

基于样本修整和支持向量机算法的并网 风电机组运行特性研究

王吉东, 许 昌, 王 欣, 韩星星, 郑 源, 刘德有

(河海大学 能源与电气学院, 南京 210098)

摘要: 针对并网风力机组运行时非线性、耦合性和大惯性的特点,提出了一种基于样本修整和支持向量机算法的系统辨识方法,并通过实例将该方法与单纯的支持向量机算法、BP(back propagation)神经网络算法进行比较.结果表明,样本修整后与修整前相比,训练速度和预测精度都有明显提高,基于样本修整和支持向量机算法的辨识方法具有明显的优越性.

关键词: 支持向量机; BP 神经网络; 系统建模; 风电

中图分类号: TK 89; O 234 **文献标志码:** A

Operating Characteristic of Wind Generation Unit Using the Method Based on Sample Modification and Support Vector Machine

WANG Ji-dong, XU Chang, WANG Xin, HAN Xing-xing, ZHENG Yuan, LIU De-you

(College of Energy and Electrical Engineering, Hohai University, Nanjing 210098, China)

Abstract: In view of that the operation characteristics of wind generation unit are nonlinear, coupled and with great inertia, a new system identification method based on sample modification and support vector machine (SVM) was put forward. By virtue of an application case, the proposed method of modified support vector machine was compared with the traditional SVM method and the back propagation (BP) neural network algorithm. The results indicate that, it is faster and more accurate than the initial one due to using modified samples. The study on the actual operation characteristics can provide reference to precise modeling and intelligent control of wind generation units.

Key words: support vector machine; BP neural network; system modeling; wind power

由于风资源的不可预测性以及机械、电气等多重因素的影响^[1],风电机组运行工况复杂多变,系统具有明显的非线性、随机性和强耦合性,因此,很难建立精确的物理模型^[2].在传统的风力机建模中,风

收稿日期: 2013-09-13

基金项目: 国家重点基础研究发展计划(973 计划)资助项目(2010CB227102-1);江苏省自然科学基金面上资助项目(2013-198);教育部留学回国人员科研启动基金资助项目(2012-940);江苏省六大人才高峰项目(2012-XNY-12)

第一作者: 王吉东(1990-),男,硕士研究生.研究方向:风电场空气动力学理论. E-mail: jidong_66@sina.com

通讯作者: 许 昌(1973-),男,教授.研究方向:风能、太阳能发电. E-mail: zhufengxu@hhu.edu.cn

能计算通常采用文献[3]中提供的风能利用系数 C_p 的经验公式,由于机组类型、参数等差异,该公式不能适用于不同机组,所以,采用经验公式建立的控制理论缺乏实用价值.而以实际运行数据为基础,通过现代智能算法探究风电机组的实际运行规律,为风力发电系统建模与控制提供了新思路.

现代智能建模领域应用最为广泛的是人工神经网络算法,而神经网络模型中最常用的是 BP(back propagation)神经网络,该方法结构简单、处理非线性问题能力强,但是,由于其基于经验风险最小化原则(ERM),预测时常出现局部误差过大的问题,对于小样本模型无法达到理想的预测精度^[4]. Vapnik 在统计学习理论上提出的支持向量机(SVM)是一类新型的机器学习算法,它基于结构风险最小化原则(SRM),克服了常规神经网络算法的缺陷,具有预测精度高、泛化能力强、能收敛于全局最优等特点^[5]. 文献[6-7]针对不同类型的训练数据,讨论了 SVM 核函数的选择方法. 文献[8]讨论了 SVM 中惩罚参数 C 和预设误差 ϵ 对回归函数的影响. 文献[9-11]分别介绍了利用双线性搜索法、网格搜索法、双线性网格搜索法、二分法及粒子群算法等对 SVM 参数进行优化的方法. SVM 算法用于回归估计时输出为实函数集,多用在非线性函数逼近、时间序列预测、过程建模和控制领域^[12-14].

