

◇ 研究报告 ◇

基于数据融合的风电机组噪声预测

张智勇¹ 余金^{2†} 常鹏² 徐其丹² 李阳²

(1 喀什大学物理与电气工程学院 喀什 844008)

(2 国网新疆电力有限公司经济技术研究院 乌鲁木齐 830049)

摘要 根据风电机组噪声信号检测复杂的情况,研究风电机组非声学参数的信息熵特征,对机组噪声进行多源数据融合预测。分析基于信息熵的非声学参数的特征提取方法,并对传统的基于遗传算法的支持向量机回归的缺陷提出改进,结合实际应用的非声学参数的信息熵特点平衡遗传算法(GA)的终止条件。通过统计分析完成了输入变量的筛选,去除了对预测影响较大的共线性因素,并实现了输入降维提高预测精度和速率。最后应用数据的信息熵特征,训练改进的遗传算法的支持向量机回归,建立最终的多源数据特征级融合预测模型。通过对比表明基于多源数据融合的预测方法精度最高,预测结果的相对误差平均值为0.7757%,具有实际可行性。

关键词 风力发电,噪声预测,多源数据融合,支持向量机回归,遗传算法

中图法分类号: TM614

文献标识码: A

文章编号: 1000-310X(2018)06-0956-07

DOI: 10.11684/j.issn.1000-310X.2018.06.018

Wind turbines noise prediction based on data fusion

ZHANG Zhiyong¹ YU Jin² CHANG Peng² XU Qidan² LI Yang²

(1 School of Electrical Engineering, Kashgar University, Kashi 844008, China)

(2 State Grid Economic and Technical Research Institute Co., LTD. of Xinjiang, Urumqi 830049, China)

Abstract According to the complex situation of wind turbine noise signal detection, the information entropy characteristics of non-acoustic parameters of wind turbine are analyzed and the noise is predicted by multi-source data fusion. The feature extraction method based on entropy of non-acoustic parameters is analyzed, and the traditional method based on the support vector regression of genetic algorithm (GA-SVR) is improved. While the termination condition of the genetic algorithm is balanced by the information entropy characteristics of the non-acoustic parameters. What's more, the input variables are filtered out through the statistical analysis, removing the collinear factors that have a greater impact on the forecast and realizing the variable dimension reduction to improve the prediction accuracy and speed. Finally, the information entropy features of the data are applied to train the improved GA-SVR to establish the final prediction model of multi-source data feature level fusion. The comparison shows that the accuracy of the prediction method based on multi-source data fusion is the highest, and the relative error of the predicted results is 0.7757%, which is feasible.

Key words Wind power, Noise prediction, Multi-source data fusion, Support vector regression, Genetic algorithm

2018-01-27 收稿; 2018-05-01 定稿

作者简介: 张智勇(1989-), 男, 内蒙古人, 硕士研究生, 研究方向: 可再生能源及其控制技术。

† 通讯作者 E-mail: 21858309@qq.com

1 引言

随着全球范围内可再生能源的大力推广与发展,我国作为能源大国对可再生能源的发展和提倡不断加强。风能作为清洁可再生能源被大力应用于风力发电,我国在风力发电方面也投入巨大。随着风电机组发展的大型化、大量化,其叶片转动产生的噪声也日益增大。国内外学者通过研究发现,风电机组辐射出的噪声信号中夹杂有机组运行时大量信息,因此,对风电机组噪声信号的研究已成为国内外学者研究的热点^[1-3]。同时,风电机组所辐射出的噪声对周边居民也产生了影响^[4],因此,进行风电机组噪声信号的预测研究就具有很大的现实意义。

目前,在噪声预测及分析研究方面的方法有多元回归预测^[5-6]、小波分析预测^[7-8]、神经网络预测^[9-10]和支持向量机回归预测^[11-12]等。但是,对于风电机组噪声信号的预测方法应用较少。通过研究发现风电机组噪声信号具有非线性、非平稳、复杂时变的特点,因此利用信息熵理论对风机噪声信号进行特征提取,再应用多元线性回归、多源数据融合理论、改进的基于遗传算法的支持向量机回归(Support vector regression of genetic algorithm, GA-SVR)三者结合的方法,对风电机组噪声信号进行声压级预测。以回归分析为基础,对预测样本进行回归分析,诊断输入变量中存在的共线性问题,并实现完成变量的筛选和降维提高预测精度。通过实测数据验证比较后表明,应用多源数据融合理论与改进的GA-SVR相结合的模型,得到的预测结果更为精确。

