

基于 CS-SVR 模型的短期风电功率预测

王小娟¹, 刘俊霞², 胡兵², 郑连清²

(1. 新疆工程学院 公共基础学院, 乌鲁木齐 830011; 2. 新疆工程学院 控制工程学院, 乌鲁木齐 830011)

摘要: 为了提高短期风电功率预测精度, 提出一种布谷鸟搜索算法 (cuckoo search algorithm, CS) 优化支持向量回归 (support vector regression, SVR) 机的预测方法, 该方法首先根据上截断点和下截断点对输入数据进行预处理, 剔除异常数据, 之后以输入数据中的风速、平均风速、风机状态等属性数据作为 SVR 算法模型的输入, 以风电功率数据作为 SVR 算法模型的输出, 建立短期风电功率的 SVR 预测模型, 针对 SVR 算法存在难以选择最优参数的缺点, 提出采用布谷鸟算法优化 SVR 参数的方法, 建立短期风电功率的 CS-SVR 预测模型; 通过与 SVR、PSO-SVR 预测模型进行了对比仿真实验, 实验结果表明, CS-SVR 预测模型具有较高的预测精度。

关键词: 功率预测; 布谷鸟搜索算法; 支持向量回归机; 参数寻优; 异常数据剔除

Short-term Wind Power Prediction Based on CS-SVR Model

Wang Xiaojuan¹, Liu Junxia², Hu Bing², Zheng Lianqing²

(1. Public Basic School, Xinjiang Institute of Electrical Engineering, Urumqi 830011, China;

2. School of Control Engineering, Xinjiang Institute of Electrical Engineering, Urumqi 830011, China)

Abstract: In order to improve the prediction accuracy of short-term wind power, a prediction method of cuckoo search algorithm (CS) to optimize support vector regression (SVR) machine is proposed. The method preprocesses the input data according to the upper cutoff point and the lower cutoff point to eliminate the abnormal data. Then, the wind speed, average wind speed, fan state and other attribute data in the input data were used as the input of the SVR algorithm model, and the wind power data was used as the output of the SVR algorithm model to establish the SVR prediction model of short-term wind power. Considering that the SVR algorithm is difficult to select the optimal parameters, a CS-SVR prediction model for short-term wind power is established by using the cuckoo algorithm to optimize the SVR parameters. Compared with SVR and PSO-SVR prediction models, the simulation experiment was conducted. Experimental results show that the CS-SVR prediction model has higher prediction accuracy.

Keywords: power prediction; cuckoo search algorithm; support vector regression machine; parameter optimization; eliminate abnormal sample data

0 引言

风能凭借储量丰富、分布广泛和清洁无污染越来越引起人们的关注。近年来, 随着我国风电装机容量的逐步增长, 风力发电规模逐渐增大, 风力发电占电网总发电量的比例逐步加大。然而因风能具有波动性、间歇性、低能量密度等特征, 导致风电功率具有同样的特征, 当大量风电并网, 会对电力系统的安全稳定运行产生严重影响, 为此需要采用适当的方法准确预测短期风电功率, 从而根据预测的风电功率对风力发电进行合理调度, 以避免风电大量并网所带来的影响。

为了准确预测风电功率, 国内外学者开展了大量研究, 对于短期风电功率的预测研究主要有物理模型法、统计模型法、智能算法模型和组合模型法^[1-2]。物理模型法需要根据风电场实际气象数据及地理物理信息预测未来的气象,

建立气象数据与风电功率的映射关系^[2-3], 具有鲁棒性强、计算时间短、不需要历史数据的支持的优点, 但是建模过程复杂, 要求对数据采集精度高。统计模型法根据风电场历史风速或者风机输出功率数据, 采用系统辨识、曲线拟合等方法预测短期风电功率, 常用的统计模型有时间序列法、灰色关联度法、聚类分析法等^[4-5]。统计模型法具有建模简单的优点, 但是该预测方法仅仅考虑历史风电功率数据或风速数据, 不考虑其他气象数据和风机运行状态数据, 预测精度受影响。智能算法模型根据某种学习算法, 通过训练、测试历史气象数据和风电数据建立输入数据与输出数据的映射关系预测短期风电功率, 常用神经网络法、支持向量机法等^[6-8]。这种预测方法不需要具体的数学表达式, 建模简单, 精确度高, 但是容易出现过学习和陷入局部最优的缺点。组合模型法综合上述方法, 采用两种以上预测方法进行等权组合或非等权组合来提高预测精度^[9-10], 具有预测精度高的优点, 但是该方法存在参数多、预测运行时间长的缺点。