本文从历史运行数据入手,针对风电机组实际运行特点,对原始样本进行分类、剔除坏点、归一化等处理,得到新的训练样本,大大增强各个输入变量和目标值之间的相关性,克服了支持向量机在处理庞大杂乱数据时速度慢、预测精度低的缺点^[15]. 通过对实际样本训练分析,比较样本修整前后和分别运用 SVM 算法与 BP 神经网络算法的训练结果,证明本文提出的方法能够缩短训练时间、具有更高的训练精度和预测精度,是一种切实有效的风电机组系统辨识方法.

1 样本预处理方法

图1是风力机标准运行曲线,据此可以看出,在运行点 B 并网以后,风电机组运行状态可分成两段:a. 额定风速以下的部分负荷区,控制风力机转速变化,跟踪最佳叶尖速比实现最大风能捕获;b. 额定风速以上的满负荷区,调节桨距角,限制风力机转速和气动功率捕获,确保发电机组输出功率稳定在额定值.

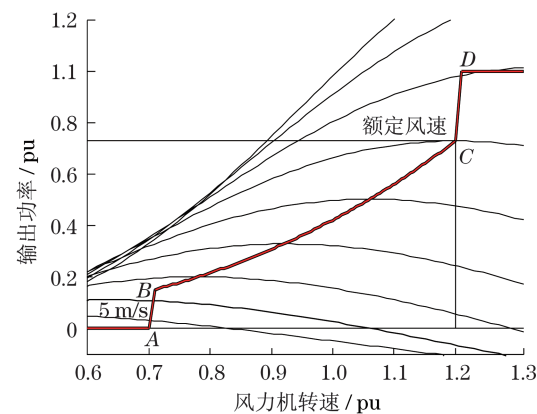


图1 风力机标准运行特性曲线

Fig.1 Standard operation characteristic curve of wind turbine

在图1中,pu 为标么值,风力机转速和输出功率分别以发电机同步转速和风力机额定功率作为基准值.

本文的研究目的是利用风电场实测数据,在低风速时,建立风力机转速给定值输出模型;在高风速时,建立实际的桨距角给定值输出模型.得到转速控制和桨距角控制的给定值,为采用智能转速控制器和桨距角控制器取代传统 PID 控制器提供理论支持.

对样本进行预处理:

a. 提取样本. 根据实际运行特性按工况提取所需样本数据,提取额定风速以下的部分负荷区数据作为样本 I,提取额定风速以上的满负荷区数据作为样本 II.

b. 剔除坏点. 训练样本包括输入和输出样本数据,输入样本 $X_k = (x_{k1}, x_{k2}, \dots, x_{ki}, \dots, x_{km})$,其中,主要变量设为 M (本文中指风速 v);输出样本 $T_h = (t_{h1}, t_{h2}, \dots, t_{hj}, \dots, t_{hm})$,根据训练目标不同,选定输出的目标变量 T (输出功率 P 、风力机转速 ω 或桨距角 β).

在样本中存在一些不合理的数据点,会严重影响训练和预测的精度,这些点大致分为两类:第一类是由于设备检测或数据记录错误等因素产生的与实际运行规律相悖的数据点;第二类是由于机组惯性或湍流风速过大等原因引起的严重偏离最优运行曲线的数据点.对两种“坏点”进行数学定义,对于样本内的第 i 个数据点:

第一类

$$\frac{T_i - T_{i-1}}{T_{i-1}} \frac{M_i - M_{i-1}}{M_{i-1}} < 0 \quad (1)$$

第二类

$$\frac{|T_{i+1} - T_i| + |T_i - T_{i-1}|}{\frac{(T_{i+1} + T_{i-1})}{2}} > \delta \quad (2)$$

式中, δ 为平整度, 可自主设定.

以江苏省某风电场额定风速以下的运行数据为例, 选择风速 v 作为样本主要属性, 输出功率 P 作为目标变量, 平整度 $\delta = 15\%$. 按照式(1)和式(2)逐个剔除样本中的“坏点”, 图 2(a)和 2(b)分别显示了样本处理前后的数据点分布情况, 比较发现, 剔除坏点后样本得到有效简化, 数据间具有更强的相关性.

