

基于 GA 优化的加权 LSSVM 短期风速预测^①

梁 涛^② 孙天一^③ 邹继行 侯振国

(河北工业大学控制科学与工程学院 天津 300130)

摘 要 精确的风电场风速预测可以提高风力发电在电力市场中的竞争力,还可提高电力系统的稳定性。为了降低短期风速预测的误差,提出了基于遗传算法(GA)优化加权最小二乘支持向量机(WLSSVM)的短期风速预测模型。该模型以风场实测风速数据作为模型的输入向量,根据遗传算法对加权最小二乘支持向量机的惩罚参数和核函数参数寻找最优解,以此建立起参数最优的风速预测模型。该模型用于研究某风电场同一季节连续的 300 个(采样间隔 1h)历史风速数据,取前 240 个数据为训练集,后 60 个数据为预测集,预测结果的平均绝对百分比误差仅为 11.88%。与只采用最小二乘支持向量机(LSSVM)进行预测的模型对比,该模型预测精度较高。

关键词 风电场, 风速预测, 遗传算法(GA), 加权最小二乘支持向量机(WLSSVM), 参数最优

0 引 言

近几年,在国家政府的大力扶持下,风力发电技术迅速崛起。据统计,截至 2015 年底,全国风电新增核准容量为 40 030 MW,比 2014 年增长了 23%^[1]。然而由于风能的间歇性、随机性和波动性使得风电产业迅速发展的同时也带来了一些问题:风能的利用率低,稳定性差,电网调度困难等^[2],所以对风速进行准确的预测能很好地使电力公司有效地利用风力发电以及保证风电场的安全运行^[3]。

常用的短期风速预测方法有持续预测法^[4]、时间序列预测法^[5]、卡尔曼滤波预测法^[6]、神经网络预测法^[7]等。这些方法虽然在风速预测方面是可行的,但是它们本身也还存在着不足和缺陷。近年来有学者提出将最小二乘支持向量机算法^[8]运用到短期风速预测中,该方法可以很好地解决小样本非线性等问题,但是这种算法受惩罚参数和核函数

参数的影响较大,且丧失了支持向量机本身具有的稀疏性。

基于此,有学者采用了复合预测法,将两种或两种以上的方法进行组合预测^[9]。文献[10]提出了基于自回归滑动平均模型和最小二乘支持向量机结合的风速预测方法。文献[11]提出了基于小波分解与支持向量机(wavelet decomposition-support vector machine,WD-SVM)的风速组合预测方法,并将该方法应用于提前 4 小时的风速预测。文献[12]提出了一种基于聚类经验模态分解(clustering empirical mode decomposition,EEMD)与最小二乘支持向量机(least square support vector machine,LSSVM)相结合的预测算法。文献[13]提出了基于快速独立分量分析算法(FastICA)和改进最小二乘支持向量机(improved LSSVM)的风速预测模型。

基于上述研究分析,本文提出了遗传算法与加权最小二乘支持向量机组合的预测模型。即依据训练集样本中的数据对模型贡献率大小赋予每个误差

① 河北省科技计划(16214510D,17214304D)和石家庄科技局(181060481A)资助项目。

② 男,1975 年生,博士,教授;研究方向:新能源及其大数据分析;E-mail: liangtao@hebut.edu.cn

③ 通讯作者,E-mail: 940901359@qq.com

(收稿日期:2017-11-26)

变量不同的权值,然后将加权最小二乘支持向量机应用在风电场短期风速预测中,并采用遗传算法对加权最小二乘支持向量机(weighted LSSVM, WLSSVM)模型的参数进行优化。

1 WLSSVM 模型算法

LSSVM 是遵循结构风险最小化(structural risk minimization, SRM)原则的核函数学习机器,它是由 SVM^[14]扩展的特殊类型。LSSVM 算法认为每一个数据点对计算过程的重要性是一致的,然而在实际中不同数据点在优化过程中的重要度却是不一样的。即每个数据对应一个权值,数据点的重要程度通过对应的权值大小体现。