上述方法分别从不同角度, 不同方面预测风电功率, 各种方法各有优缺点, 由于风电功率数据为波动性和随机性很强的非线性数据, 智能算法模型中的支持向量回归机具有良好的非线性映射能力, 鲁棒性强, 本文选取支持向

收稿日期: 2019-05-23; 修回日期: 2019-07-05。

基金项目: 新疆自然科学基金项目(2019D01A30); 新疆高校科研计划自然科学基金项目(XJEDU2018Y056); 2016年度新疆工程学院科研基金项目(2016xgy141812)。

作者简介: 王小娟(1984-), 女, 湖北孝感人, 硕士, 主要从事运筹学与控制论相关方向的研究。

量回归机预测短期风电功率, 针对支持向量回归机预测过程中会出现参数寻优难和容易陷入局部最优的缺点, 选取具有全局搜索能力和局部搜索能力强的布谷鸟算法优化模型参数, 并通过算例对比分析, 验证本文所提算法预测精度较高。

1 支持向量回归机

支持向量机回归机是一种基于统计学习理论的监督学习算法, 该算法由 Vapnic^[11] 自 2002 年提出以来得到广泛的应用, 算法假定训练集数据为 $D = \{(x_i, y_i)\}$, 每个 $x_i \in R^m$ 代表输入空间的一个样本, $y_i \in R$ 是该样本对应的输出值, 其中 $i=1, 2, L, N, m$ 为训练数据的维数, SVR 试图找到一种遵循与训练样本点相同概率分布的函数 $f(x)$, 它能够对给定训练样本之外的输入, 得到近似输出值, 其数学表达如下:

$$f(x) = \langle w, x \rangle + b \quad (1)$$

其中: $w \in R^m$ 为权值; $b \in R$ 为阈值;

实际应用过程通常为非线性, 需要一个非线性的公式, SVR 方法应用一个非线性映射函数将输入值转换为具有更高维度的特性空间, 其数学表达如下:

$$f(x) = \langle w, \phi(x) \rangle + b \quad (2)$$

其中: $\phi(x)$ 为非线性映射函数。

SVR 模型的训练需要最小化权值 w 和 b 阈值, 可以通过最小化欧几里得范数来实现。从而问题转化为一个凸优化问题, 即:

$$\min \frac{1}{2} \|w\|^2$$

$$\begin{aligned} \text{s.t.} \quad & y_i - \langle w, \phi(x) \rangle - b \leq \omega \quad i = 1, L, n \\ & \langle w, \phi(x) \rangle + b - y_i \leq \omega \quad i = 1, L, n \end{aligned} \quad (3)$$

凸优化的约束假设问题是可行的。然而实际应用过程中因为有通信、误差等因素的影响会出现错误和不可行的约束, 为此需要引入了两个松弛变量来扩大回归的容忍度。因此, 最小化问题转换为以下二次规划^[12]。

$$\min \frac{1}{2} \|w\|^2 + C \sum (\xi_i + \xi_i^*)$$

$$\begin{aligned} \text{s.t.} \quad & y_i - \langle w, \phi(x) \rangle - b \leq \omega + \xi_i \quad i = 1, L, n \\ & \langle w, \phi(x) \rangle + b - y_i \leq \omega + \xi_i^* \quad i = 1, L, n \\ & \xi_i, \xi_i^* \geq 0, i = 1, L, n \end{aligned} \quad (4)$$

式中, C 为误差惩罚系数, ϵ 为不敏感损失系数, ξ_i 和 ξ_i^* 为松弛变量。

引入拉格朗日函数, 经对偶转换得到的回归函数的表达式为:

$$f(x) = \sum_{i=1}^n (a_i - a_i^*) \langle \phi(x_i), \phi(x) \rangle + b \quad (5)$$