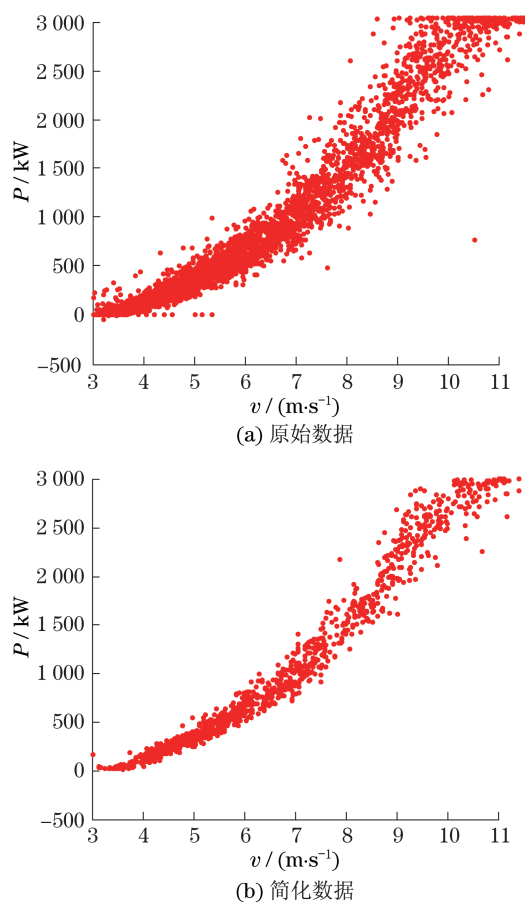


图 2 剔除坏点前后样本数据比较

Fig.2 Comparison of initial data and reduced data

c. 数据归一化处理. 为了避免各指标的数值之间相差过大影响学习训练的速度, 对历史数据进行归一化处理, 即将原始数据经过线性变化至 $[0, 1]$ 区间中, 构成训练数据集^[16].

2 支持向量机回归算法

在支持向量机回归方法中, 分为线性回归与非线性回归. 对于线性回归问题, 根据给定的训练集样

本 $\{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\}$, $x_i \in \mathbb{R}^n$, $y_i \in \mathbb{R}$, $i = 1, 2, \dots, l$. x_i 为输入量, y_i 为对应的输出量, 寻求一个反映样本数据输出和输入的最优函数关系 $f(x)$, 并且对于预测集样本的输入量 $\{x_{l+1}, x_{l+2}, \dots, x_m\}$, 能够由该函数得出较为准确的预测值 $\{y_{l+1}, y_{l+2}, \dots, y_m\}$.

函数“最优”是指按照某一确定的误差函数来计算, 所得的函数关系对样本数据集拟合的累计误差最小, SVM 算法用于回归估计采用 ϵ -insensitive 损失函数作为误差函数, 形式为

$$\left. \begin{aligned} y_i - f(x_i) &\leq \epsilon + \xi_i \\ f(x_i) - y_i &\leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* &\geq 0, i = 1, 2, \dots, n \end{aligned} \right\} \quad (3)$$

式中, ξ_i, ξ_i^* 为松弛因子.

这种误差函数中间有一个 2ϵ 的不敏感带, 当划分有误差时, ξ_i, ξ_i^* 都大于 0; 当不存在划分误差时, ξ_i, ξ_i^* 都取 0.

训练学习的过程是构造最优的线性拟合函数 $f(x) = wx + b$, 其中, $w \in \mathbb{R}^n$, $b \in \mathbb{R}$. 根据结构风险最小化原理, 将训练过程进行转化, 通过最小化目标函数来训练 w 和 b ^[17].

$$\left. \begin{aligned} \min_{w, b, \xi} \left\{ \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \right\} \\ \text{s.t. } \left. \begin{aligned} y_i - w\varphi(x_i) - b &\leq \epsilon + \xi_i \\ w\varphi(x_i) + b - y_i &\leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* &\geq 0, i = 1, 2, \dots, l \end{aligned} \right\} \end{aligned} \right\} \quad (4)$$