事实上,风速随时间的改变有较大的随机性和波动性,不同时刻的风速数据都有一定的差异。一般,在模型预测中离预测点较近的风速点对预测值影响较大。鉴于此,本文提出了一种基于加权最小二乘支持向量机(WLSSVM)的算法。WLSSVM 算法实质是在 LSSVM 模型的目标函数中引入每个数据点的权系数 v_i , 对误差平方加权,其中, $0 < v_i < 1$, v_i 值越大表明该数据点对训练过程越重要。其具体算法为:设给定的样本训练集为 $(\mathbf{x}_i, \mathbf{y}_i)$, 则样本权值的数据集为 $\{(\mathbf{x}_i, \mathbf{y}_i, v_i)\}_{i=1}^l$, 其中 $i = 1, 2, \dots, l$, $\mathbf{x} \in R^l$, $\mathbf{y} \in R$ 。 $\mathbf{x}_i$ 为第 i 个输入向量, $\mathbf{y}_i$ 为第 i 个输出向量, $\delta < v_i < 1$ 是样本数据 $\mathbf{x}_i$ 的权重,用非线性映射 $\varphi(\cdot)$ 的原则,把样本集从输入空间 R^l 最终映射在特征空间 $\varphi(\mathbf{x}_i)$ 中。则在 $\varphi(\mathbf{x}_i)$ 中构造最优决策函数:

$$\mathbf{y} = \mathbf{w}^T \cdot \varphi(\mathbf{x}) + b \quad (1)$$

式中, $\mathbf{w}^T$ 为特征空间权系数向量, b 为偏置。由结构风险最小化原则,可将最优问题表述为

$$\begin{aligned} \min_{\mathbf{w}, b, e} J(\mathbf{w}, e) &= \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{1}{2} \gamma \sum_{i=1}^l v_i e_i^2 \\ \text{s. t. } \mathbf{y}_i &= \mathbf{w}^T \varphi(\mathbf{x}_i) + b + e_i \quad (i = 1, 2, \dots, l) \end{aligned} \quad (2)$$

为求解上述问题,需要将约束最优化问题转化为无约束最优化问题。然后以无约束的方法处理,由此引入 Lagrange 函数,将上式中的优化问题转变到对偶空间中,得到如下式子:

$$\begin{aligned} L(\mathbf{w}, b, e, \alpha) &= \frac{1}{2} \mathbf{w}^T \mathbf{w} + c \sum_{i=1}^l v_i e_i^2 \\ &\quad - \sum_{i=1}^l \alpha_i [\mathbf{w}^T \varphi(\mathbf{x}_i) + b + e_i - \mathbf{y}_i] \end{aligned} \quad (3)$$

式中, $e_i = \alpha_i / \gamma$, α_i 为 Lagrange 乘子, $v_i e_i$ 为不同权值的误差平方项, γ 为调节因子。由库恩-塔克条件,分别求出 $L(\mathbf{w}, b, e, \alpha)$ 对 $\mathbf{w}$ 、 b 、 e 、 α 的偏导,可得:

$$\begin{cases} \frac{\partial L}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^l \alpha_i \varphi(\mathbf{x}_i) \\ \frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^l \alpha_i = 0 \\ \frac{\partial L}{\partial e_i} = 0 \Rightarrow \alpha_i = \gamma v_i e_i \\ \frac{\partial L}{\partial \alpha_i} = 0 \Rightarrow \mathbf{w}^T \varphi(\mathbf{x}_i) + b + e_i - \mathbf{y}_i = 0 \end{cases} \quad (4)$$

消去参数 $\mathbf{w}$ 和 e_i , 则可以得到:

$$\begin{bmatrix} 0 & \mathbf{e}_1^T \\ \mathbf{e}_1 & \mathbf{Q} + \gamma^{-1} \mathbf{V} \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix} \quad (5)$$