用核函数 $k(x_i, x)$ 代替上述等式中映射函数的点积, 重新定义 SVR 函数为:

$$f(x) = \sum_{i=1}^n (a_i - a_i^*) k(x_i, x) + b \quad (6)$$

系统选用 RBF 核函数为:

$$k(x_i, x) = \exp\left(-\frac{\|x_i - x\|^2}{2g^2}\right) \quad (7)$$

式中, g 为核函数参数。

从上述介绍中可以看出, 误差惩罚系数 C 和核函数参数 g 关系到 SVR 的预测性能, 选择合适的参数可以很好的提高 SVR 的泛化能力和预测精度。本文采用布谷鸟算法优化 SVR 参数。

2 布谷鸟算法

布谷鸟搜索算法是一种自然启发的元启发式搜索算法, 该算法由 Yang 和 Deb^[13] 于 2009 年提出, 算法基于布谷鸟的巢寄生繁殖策略和 Levy 飞行策略发展而来^[14]。Levy 飞行是指自然界中的动物或昆虫在寻找食物过程中表现出的方向随机、步长符合 Levy 分布的飞行行为, 具有频繁的短距离的局部搜索能力和少数的长距离的全局搜索能力^[15]。

布谷鸟搜索算法假定鸟巢中的卵为解, 布谷鸟卵为新解, 寻优过程为用更好的新解代替不好的解的过程, 算法对全局最优问题非常有效, 因为它维持了局部漫步与全局漫步的平衡, 这个平衡由参数 p_α 来控制, 其局部漫步定义如公式 (8)。

$$x_i^{t+1} = x_i^t + \alpha \otimes H(p_\alpha - \epsilon) \otimes (x_j^t - x_k^t) \quad (8)$$

式中, x_j^t, x_k^t 为随机选择的当前位置, α 为 Levy 飞行步长比例因子, $H(\mu)$ 为阶跃函数, ϵ 是从均匀分布中抽取的随机数, s 是步长。其全局漫步采用 Levy 飞行, 定义如公式 (9)。

$$x_i^{t+1} = x_i^t + \alpha L(s, \lambda) \quad (9)$$

$L(s, \lambda)$ 是 Levy 分布, 用于定义随机漫步的步长, 表达式如公式 (10)。

$$L(s, \lambda) = \frac{\Gamma(\lambda) \sin(\pi\lambda/2)}{\pi} \frac{1}{S^{1+\lambda}}, (s \cdot s_0 > 0) \quad (10)$$

λ 是步长的标准差, $\Gamma(\lambda) = \int_0^\infty t^{\lambda-1} e^{-t} dt$, 为阶乘函数的扩展, 当 λ 为正整数时, $\Gamma(\lambda) = (\lambda-1)!$, Levy 飞行确保系统不会被困在局部最优状态。

3 布谷鸟算法优化 SVR

本文采用布谷鸟算法具有良好的局部寻优能力和全局寻优能力对 SVR 的误差惩罚系数、核函数参数进行优化, 算法流程图如图 1 所示。

寻优步骤如下:

1) 风电功率数据预处理。对采集的风电样本数据进行异常数据剔除, 并保存处理好的样本数据。

2) 参数初始化。设置迭代次数 T 、鸟巢数量 n 、被宿主发现的概率 P_α 、设置 C 、 g 的取值范围。

3) 随机初始化鸟巢。随机生成 N 个鸟巢的位置 $X_i^{(0)} = [x_1^{(0)}, x_2^{(0)}, L, x_N^{(0)}]^T$, 所有鸟巢位置均由参数 (C, g) 组成。