2 噪声预测方法

风电机组噪声检测技术标准 IEC 61400-11 中规定用于测量评估的非声学参数有轮毂风速、变桨角度、输出功率等,同时标准中也规范了风机噪声声压级的检测^[13]。在风机噪声检测过程中,需要用测量仪进行反复、长期地采样收集,在这个周期里的噪声采样数据极易受外界干扰,从而影响数据有效性。为了有效地进行风电机组运行分析,研究噪声的实时预测就十分重要。

2.1 多元线性回归诊断

多元线性回归是一种最基本的回归预测方法,该法通过数理统计算法估算出单一因变量与多自

变量间的相对线性关系,由此得到变量间的预测模型^[14]。

在得到回归方程后,方程中回归系数的绝对值大小是判定自变量对因变量影响的重要指标。但如果自变量间存在较严重的共线性时,得到的回归方程的准确性和有效性将受到严重的影响。为此,可通过多元线性回归诊断的方法对共线性进行判定和处理。常用的方法有变量筛选法、岭回归估计、主成分回归和偏最小二乘回归,其中只有变量筛选法具有降维处理的效果^[15]。

在实测数据中,也可以应用多元线性回归异常点诊断的方法,对数据进行优化处理以实现样本数据的优化。较常用的判定指标有计算样本点的学生化残差和 COOKD 统计量^[16]。

2.2 信息熵

设 \mathbf{M} 是由可被观察的集合 \mathbf{H} 所构成的勒贝格空间 (Lebesgue space), 并设空间测度为 μ , 且有 $\mu(\mathbf{M}) = 1$ 。设空间 \mathbf{M} 可被 $\mathbf{A} = (\mathbf{A}_i)$ 有限划分表示为其互不相容的集合, 即 $\mathbf{M} = \bigcup_{i=1}^n \mathbf{A}_i$, 且有 $\mathbf{A}_i \cap \mathbf{A}_j = 0, \forall i \neq j$, 其中 \mathbf{A}_i/\mathbf{M} 是划分后任意子集 \mathbf{A}_i 所得配额。由此条件, 即得出 \mathbf{M} 的信息熵:

$$S(\mathbf{A}) = - \sum_{i=1}^n \mu(\mathbf{A}_i) \lg \mu(\mathbf{A}_i), \quad (1)$$

式(1)中, $\mu(\mathbf{A}_i) = \mathbf{A}_i/\mathbf{M}$ 是任意子集 \mathbf{A}_i 的测度, $i = 1, 2, \dots, n$ 。

通过研究表明,利用奇异谱熵、功率谱熵、小波能谱熵和小波空间谱熵四种信息熵对噪声进行特征量提取,可以较完整地保留信号信息^[17]。

奇异谱熵可以对风机噪声信号进行时域的特征提取。设有空间 \mathbf{A} 为 $M \times N$ 的矩阵, 对其完成奇异分解可得对应的奇异值谱 $\{\sigma_i\}$, $1 \leq i \leq M$, 对这些值进行一定方式的比例划分, 遵循 $p_i = \sigma_i / \sum_{i=1}^M \sigma_i$, 则可定义得奇异谱熵 H_t :

$$H_t = - \sum_{i=1}^M p_i \lg p_i. \quad (2)$$

功率谱熵可以对风机噪声信号进行频域的特征提取。设 $S = \{S_1, S_2, \dots, S_N\}$ 为信号在频域空间的能量划分, 并对在此划分下的功率谱概率进行 $q_i = S_i / \sum_{i=1}^N S_i$ 的计算。最后, 对这些划分进行功率

谱熵 H_f 计算:

$$H_f = - \sum_{i=1}^N q_i \lg q_i. \quad (3)$$

小波能谱熵和小波空间谱熵则可以对风机噪声信息进行基于时频域的特征提取。假设能量函数 $f(t)$ 满足小波变换后的能量守恒定律, 则:

$$\int_{-\infty}^{+\infty} |f(t)|^2 dt = \frac{1}{c_\psi} \int_0^{+\infty} a^{-2} E(a) da, \quad (4)$$

$$c_\psi = \int_{-\infty}^{+\infty} \frac{|\psi(w)|^2}{w} dw, \quad (5)$$

$$E(a) = \int_{-\infty}^{+\infty} |W_f(a, b)|^2 db, \quad (6)$$

其中, $E(a)$ 是函数 $f(t)$ 在尺度为 a 时的能量表示。则得 $E = \{E_1, E_2, \dots, E_n\}$ 是信号 $f(t)$ 在 n 个尺度下的小波能谱。而 $q_i = E_i / \sum_{i=1}^n E_i$ 为任意尺度下的小波能谱在能量谱中的占有。由此计算出小波能谱熵:

$$H_{we} = - \sum_{i=1}^n q_i \lg q_i. \quad (7)$$

设 $W = [|W_f(a, b)|^2 / c_\psi a^2]$ 是信号在二维空间中的能量分布阵, 因此, 可以求得对应的奇异值谱 $\{\sigma_i\}$, $i = 1, 2, \dots, n$ 和对应划分比例占有 $p_i = \sigma_i / \sum_{i=1}^n \sigma_i$ 。由此计算出小波空间谱熵:

$$H_{ws} = - \sum_{i=1}^n p_i \lg p_i. \quad (8)$$

2.3 多源数据融合

多源数据融合技术在处理多层次、多方面等过程上有着广泛应用, 因该技术能全面地对多传感器采集的多数据进行检测、相关、组合和估计等处理, 提高了将其应用于状态、身份识别中的精确度, 并且该技术也能实现对极复杂的多变态势的实时评测。该技术在应用中, 由多个传感单元获得证据信息, 然后将这些证据信息进行全面融合, 从而完成了较单一传感更精准、更稳当、更有效的解析和判别 [18]。

多源数据特征级融合, 在传感单元处进行数据的特征提取, 然后通过对提取特征量进行分析完成最终的融合, 其流程如图1所示。

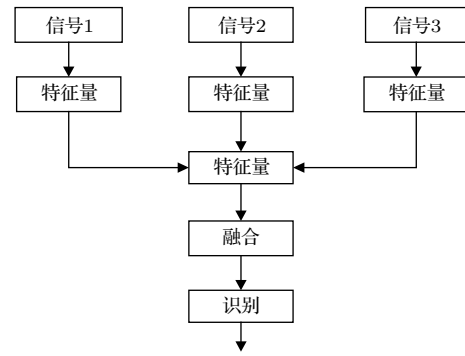


图1 特征级融合

Fig. 1 Feature-based fusion

2.4 改进的 GA-SVR 回归预测

一般的支持向量机回归 (SVR), 在进行二次规划问题时, 其核函数可以完成矩阵运行及相应的乘积运算等, 这相较其他一般算法效率和功能有明显的提升 [19]。

基于遗传算法 (GA) 的 SVR 规划, 是由 GA 完成优化。该算法可以自适应地随机搜寻, 能在全局大概率地得到最优解, 确定出有效的回归参数惩罚因子 c 与核函数半径 g 等。同时, 基于遗传算法的支持向量机回归 (GA-SVR) 基于 ϵ -不灵敏损失函数, 提高了算法的鲁棒性和泛化性 [20]。

GA-SVR 的预测流程如图2所示。

传统 GA-SVR 的回归预测, 会受算法的反复迭代计算、因子选择和函数选择的影响, 使其搜索优化力受限, 降低了收敛速率和效力 [21]。通过对风电机组噪声信号的大量数据特征进行分析, 提出

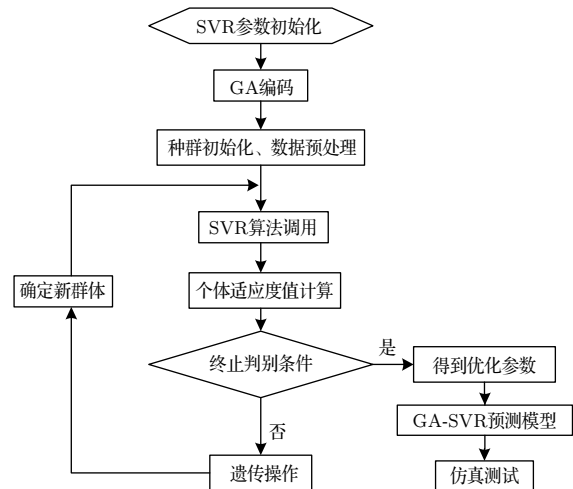


图2 GA-SVR 的预测流程

Fig. 2 Prediction process of GA-SVR

GA-SVR 回归预测模型的改进, 在传统的终止判别中加入能试用于噪声的平均适度值精度, 以弥补传统遗传算法引起的缺陷。

设平均适度值精度为 SMSE, 先求出回归模型的平均平方误差 (Mean-square error, MSE), 即

$$MSE = \frac{1}{m} \left(\sum_{i=1}^m y_i - \bar{y}_i \right)^2, \quad (9)$$

式 (9) 中, m 为样本点数, y_i 为实际值, \bar{y}_i 为预测值。然后由噪声信号特征计算出适度值精度判断条件:

$$SMSE = \left(\frac{1}{m} \left(\sum_{i=1}^m y_i - \bar{y}_i \right)^2 + \beta \right)^{-1} < \alpha, \quad (10)$$

