度, 从而提高概率性预测的整体性能^[11]。

3.3 对比实验

为了验证 LS-SVM 模型的预测精度, 采用 ARMA 模型和 BP 神经网络模型的预测结果进行对比。

采用 AIC (Akaike Information Criterion) 准则^[12] 确定模型阶数为 $p=6, q=10$ 。经检验一次差分后, 原先不平稳的时间序列变为平稳的时间序列, 将差分后的时间序列代入模型得到的预测结果如图 6 所示。

以上实验结果说明使用 ARMA 线性建模方法对非线性的风电功率时间序列进行预测的效果不佳。

设定学习次数为 5000, 学习速率为 0.02, 训练的误差

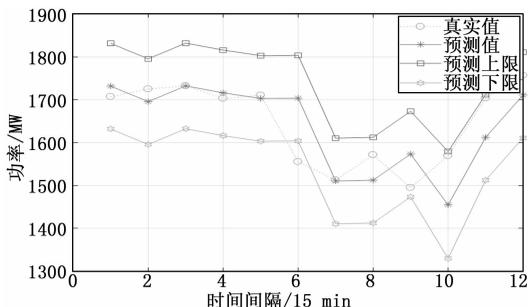


图 5 概率性预测结果 (置信度 0.7)

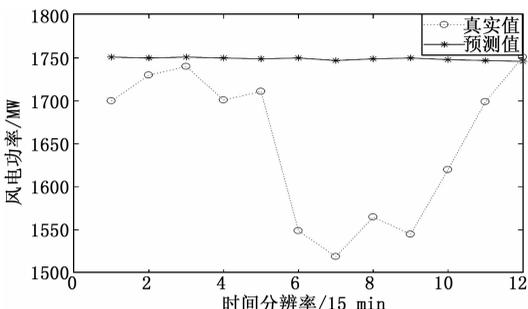


图 6 ARMA 预测模型的实验结果

精度为 0.01, 得到的 BP 神经网络预测结果^[13] 如图 7 所示。

用 LS-SVM 模型、BP 神经网络模型^[14] 和 ARMA 模型, 基于相同时段的历史数据, 预测相同的时段, 得出的性能指标可以充分地说明三者的优劣。三者预测性能对比情况如表 1 所示。

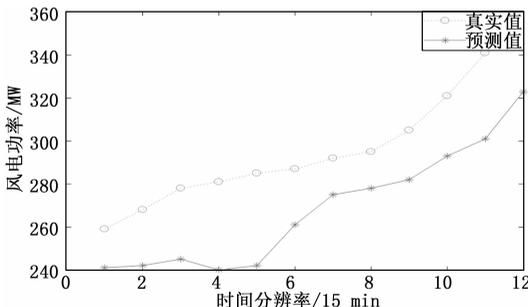


图 7 BP 神经网络预测模型的实验结果

表 1 三种预测模型性能对比

模型	MAE	NRMSE
LS-SVM	42.6388	1.008 0
ARMA	162.1562	4.094 3
BP	78.5684	2.125 3

预测结果表明, 对风电功率进行确定性预测时, LS-SVM 模型的预测精度优于其他两种模型。因此, 在 LS-SVM 预测模型的基础上进行概率性预测会得到更精确的结果。

4 结论

应用最小二乘支持向量机和核密度估计相结合的方法

实现对风电功率的概率性预测。首先利用了最小二乘支持向量机模型实现了对风电功率的单步预测,在单步预测的基础上利用迭代法实现了多步预测,多步预测的平均绝对误差为 42.6,预测效果较好,具有一定的实用价值。

但本文的研究方法还存在着一定的不足,首先在输入向量构造方面,本文采用的是经验法,选取连续的 13 个点作为输入向量,这种方法完全基于实验效果,缺乏量化标准和选取原则,后续可考虑采用其他方法来构造输入向量,例如:自相关分析方法、功率谱分析方法。

使用了非参数核密度估计法对风电功率进行概率性预测,首先对确定性预测模型预测误差进行了分布拟合,与传统的参数统计不同的是非参数统计不需要事先假设误差的分布,模型完全基于数据,提高了预测结果的可靠性。但核密度估计方法的准确性是以样本数据量为基础,其进行估计时需要大量的误差数据,实验中各区间的误差样本数量太小,无法真实地反映出实际中误差的分布情况,这样对区间预测的效果会有一定的影响,文献 [10] 中采用 Bootstrap 方法与核密度估计相结合的方法来解决这一问题,利用 Bootstrap 法对每个功率区间的误差样本值进行多次重抽样,产生大量新的误差样本,然后在新误差样本的基础上进行概率密度的估计,求各个功率区间的置信区间,这种方法为本文后续研究提供参考和借鉴。

参考文献:

- [1] Barbosa de Alencar D, de Mattos Affonso C, Limao de Oliveira R C, et al. Different Models for Forecasting Wind Power Generation: Case Study [J]. *Energies*, 2017, 10 (12): 1976.
- [2] Juban J, Siebert N, Kariniotakis G N. Probabilistic short-term wind power forecasting for the optimal management of wind generation [A]. 2017 IEEE Lausanne Power Tech. [C]. 2007: 683-688.
- [11] Low C Y, Teoh B J. Stacking-based deep neural network: deep analytic network on convolutional spectral histogram features [J]. *Multimedia University*, 2017.
- [12] Wang L, You Z H, Chen X, et al. A computational-based method for predicting drug-target interactions by using stacked autoencoder deep neural network [J]. *Journal of Computational Molecular Cell Biology*, 2017, 25 (3).
- [13] Wang Y B, You Z H, Li X, et al. Predicting protein-protein interactions from protein sequences by a stacked sparse autoencoder deep neural network [J]. *Mol. BioSyst.* 2017: 10. 1039. C7MB00188F.
- [14] Adavanne S, Virtanen T. Sound event detection using weakly labeled dataset with stacked convolutional and recurrent neural network [Z]. 2017.
- [15] 侯文擎, 叶鸣, 李巍华. 基于改进堆叠降噪自编码的滚动轴承故障分类 [J]. *机械工程学报*, 2018, 54 (7): 87-96.

- [3] 胡汉. 风电场风电功率概率预测研究 [D]. 南京: 东南大学, 2016.
- [4] Shi Z, Liang H, Dinavahi V. Direct interval forecast of uncertain wind power based on recurrent neural networks [J]. *IEEE Transactions on Sustainable Energy*, 2018, 9 (3): 1177-1187.
- [5] You L, Ming Y, Wan C, et al. A Multi-model Combination Approach for Probabilistic Wind Power Forecasting [J]. *IEEE Transactions on Sustainable Energy*, 2017, 99: 1-1.
- [6] 刘琦. 基于 LS-SVM 的时间序列预测方法及其应用研究 [D]. 哈尔滨: 哈尔滨工业大学, 2011.
- [7] 孙志华, 尹俊平, 陈菲菲, 等. 非参数与半参数统计 [M]. 清华大学出版社, 2016: 9-19.
- [8] Vasan A S S, Pecht M G. Electronic circuit health estimation through Kernel learning [J]. *IEEE Transactions on Industrial Electronics*, 2018, 65 (2): 1585-1594.
- [9] Wang S, Zhang N, Wu L, et al. Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method [J]. *Renewable Energy*, 2016, 94: 629-636.
- [10] 王晓曦. 基于非参数方法的短期风电功率预测研究 [D]. 大连: 大连理工大学, 2014.
- [11] Pang J, Liu D, Yu P, et al. Anomaly detection based on uncertainty fusion for univariate monitoring series [J]. *Measurement*, 2017, 95: 280-292.
- [12] Song Y, Li A, Quan Y. Topics' popularity prediction based on ARMA model [A]. 2018 ACM International Conference on Mathematics and Artificial Intelligence [C]. 2018: 68-72.
- [13] 隋惠惠. 基于 BP 神经网络的短期电力负荷预测的研究 [D]. 哈尔滨: 哈尔滨工业大学, 2015.
- [14] Dong S, Zhang Y, He Z, et al. Investigation of support vector machine and back propagation artificial neural network for performance prediction of the organic Rankine cycle system [J]. *Energy*, 2018, 144: 851-864.
- [16] 蔺素珍, 韩泽. 基于深度堆叠卷积神经网络的图像融合 [J]. *计算机学报*, 2017 (11): 76-88.
- [17] Li N, Lei Y, Lin J, et al. An improved exponential model for predicting remaining useful life of rolling element bearings [J]. *IEEE Transactions on Industrial Electronics*, 2015, 62 (12): 1-1.
- [18] 张西宁, 向宙, 唐春华. 一种深度卷积自编码网络及其在滚动轴承故障诊断中的应用 [J]. *西安交通大学学报*, 2018, 52 (7): 6-13, 64.
- [19] 陈仁祥, 杨星, 杨黎霞, 等. 栈式稀疏加噪自编码深度神经网络的滚动轴承损伤程度诊断 [J]. *振动与冲击*, 2017 (21): 132-138, 144.
- [20] 汤芳, 刘义伦, 龙慧. 稀疏自编码深度神经网络及其在滚动轴承故障诊断中的应用 [J]. *机械科学与技术*, 2018, 37 (3): 30-35.
- [21] 陈仁祥, 等. 加噪样本扩展深度稀疏自编码神经网络的滚动轴承寿命阶段识别 [J]. *振动工程学报*, 2017 (5).

(上接第 33 页)