4) 找最优鸟巢位置。以均方误差作为适应度函数, 计算适应度值, 即:

$$F_i = \frac{1}{n} \sum_{i=1}^n (y_i^s) \quad (11)$$

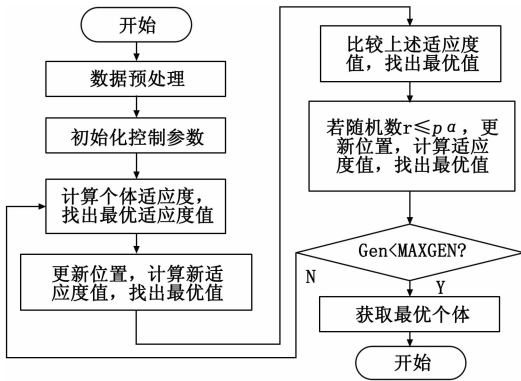


图 1 CS 算法寻优流程图

式中, n 为训练样本数, y_i^s 为风电功率预测值, y_i 实际风电功率值, 计算每个鸟巢的适应度值, 找出当前最小适应度值和对应的位置。

5) 更新鸟巢位置。用 Levy 飞行更新 N 鸟巢的位置, 判断新鸟巢适应度值与上一代适应度值, 如果新鸟巢适应度值更优, 就更新鸟巢位置, 否则保持上次位置。

6) 用 $[0, 1]$ 随机数 r 与 p_{α} 比较, 不改变 $r \leq p_{\alpha}$ 的鸟巢位置, 以随机步长改变其他鸟巢位置, 判断新鸟巢与原鸟巢的适应度值, 如果新鸟巢值更优, 则替换原鸟巢位置, 否则保留原鸟巢位置, 这样产生了最新的鸟巢位置。

7) 找出 6) 中最优鸟巢位置, 比较是否到达设置迭代次数, 到达设置迭代次数则该位置为最优位置, 否则要求从 4) 重新执行。最后得到的最优鸟巢位置为 SVR 中参数和参数的最优值。

4 试验验证

4.1 数据采集

实验数据采集来源于新疆某风电场三期工程, 该工程中共有 33 台 1.5 MW 的永磁直驱风力发电机, 实验数据为其中 1 号风力发电机每隔 10 min 采集一次的数据, 共采集 10 天 1 440 个数据。数据集详细记录了风力发电机平均功率 (kW)、最大功率 (kW)、平均风速 (m/s)、最大风速 (m/s)、电机转速 (rpm)、最高机舱温度 (°C)、最高环境温度 (°C)、最大桨叶角度 (°)、平均桨叶角度 (°)、风机状态、风速 (m/s)、功率 (kW)、发电量 (kWh)、发电时间 (h) 等共计 50 个物理因素的数据, 实验选取其中平均功率 (kW)、最大功率 (kW)、平均风速 (m/s)、最大风速 (m/s)、电机转速 (rpm)、风机状态、风速 (m/s) 七个物理因素与风电功率 (kW) 建立基于 CS-SVR 的数学模型, 并对模型的性能进行评价。

4.2 数据预处理

由于上述实验数据中存在异常数据, 采用上截断点和下截断点判断异常数据的方法对异常数据进行剔除^[16]。其数学表达式如下:

$$\begin{aligned} D_{up} &= Q_1 + 1.5R_1 \\ D_{down} &= Q_1 - 1.5R_1 \end{aligned} \quad (12)$$

式中, D_{up} 、 D_{down} 分别为上截断点和下截断点, Q_1 为下四分

位数, R_1 为四分位极差, 即上四分位数与下四分位数只差。

将实验数据分别与截断点进行比较, 大于上截断点或小于下截断点的数据均为异常数据, 需要剔除异常数据。经过异常数据剔除, 最终实验样本数据为 1 433 个, 随机选取 1 333 个数据作为训练集, 100 作为测试集, 并对训练集和测试进行归一化处理。

4.3 实验结果分析

经过数据预处理后, 分别采用 SVR、PSO-SVR 和 CS-SVR 神经网络对样本数据进行训练和测试, 并对预测结果进行对比分析, 验证所提算法具有优越性。实验过程中, 设置布谷鸟优化的迭代次数、鸟巢数量、被宿主发现的概率、寻优参数、取值范围均为 $[0.01, 1000]$ 。设置粒子群优化迭代次数为, 种群规模, 学习因子, 寻优参数、取值范围均为 $[0.01, 1000]$, 最终寻优得出的最佳参数如表 1 所示。