式(4)中第一项使拟合函数更为平坦, 以提高泛化能力; 第二项为减小误差; 惩罚参数 $C > 0$, 表示对超出预设误差 ϵ 的样本的惩罚程度, 这是一个对偶二次凸规划问题, 引入 Lagrange 函数, 将式(4)的对偶最优化问题转化为最大化目标函数

$$W(\alpha, \alpha^*) = \frac{1}{2} \sum_{i=1, j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \cdot (x_i x_j) + \sum_{i=1}^n (\alpha_i - \alpha_i^*) y_i - \sum_{i=1}^n (\alpha_i + \alpha_i^*) \epsilon \quad (5)$$

约束条件为

$$\left. \begin{aligned} \sum_{i=1}^n (\alpha_i - \alpha_i^*) &= 0 \\ 0 &\leq \alpha_i, \alpha_i^* \leq C \end{aligned} \right\} \quad (6)$$

式中, α_i, α_i^* 为 Lagrange 乘子, 解之得

$$\begin{aligned} \alpha, \alpha^* = \arg \min \left\{ \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i - \alpha_i^*) \cdot \right. \\ \left. (\alpha_j - \alpha_j^*) (x_i x_j) + \sum_{i=1}^l (\alpha_i - \alpha_i^*) y_i - \right. \\ \left. \sum_{i=1}^l (\alpha_i + \alpha_i^*) \epsilon \right\} \end{aligned} \quad (7)$$

参数 b 的求解,对于满足 $0 < \alpha_i < C (\alpha_i^* = 0)$ 条件的标准支持向量,有

$$b_1 = y_i - \sum_{j=1}^l (\alpha_j - \alpha_j^*) x_j x_i - \epsilon = y_i - \sum_{x_j \in SV} (\alpha_j - \alpha_j^*) x_j x_i - \epsilon \quad (8)$$

式中, SV 表示标准支持向量的集合,包含所有在间隔区边缘的训练样本点.

对于满足 $0 < \alpha_i^* < C (\alpha_i = 0)$ 条件的标准支持向量,有

$$b_2 = y_i - \sum_{x_i \in SV} (\alpha_j - \alpha_j^*) x_j x_i - \epsilon \quad (9)$$

通过式(8)和式(9)对所有标准支持向量分别计算 b 的值,然后求平均值,即

$$b = \frac{1}{N_{SV}} \left\{ \sum_{0 < \alpha_i < C} \left[y_i - \sum_{x_j \in SV} (\alpha_j - \alpha_j^*) x_j x_i - \epsilon \right] + \sum_{0 < \alpha_i^* < C} \left[y_i - \sum_{x_j \in SV} (\alpha_j - \alpha_j^*) x_j x_i - \epsilon \right] \right\} \quad (10)$$

式中, N_{SV} 表示标准支持向量的个数.

最终求得的线性拟合函数为

$$f(x) = wx + b = \sum_{i=1}^n (\alpha_i - \alpha_i^*) x_i x + b \quad (11)$$

对于非线性回归问题,首先通过一个非线性映射 $\Phi(x)$ 将样本数据映射到某个高维特征空间中,在高维特征空间中进行线性回归,这样将原空间中的非线性回归问题转化为高维特征空间中的线性回归问题.为了从原始空间中抽取特征,将原始空间中的样本映射为高维特征空间中的一个向量,引入的映射函数 $K(x_i, x_j)$,称为核函数^[18],核函数是连接低维与高维之间的桥梁.

常用的核函数有多项式核、S型核、RBF(radial basis function)核等^[19].本文选择 Gaussian 型 RBF^[20-21]作为SVM的核函数,其形式为 $K(x, x') = \exp(-\|x - x'\|^2 / 2\sigma^2)$,其中, x' 为核函数中心, σ 为函数的宽度参数,控制了函数的径向作用范围,通过径向基函数实现了用一元函数表示多元函数的目的.

于是,式(5)转化为

$$W(\alpha, \alpha^*) = \frac{1}{2} \sum_{i=1, j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \cdot$$

$$K(x_i, x_j) + \sum_{i=1}^n (\alpha_i - \alpha_i^*) y_i - \sum_{i=1}^n (\alpha_i + \alpha_i^*) \epsilon \quad (12)$$

约束条件与线性问题相同,这样就得到非线性拟合函数的表达式为

$$f(x) = w\Phi(x) + b = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x, x_i) + b \quad (13)$$

3 样本训练结果

以江苏省某风电场 2013 年 1 月的 10 min 运行间隔数据作为初始样本,选定机组的额定风速为 11 m/s,按工况条件提取样本 I 和 II,剔除样本 I, II 中的坏点,再进行归一化处理.用 SVM、BP 神经网络分别对预处理(DPP)后的样本和处理前的样本进行训练,选取均方根误差(RMSE)和平均相对误差(MRE)作为训练和预测效果的主要判别依据.