其中, $\mathbf{e}_1 = [1, 2, \dots, l]$, $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_l]^T$, 对角阵 $\mathbf{V} = \text{diag}[v_1^{-1}, v_2^{-1}, \dots, v_l^{-1}]$, $\mathbf{y} = [y_1, y_2, \dots, y_l]^T$, $\mathbf{Q} = \varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}_i)$ 。由正定矩阵定理可以用核函数 $k(\mathbf{x}_i, \mathbf{x}_j)$ 代替 $\mathbf{Q} = \varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}_i)$ 的内积。由式(5)可以求出系数矩阵 α 和 b :

$$\begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{e}_1^T \\ \mathbf{e}_1 & \mathbf{Q} + \gamma^{-1} \mathbf{V} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix} \quad (6)$$

由上述得到的 α 和 b 求出误差变量的权值 v_i , 然后再重新组合加权最小二乘支持向量机的目标函数,多次迭代求解,最终得到稳定的 α 值。

综上所述,可以得到训练集的 WLSSVM 非线性逼近函数为

$$f(\mathbf{x}) = \sum_{i=1}^l \alpha_i k(\mathbf{x}_i, \mathbf{x}_j) + b \quad (7)$$

常用的核函数种类有:

线性核函数(linear kernel): $k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$;

多项式核函数(polynomial kernel): $k(\mathbf{x}_i, \mathbf{x}_j) = [(\mathbf{x}_i \cdot \mathbf{x}_j) + 1]^q$, 其中 q 为多项式的阶次;

两层感知器核函数(sigmoid tanh): $k(\mathbf{x}_i, \mathbf{x}_j) = \tanh[\nu(\mathbf{x}_i \cdot \mathbf{x}_j) + c]$, 其中 ν, c 为该核函数的参数;

径向基函数(radical basic function, RBF): $k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right)$, 其中 σ 为核函数的参数。

在上述核函数中, 径向基核函数(RBF)结构简单且泛化能力好。因此, 本模型的核函数选择的是径向基函数。

2 遗传算法

遗传算法(genetic algorithm, GA)是由美国 Holland^[15]基于达尔文生物进化论的自然选择和遗传学机理提出的一种寻求最优解的算法。遗传算法具有实用、鲁棒性强、高效等特点, 在很多领域都已得到很好应用。它的主要过程是将待解决问题的解(称为“染色体”)进行编码, 常见个体编码方式有两种, 分别是二进制编码和实数编码, 编码的实质是把解空间映射到染色体空间。随后在这些解空间中产生合理的初始种群, 根据适应度函数以及遗传中的选择(selection)、交叉(crossover)和变异(mutation)的操作对个体进行选择。保留适应度值高的个体, 反之, 剔除该个体。这样新产生的后代之中保留了上一代优点, 且又有上一代没有的优点, 多次循环, 直至得到最优解。