表 1 最佳寻优参数

模型	C	g
SVR	16	0.5
PSO-SVR	808.69	0.44
CS-SVR	1 000	0.26

在最佳参数下各模型对测试集的预测结果及对比分析如图 2~图 5 所示。

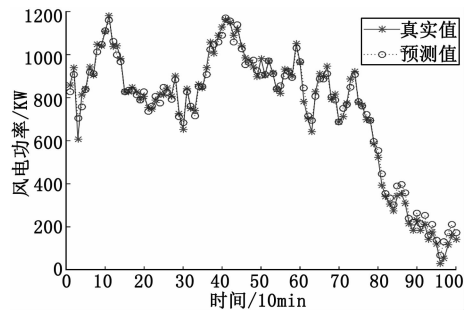


图 2 SVR 模型预测结果

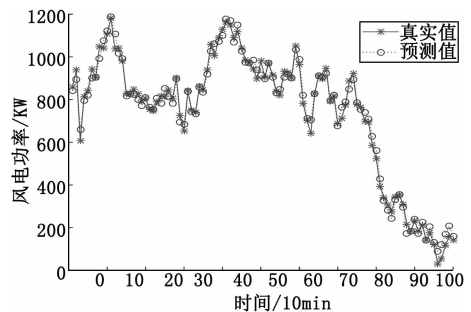


图 3 PSO-SVR 模型预测结果

通过对比图 2~5 可以看出, SVR、PSO-SVR 和 CS-SVR 模型对测试集的风电功率预测结果均很准确, 这说明 SVR 算法可以很好的用于短期风电功率的预测, 预测模型在功率为 200 kW 以下的预测误差较大, 这与实际风机运行的状态紧密相关, 风机在待机或者启停过程中功率波动的变化对风电功率预测的影响较大。从图 5 的模型性能对比中可以看出, CS-SVR 模型的预测效果要优于 SVR、

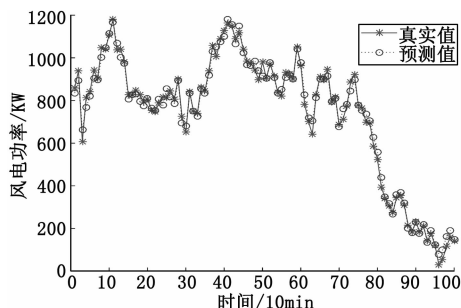


图 4 CS-SVR 模型预测结果

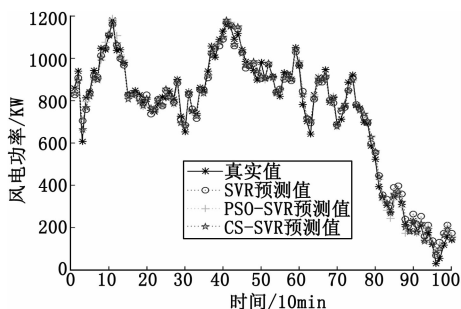


图 5 SVR、PSO-SVR 和 CS-SVR 模型的性能对比

PSO-SVR 模型, 说明 CS 算法具有良好的局部寻优和全局寻优能力。

为了更加直观的分析模型的预测精度, 实验过程中分别计算了均方根误差 (RMSE)、平均绝对误差 (MAE)、平均绝对百分比误差 (MAPE), 三者数学表达式如下:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \hat{e}_i)^2} \times 100\% \quad (13)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_i - \hat{e}_i}{x_i} \right| \times 100\% \quad (14)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - \hat{e}_i| \times 100\% \quad (15)$$

式中, x_i 和 \hat{e}_i 分别为样本数据的实际值和预测值, 3 种预测模型误差的评价结果见表 2 所示。

表 2 预测模型的预测误差比较

模型	RMSE	MAE	MAPE
SVR	4.26	3.43	6.25
PSO-SVR	3.97	3.10	2.33
CS-SVR	3.62	2.84	3.14

从表 2 中可以看出, PSO-SVR 模型的 RMSE、MAE、MAPE 三种误差的数据分别为 3.97%、3.10% 和 2.33%, CS-SVR 模型的 RMSE、MAE、MAPE 三种误差的数据分别为 3.62%、2.84% 和 3.14%, 其预测误差均小于 SVR 的预测误差, 说明参数优化具有明显提高算法精度的效果。对比 CS-SVR 和 PSO-SVR 模型的预测误差, CS-SVR 预测模型的均方根误差 (RMSE) 比 PSO-SVR 模型的均方根误差 (RMSE) 减少 8.8%。CS-SVR 预测模型的平均绝对误差 (MAE) 比 PSO-SVR 模型的平均绝对误差 (MAE) 减少 8.3%。说明 CS-SVR 预测性能优于 PSO-