均方根误差

$$e_{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (F(i) - Y(i))^2} \quad (14)$$

平均相对误差

$$e_{MRE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{F(i) - Y(i)}{Y(i)} \right| \times 100\% \quad (15)$$

式中, $Y(i)$ 和 $F(i)$ 分别为实际值和预测值.

选定风速作为输入变量,即 $X = v$,输出功率作为输出变量 $Y = P$,对样本 I 进行预处理后样本点从 3 608 个缩减到了 1 831 个.由于支持向量的数目随着预设误差 ϵ 的增大而减小,根据训练过程,取 $\epsilon = 0.001$,此时支持向量数目较少,数据点的逼近精度较高.选用双线性网格搜索法,对 SVM 参数 σ^2 和 C 进行优选,得到最优参数 $\sigma^2 = 0.328 2$, $C = 1 028$.BP 神经网络采用标准的 3 层网络模式,SVM 和 BP 神经网络的输入层节点数均为 3,输出层节点数为 1,训练得到额定风速以下 $P - v$ 运行曲线,结果如图 3 和图 4 所示(见下页).

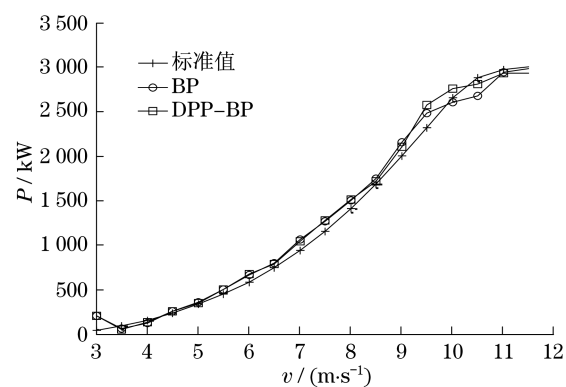


图3 BP神经网络算法下 $P - v$ 曲线预测结果

Fig.3 $P - v$ prediction curves by BP neural network algorithm

图 3 和图 4 分别为使用 BP 网络和 SVM 算法对处理前和处理后的样本进行训练,并将得到的预测曲线与标准风功率曲线进行对比,进一步计算得

到表1中的预测误差.比较图3中的BP和DPP-BP曲线以及图4中的SVM和DPP-SVM曲线可以看出,数据预处理后的样本训练得到的曲线比处理前贴合得更好;比较图3、图4和表1可以看出,与BP神经网络算法相比,运用SVM算法的训练结果更加靠近标准风功率曲线,训练误差更小并且训练时间更短.

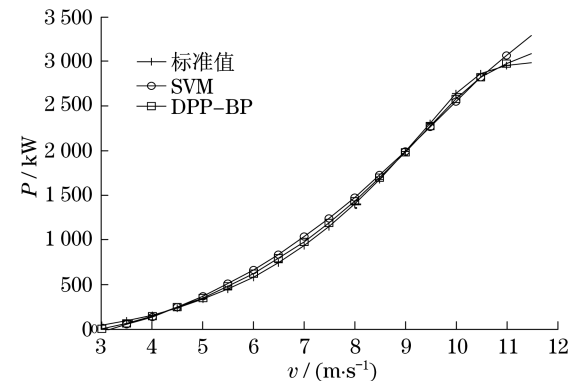


图4 SVM算法下 $P-v$ 曲线预测结果

Fig.4 $P-v$ prediction curves by SVM algorithm

表1 $P-v$ 曲线在不同算法下的预测误差和训练时间

Tab.1 Prediction error and training time of $P-v$ curves by different algorithms

算法	预测误差		训练时间/s
	$e_{\text{RMSE}}/\text{kW}$	$e_{\text{MRE}}/\%$	
BP	8.867	14.69	104
DPP-BP	4.892	9.08	58
SVM	7.482	10.99	34
DPP-SVM	2.621	3.42	18

应用上述方法,分别通过样本 I 和 II 训练得到转速-风速($\omega-v$)曲线和桨距角-风速($\beta-v$)曲线,剔除坏点后样本 I($\omega-v$)数据点从 3 573 缩减到 1 424 个,样本 II($\beta-v$)数据点从 2 108 缩减到 971 个,训练结果如图 5 和图 6 所示.