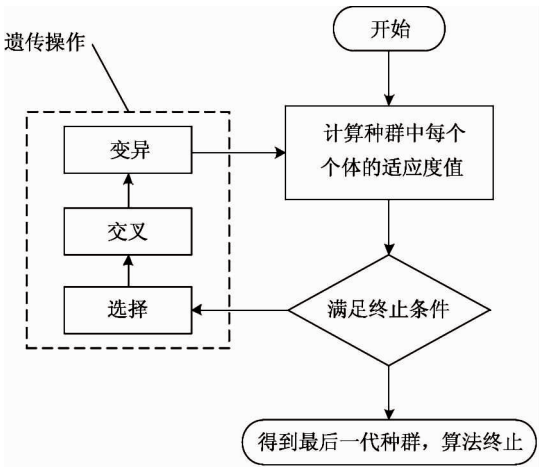


图 1 GA 基本流程图

3 风速预测模型的建立

鉴于上述模型优点, 本文提出了基于遗传算法

优化加权最小二乘支持向量机(GA-WLSSVM)的短期风速预测模型。该模型主要包括 3 部分: 数据预处理、GA 参数寻优化及模型预测, 如图 2 所示。

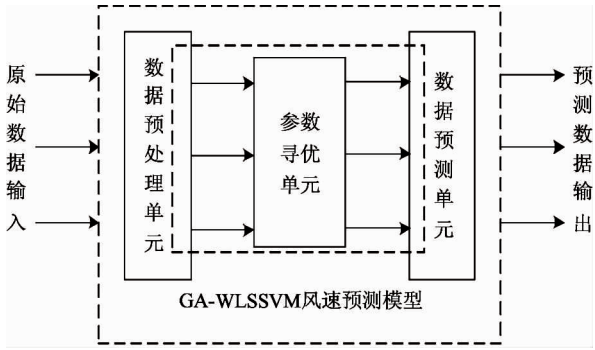


图 2 模型示意图

3.1 数据预处理

样本数据的偏大偏小或对预测精度有很大的影响, 因此, 首先需要对样本做归一化预处理。归一化的映射关系为

$$f: \mathbf{x} \rightarrow \mathbf{y} = \frac{\mathbf{x} - \mathbf{x}_{\min}}{\mathbf{x}_{\max} - \mathbf{x}_{\min}} \quad (8)$$

其中, $\mathbf{x}, \mathbf{y} \in R^n$, $\mathbf{x}_{\min} = \min(\mathbf{x})$, $\mathbf{x}_{\max} = \max(\mathbf{x})$, 将原始数据映射到 $[0, 1]$, 即 $y_i \in [0, 1], i = 0, 1, \dots, n$ 。

由于风速随时间的改变波动性较大, 前一刻对后一刻的影响最大。另一方面, 在对风速预测时, 训练样本与预测样本的相似程度越高, 那么它在预测时的贡献率也越大。因此, 出于对这两方面的考虑, 则各训练样本的权重 v_i 表示为

$$v_i = r v_{T,i} + (1 - r) v_{D,i} \quad i = 1, 2, \dots, N \quad (9)$$

式中, $v_{T,i}$ 表示只有时间影响下训练样本权重值; $v_{D,i}$ 表示只有相似度影响下训练样本权重值; r 表示常数, $0 \leq r \leq 1$, 作用是调节 $v_{T,i}$ 和 $v_{D,i}$ 。

随着训练样本离预测样本越来越远, 它的权值会逐渐减小。所以, 权值 $v_{T,i}$ 可以线性地表示为

$$v_{T,i} = \beta + (1 - \beta) \frac{i - 1}{N - 1} \quad (10)$$

式中, β 表示与预测样本时间相隔最远样本的初始权值, $0 \leq \beta \leq 1$ 。

对于相似度权值 $v_{D,i}$, 本文把欧氏距离作为衡量的标准。设输入向量 $\mathbf{x}_i$ 和预测向量 $\hat{\mathbf{x}}_i$ 之间的欧氏距离为

$$v_{D,i} = \sqrt{\sum_{j=1}^N (x_i, \hat{x}_i)^2} \quad (11)$$

由式(9)、(10)、(11)这三个表达式可以确定出训练样本的权值。

3.2 GA 寻优 WLSSVM 参数

采用 WLSSVM 模型进行预测时,欲想得到理想的预测效果,必须选择合适的惩罚参数 γ 和核函数参数 σ [16]。基于此,本文采用遗传算法对 WLSSVM 的 γ 和 σ 全局寻优。图 3 所示为基于 GA 对 WLSSVM 参数寻优的整体算法过程。