SVR 模型, 因此本文所提预测模型具有较高的预测精度。

5 结论

1) 通过 SVR、PSO-SVR 和 CS-SVR 模型对测试集的短期风电功率进行预测, 实验验证结果表明 SVR 算法可以很好的用于风电功率预测, 采用其它启发式算法对 SVR 参数进行优化可以明显提高预测的精度。

2) 通过对比 PSO-SVR 和 CS-SVR 模型对测试集的短期风电功率的预测结果, 验证了 CS 算法的寻优能力优于 PSO 算法, 验证了本文所提预测模型具有较高的预测精度。

3) 在对输入样本数据进行预处理时引入了上截断点和下截断点的异常数据判别方法, 有效剔除了无效的样本数据, 为了进一步提高预测精度, 可以加强对输入样本数据预处理方面的研究。

参考文献:

- [1] 黎静华, 桑川川, 甘一夫, 等. 风电功率预测技术研究综述 [J]. 现代电力, 2017 (3): 1-11.
- [2] 冯双磊, 王伟胜, 刘纯, 等. 风电场功率预测物理方法研究 [J]. 中国电机工程学报, 2010 (2): 1-6.
- [3] 王丽婕, 冬雷, 高爽. 基于多位置 NWP 与主成分分析的风电功率短期预测 [J]. 电工技术学报, 2015 (5): 79-84.
- [4] 王聪, 杨建斌, 邓颖. 数学建模方法在风电功率预测中的应用 [J]. 太阳能学报, 2015 (5): 1081-1087.
- [5] Chen H, Wan Q, Li F, et al. GARCH in mean type models for wind power forecasting [A]. Power and Energy Society General Meeting [C]. 2013.
- [6] Xu L, Mao J. Short-term wind power forecasting based on Elman neural network with particle swarm optimization [A]. Control and Decision Conference [C]. 2016.
- [7] Devi M R, Sridevi S, Devi M R, et al. Short-term wind power forecasting and predominant wind direction using SVM kernel function [J]. International Journal of Civil Engineering & Technology, 2017, 8 (7): 256-263.
- [8] 曹天行, 刘三明, 王致杰, 等. 基于集合经验模态和深浅层学习组合的风电场功率短期预测研究 [J]. 电测与仪表, 2018 (13): 84-88.
- [9] 张露. 风电功率组合预测模型的研究 [D]. 重庆: 重庆大学, 2015.
- [10] 李玲玲, 许亚惠, 田晓越, 等. 基于组合模型的风电功率短期预测 [J]. 电工技术学报, 2014, 29 (S1): 475-480.
- [11] Vapnik V N. The nature of statistical learning theory [J]. Technometrics, 2002, 8 (6): 1564.
- [12] 赫卫国, 郝向军, 郭雅娟, 等. 基于 ARIMA 和 SVR 的光伏电站超短期功率预测 [J]. 广东电力, 2017, 30 (8): 32-37.
- [13] Yang X S, Deb S. Cuckoo Search via Lévy flights [A]. Nature & Biologically Inspired Computing, 2009. NaBIC 2009. World Congress on [C]. 2010.
- [14] 战非, 张少茹. 适应云计算的混沌布谷鸟算法应用优化研究 [J]. 控制工程, 2017 (7): 1486-1492.
- [15] 严小飞, 叶东毅. 基于 Levy 飞行的改进菌群觅食算法 [J]. 计算机系统应用, 2015 (3): 124-132.
- [16] 李柏年, 吴礼斌. MATLAB 数据分析方法 [M]. 北京: 机械工业出版社, 2012.