由于 $\omega-v$ 、 $\beta-v$ 标准曲线无法获得,选择训练误差作为准确性的判别依据,表 2 和表 3 分别是训练 $\omega-v$ 和 $\beta-v$ 曲线得到的训练误差和训练时间.

比较图 3 和图 4 可知,训练得到的 $P-v$ 曲线与图 1 的标准运行曲线 BC 段的走势基本一致;在图 5 中, $\omega-v$ 曲线的风速 v 在 5~8 m/s 变化时,转速和风速基本符合一次函数关系,说明在这一工况区间机组是严格按照恒叶尖速比规律运行的;在图 6 中, $\beta-v$ 曲线的风速未超过额定风速时已经开始进行变桨操作,风速较低时桨距角变化较慢,随着风

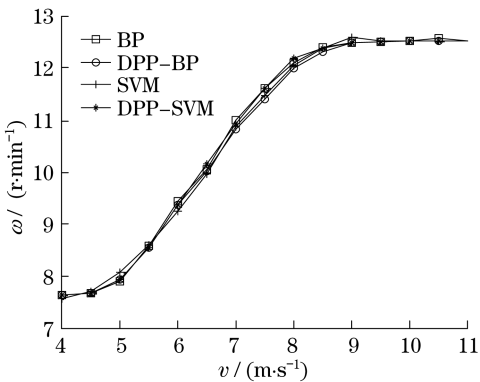


图5 不同算法下 $\omega-v$ 曲线预测结果

Fig.5 $\omega-v$ prediction curves by different algorithms

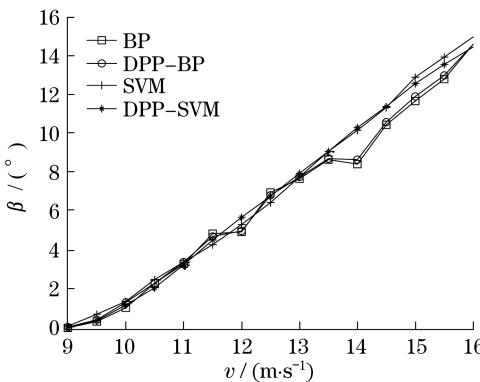


图6 不同算法下 $\beta-v$ 曲线预测结果

Fig.6 $\beta-v$ prediction curves by different algorithms

表2 $\omega-v$ 曲线在不同算法下的训练误差和训练时间

Tab.2 Training error and training time of $\omega-v$ curve by different algorithms

样本	算法	训练误差		训练时间/s
		$e_{\text{RMSE}}/(\text{r} \cdot \text{min}^{-1})$	$e_{\text{MRE}}/\%$	
A($\omega-v$)	BP	2.717	13.31	83
	DPP-BP	1.343	7.51	62
	SVM	1.221	7.46	37
	DPP-SVM	0.836	4.97	20

表3 $\beta-v$ 曲线在不同算法下的训练误差和训练时间

Tab.3 Training error and training time of $\beta-v$ curve by different algorithms

样本	算法	训练误差		训练时间/s
		$e_{\text{RMSE}}/(\text{°})$	$e_{\text{MRE}}/\%$	
B($\beta-v$)	BP	3.759	16.38	65
	DPP-BP	2.231	10.17	31
	SVM	1.928	8.69	28
	DPP-SVM	1.076	5.45	11

速增大,变桨幅度呈现先增大后减小的趋势.进一步比较表1、表2和表3中的误差和训练时间,由于计

算训练误差时用到的样本点远远多于计算预测误差用到的样本点,因此,训练误差普遍大于预测误差.比较可得本文提出的基于数据预处理的支持向量机算法的预测精度和训练精度最好,训练速度最快,并且没有出现BP神经网络中局部误差过大的问题,具有明显的优越性.