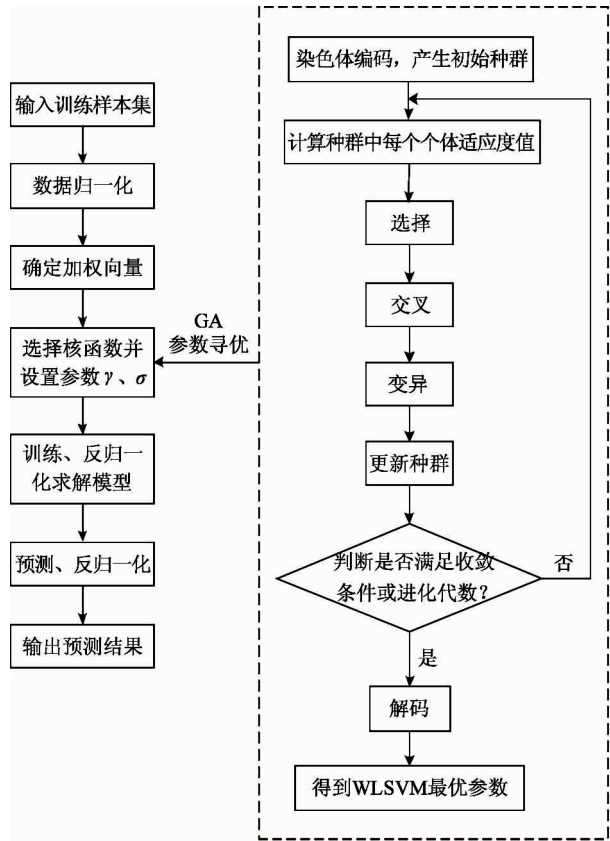


图 3 基于 GA 优化 WLSSVM 参数算法流程图

遗传优化过程的参数设置如下:

(1)遗传编码方式。设置 γ 和 σ 的寻找范围分别为 $\gamma \in (0.01, 1000)$ 和 $\sigma \in (0.01, 100)$, 为了编码方便,对参数 γ 和 σ 取对数 $\lg x$ 变换,此时参数范围为 $(-2, 3)$ 和 $(-2, 2)$ 。因此, $\lg \gamma$ 和 $\lg \sigma$ 可以用二进制编码表示,编码的长度取 10。将编码长度离散为 1023 个均匀区域,共有 1024 个不同的离散点。将 $\lg \gamma$ 和 $\lg \sigma$ 组成一个 20 位二进制编码串,即为产生的初始种群。

(2)适应度函数值。为达到最优的准确率,将预测样本交叉验证下预测值与实际值的均方误差 MSE 作为 GA 算法的适应度函数值。

$$MSE = \frac{\sum_{i=1}^n (\hat{x}_i - x_i)^2}{n} \quad (12)$$

其中 $\hat{x}_i$ 为预测值, x_i 为实际值。

(3)遗传算法运行参数,如表 1 所示。

表 1 遗传算法参数

参数	设置范围
遗传代数	300
种群大小	30
编码方式	二进制编码及 $\lg x$ 转换
选择方式	轮盘赌选择
交叉方式	均匀交叉
变异方式	多点变异
交叉概率	0.7
变异概率	0.2
γ 搜寻范围	(0.01, 1000)
σ 搜寻范围	(0.1, 100)

3.3 模型预测

把 GA 算法优化的 γ 和 σ 赋给 WLSSVM 模型,然后利用 WLSSVM 模型进行预测。

4 预测结果及分析

4.1 模型误差评价指标

误差评价指标的主要目的是检验模型的好坏以及预测结果精度的高低。这些指标包括:均方根误差(RMSE)、平均绝对误差(MAE)、绝对百分比误差(APE)以及平均绝对百分比误差(MAPE)等。本文模型以 APE 和 MAPE 这两个指标对预测结果进行判断和分析。

绝对百分比误差(APE):

$$E_{APE} = \frac{|W_R - W_P|}{W_R} \times 100\% \quad (13)$$

平均绝对百分比误差(MAPE):

$$E_{MAPE} = \frac{1}{N} \sum E_{APE} \quad (14)$$

式中, W_R 表示实际风速, W_p 表示模型预测的风速, N 为样本总数。

4.2 实例分析

本文以某风电场为例。将连续采样间隔为 1h 的 300 个历史风速数据为研究对象,如图 4 所示。由相关性分析法,可以得出,当采样时间为 1h 采集风速时, t 时刻的风速值与 $t - 7$ 个时刻的风速值有密切相关性^[17,18]。选取前 240 个观测点数据作为训练集样本,剩下后 60 个观测点为测试集样本。

因此,将测试集的 240 个数据构造样本对 $\{x_{t-7}, x_{t-6}, x_{t-5}, x_{t-4}, x_{t-3}, x_{t-2}, x_{t-1}, x_t\}$, 共有 233 组样本对,用于模型的训练过程,即训练集为 233 组,每组前 7 列风速值为输入向量,最后 1 列为模型的输出值。由式(9) ~ 式(11)可以计算样本对的权重向量,则得到第一个样本对的权重向量 $v = [0.916, 1.062, 1.282, 1.379, 1.548, 1.807, 2.006]$ 。

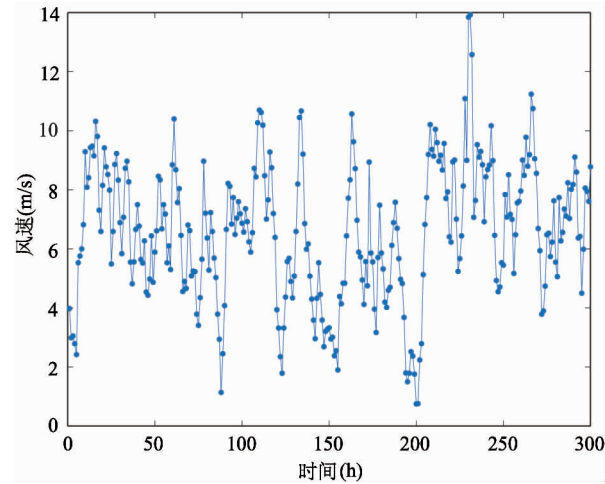


图 4 原始风速数据

对模型进行构建,图 5 为 GA 寻优 WLSSVM 过程,寻优之后的最佳惩罚参数和最佳核函数参数分别为 $\gamma = 75.098$ 、 $\sigma = 0.4672$ 。根据得到的参数对预测样本进行预测,为体现本模型的优越性,对比相同训练数据下采用 LSSVM 算法预测时的结果,如图 6 所示。从图中可以得出,本文所提出的 GA-WLSSVM 模型预测出的风速值与实际风速值基本保持一致,因此,该算法的预测能力更强。其最优值为 $\gamma = 75.098$ 、 $\sigma = 0.4672$ 。