4 结 论

充分考虑了并网风电机组的运行特性,对实测大样本进行数据预处理,分别使用SVM和BP神经网络方法对修整后的样本进行训练.结果表明,在非线形曲线拟合方面,基于结构风险最小化原则的SVM算法致力于寻找全局最优解,而基于经验风险最小化原则的BP神经网络算法容易出现局部误差过大问题.本文提出的基于样本预处理方法和SVM算法的非线形拟合方法在预测精度和训练速度上都存在明显优势.由于本方法原理简单,易于实现,在风电机组运行特性研究领域具有很强的实用性.

参考文献:

- [1] 李辰奇,许昌,杨建川,等.基于CFD的复杂地形风能分布研究[J].上海理工大学学报,2013,35(3):270-274.
- [2] Hippert H S, Pedreira C E, Souza R C. Neural networks for short-term load forecasting: A review and evaluation[J]. IEEE Trans, Power Systems, 2001, 16(1): 44-54.
- [3] Hand M M. Variable-speed wind turbine controller systematic design methodology: a comparison of nonlinear and linear model-based designs[R]. Colorado: National Renewable Energy Laboratory, 1999.
- [4] Vapnik V N. Statistical learning theory [M]. New York: Wiley, 1998.
- [5] 李盼池,许少华.支持向量机在模式识别中的核函数特性分析[J].计算机工程与设计,2005,26(2):302-304.
- [6] 周洪利,刘培玉.支持向量机中的模型选择研究[J].信息技术与信息化,2006(6):61-63.
- [7] 王鹏,朱小燕.基于RBF核的SVM的模型选择及其应用[J].计算机工程与应用,2003,39(24):72-73.
- [8] 刘东辉,卞建鹏,付平,等.支持向量机最优参数选择的研究[J].河北科技大学学报,2009,30(1):58-61.
- [9] 赵璐华,彭涛.一种有效的SVM参数优化选择方法[J].制造业自动化,2010,32(9):146-149.
- [10] 徐海龙,王晓丹,廖勇,等.一种基于PSO的RBF-SVM模型优化新方法[J].控制与决策,2010,25(3):367-370.
- [11] 贺媛媛.基于支持向量机的电力系统短期负荷预测模型研究[D].大连:大连理工大学,2005.
- [12] 牛东晓,谷志红,刑棉,等.基于数据挖掘的SVM短期负荷预测方法研究[J].中国电机工程学报,2006,26(18):6-12.
- [13] 徐勤兰,樊重俊,张鹏.灰色支持向量机在机场吞吐量预测中的应用[J].上海理工大学学报,2012,34(4):347-350.
- [14] 靳忠伟,陈康民,闫伟,等.基于支持向量机的中长期电力负荷预测研究与应用[J].上海理工大学学报,2008,30(2):129-132.
- [15] 崔瑜琦.基于最小二乘支持向量机风电机组的预测研究[D].北京:华北电力大学,2010.
- [16] Gunn S R. Support vector machines for classification and regression [R]. Southampton: University of Southampton, 1998.
- [17] Atthajariyakul S, Leephakpreeda T. Neural computing thermal comfort index for HVAC systems[J]. Energy Conversion and Management, 2005, 46(15/16): 2553-2565.
- [18] 谢宏,魏江平,刘鹤立.短期负荷预测中支持向量机模型的参数选取和优化方法[J].中国电机工程学报,2006,26(22):17-22.
- [19] Wendland H. Piecewise polynomial, positive definite and compactly supported radial basis functions of minimal degree[J]. Adv Comput Math, 1995, 4(1): 389-396.
- [20] Wu Z M. Generalized Bochner's Theorem for radial function[J]. Approximation Theory and its Applications, 1997, 13(3): 47-57.
- [21] Lu Y W, Sundararajan N, Saratchandran P. Performance evaluation of a sequential minimal radial basis function (RBF) neural network learning algorithm[J]. IEEE Trans, Neural Networks, 1998, 9(2): 308-318.

(编辑:石 瑛)