为对比 LSSVM 和 GA-WLSSVM 两种算法的预测准确性,本文用 APE 和 MAPE 这两个误差评价指

标对各预测方法进行评估。误差结果如图 7 和表 2 所示。

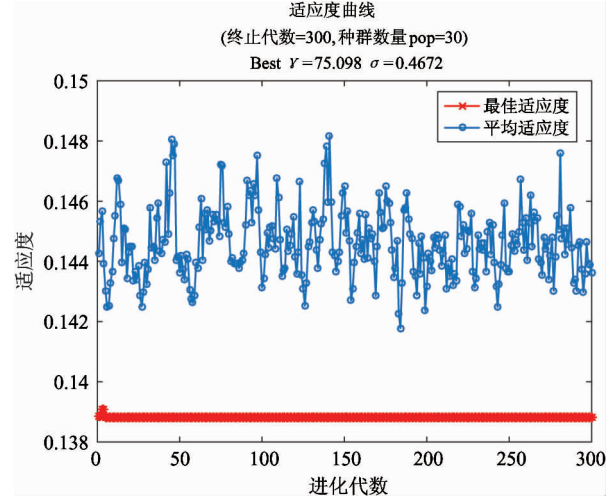


图 5 GA 寻优 WLSSVM 过程

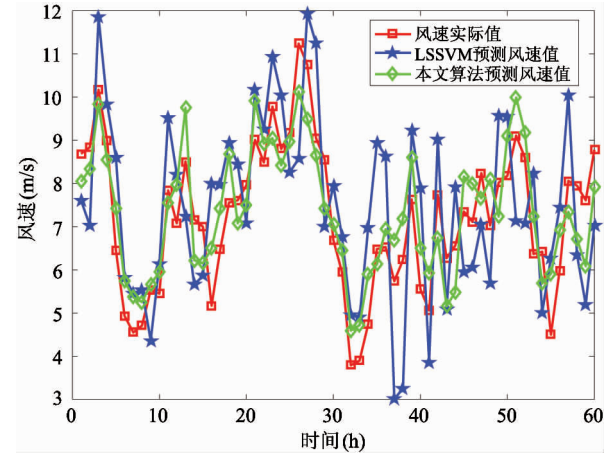


图 6 不同算法预测短期风速结果

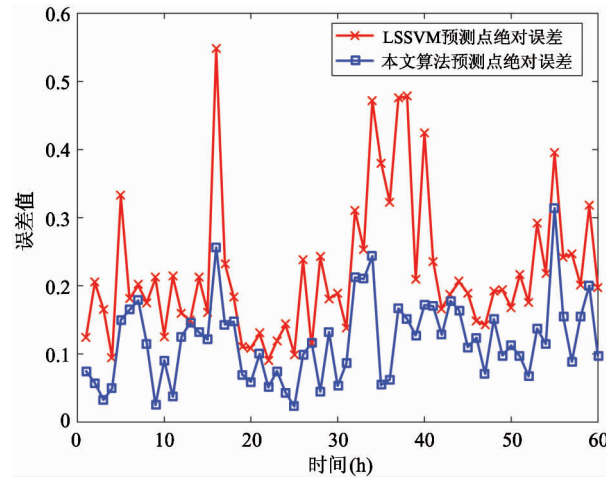


图 7 不同算法预测短期风速误差

表 2 不同算法风速预测误差比较

预测模型	相对误差		
	最大相对误差	最小相对误差	平均相对误差
LSSVM	0.5463	0.0909	0.2206
GA-WLSSVM	0.3137	0.0229	0.1188

根据图 7 和表 2 可以看出,基于 GA 算法优化 WLSSVM 的短期风速预测模型的平均相对误差为 0.1188,且比较平稳,明显优于 LSSVM 模型的预测。因此基于 GA-WLSSVM 模型的预测精度与稳定性较好。

5 结 论

由于不同历史时刻的风速对风速预测点的影响性有差异性,为了体现这种差异,本文首先对样本数据进行加权处理,通过引入每个样本的误差权重值,发挥了样本对模型的最大作用。最后,提高采用遗传算法对模型的惩罚参数 γ 和核函数参数 σ 进行选择优化,进一步提高算法的预测精度。

通过将该风速预测模型应用在某风场,结果表明,该模型在 1h 采样间隔下的短期风速预测的效果较好且具有良好的泛化性,有较高实用价值,且对不同风电场数据有普遍适应性。

参考文献

[1] 刘璟怡,王振占,殷晓斌,等. 基于全极化微波辐射计 WindSat 亮温数据的海面风场反演[J]. 高技术通讯, 2008,18(5):519-524

[2] 吴栋梁,王扬,郭创新,等. 电力市场环境考虑风电预测误差的经济调度模型[J]. 电力系统自动化, 2012, 36(6):23-28

[3] 颜晓娟,龚仁喜,张千峰. 优化遗传算法寻优的 SVM 在短期风速预测中的应用[J]. 电力系统保护与控制, 2016, 44(9): 38-42

[4] Alexiadis M, DokopoulosP, SahsamanoglouH, et al. Shortterm forecasting of wind speed and related electrical power [J]. *Solar Energy*, 1998, 63(1):61- 68

[5] Torres J L, Garcia A, Blas M D, et al. Forecast of hourly

average wind speed with ARMA model in navarre(Spain) [J]. *Solar Energy*, 2005, 79(1):65-77

[6] Bossanyi E A. Short-term wind speed using Kalman filters [J]. *Wind Engineering*, 1985, 9(1):1-7

[7] Monfared M, Rastegar H, Kojabadi H M. A new strategy for wind speed forecasting usingarti? cial intelligent methods[J]. *Renewable Energy*, 2009, 34(5):845-848

[8] 潘学萍,史宇伟,张弛. 双加权最小二乘支持向量机的短期风速预测[J]. 电力系统及其自动化学报, 2014, 26(1):13-17

[9] 张国强,张伯明. 基于组合预测的风电场风速及风电机组功率预测[J]. 电力系统自动化, 2009, 33(18): 92-95

[10] Zhao H, Li B, Zhao Z Q. Short-term wind speed forecasting simulation researchbased on ARIMA-LSSVM combination method[C]. In: *Proceedings of the International Conference on Materials for Renewable Energy & Environment*, Shanghai, China, 2011, 583-586

[11] 吴璟莉,王军伟,胡资鹏. 求解两物种小系统发育问题的遗传算法题. 高技术通讯,2015,25(3):219-227

[12] 王贺,胡志坚,张翌晖,等. 基于聚类经验模态分解和最小二乘支持向量机的短期风速组合预测[J]. 电工技术学报, 2014, 29(4):237-245

[13] 孙斌,姚海涛,李田,等. 基于 FastICA 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

[14] 田丽,曹安照,王蒙,等. 基于 SVM 和神经网络的组合预测模型物流需求预测[J]. 重庆工商大学学报: 自然科学版, 2012,29(9):61-69

[15] 张文修,梁怡. 遗传算法的教学基础[M]. 西安:西安交通大学出版社, 2000

[16] 栗然,陈倩,徐宏锐. 考虑相关因素的最小二乘支持向量机风速预测方法[J]. 电力系统保护与控制, 2010, 38(21): 146-151

[17] 方必武,刘涤尘,王波,等. 基于小波变换和改进萤火虫算法优化 LSSVM 的短期风速预测[J]. 电力系统保护与控制, 2016, 44(8): 37-43

[18] 郑丰,方必武. 基于 CFA 优化最小二乘支持向量机的短期风速预测[J]. 陕西电力:新能源, 2015, 43(6): 15-19

Short-term wind speed forecasting study based on GA optimized weighted LSSVM model

Liang Tao, Sun Tianyi, Zou Jihang, Zhang Yingjuan

(School of Control Science and Engineering, Hebei University of Technology, Tianjin 300130)

Abstract

The accurate wind farm wind speed prediction can improve the competitiveness of wind power in the electricity market and improve the stability of the power system. In order to reduce the error of short-term wind speed prediction, a short-term wind speed forecasting method based on genetic algorithm (GA) optimized weighted least square support vector machine (WLSSVM) model is proposed. In this model, genetic algorithm is employed to optimize the penalty factor and kernel parameter of weighted support vector machines, in which, the wind farm of actual measurement wind speed is taken as input vector, and then the prediction model of wind with optimal parameters is established. The model is used to study the 300 continuous historical wind speed data (sampling interval 1h) of a wind farm in the same season. The first 240 data are the training set, and the last 60 data are the prediction set. The average absolute percentage error of the prediction result is only 11.88%. The accuracy of GA-WLSSVM model is much higher than that of LSSVM.

Key words: wind farm, wind speed forecasting, genetic algorithm(GA), weighted least square support vector machine(WLSSVM), optimal parameter