式 (10) 中, α, β 均为正数, β 是一较小的正数, α 是经验所得判别正数。

2.5 多源数据特征级融合与改进的 GA-SVR 相结合

多元线性回归诊断能够提高数据样本的有效性, 再结合信息熵的噪声信号特征保留性, 将这些特点通过改进的 GA-SVR 完成多源数据特征级融合, 建立有效模型实现风电机组噪声预测, 其流程如图 3 所示。

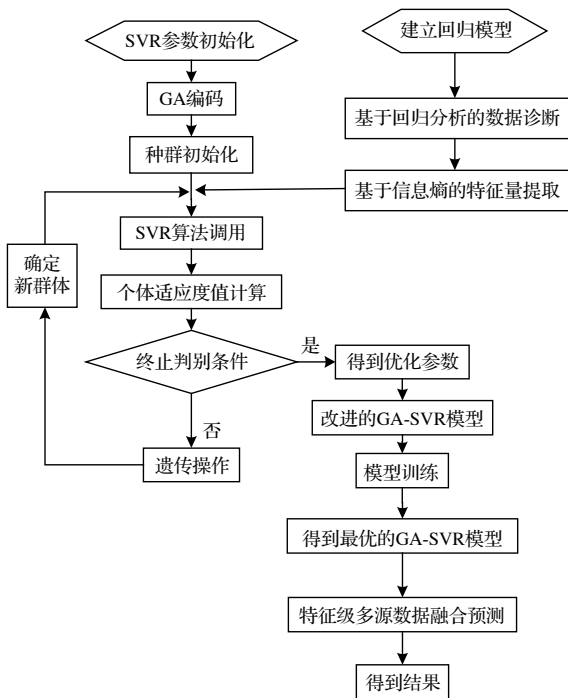


图3 多源数据融合结合改进的 GA-SVR

Fig. 3 Multi-source data fusion with improved GA-SVR

3 仿真实验

3.1 样本数据

2014年8月, 对新疆某风电场 GW87/1500 型风电机组, 在其正常运行状态时进行了噪声数据的采集, 在采集中设定了标准测试位 1 个、备用测试位 3 个, 用于现场麦克风设备的置放, 测量示意图如图 4 所示。机组相关参数如表 1 所示, 实验依据国标持续采集了 30 min。从采集样本中选取 250 个数据点作为训练样本, 另对应选取 250 个样本点作为测试样本, 进行实验仿真。

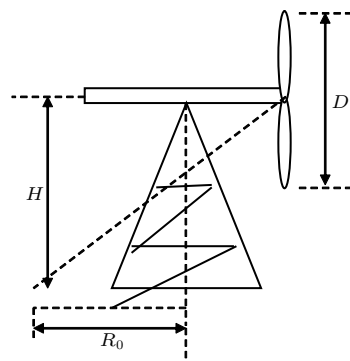


图4 麦克风测点位置侧视图

Fig. 4 Side view of microphone measuring point position

表1 噪声测量中的相关参数

Table 1 The related parameters in noise measurement

参数	数值
轮毂高度 H	70 m
风轮直径 D	87 m
测点位置 R	115 m
载放铝板直径	1 m
海拔高度	1500 m
地表粗糙度 (常数, 无量纲)	0.05
麦克风相对风机地基平面高度	1 m

3.2 多元线性回归诊断

非声学参数 X_1 表示轮毂风速、 X_2 表示环境温度、 X_3 表示变桨角度、 X_4 表示偏航角度、 X_5 表示发电机温度、 X_6 表示输出功率、 X_7 表示发电机转速、 X_8 表示齿轮箱转速、 X_9 表示大气气压, 用这 9 个非声学参数进行噪声预测时, 不可避免地会因输入变量维度过多、关系密切而产生共线性的影响, 因此进行回归诊断可提高实验可靠性。

回归诊断分为回归诊断与异常点诊断。本次仿真以条件指数 k (无量纲) 与方差膨胀因子 vif (无量纲) 为回归诊断指标, 应用统计软件 SAS 完成指标计算 [22]。

最终计算出 k 和 vif 的值如表 2 所示。

当 $1 \leq k \leq 10$ 时, 变量间为无共线性关系; 当 $10 < k \leq 30$ 时, 变量间有中等共线性关系; 当 $k > 30$ 时, 变量间有严重共线性关系。当 $vif > 10$ 时, 变量间有严重共线性关系。由计算值可进行判别, 本次仿真输入变量间存在严重共线性影响。

对样本进行学生化残差以及 COOKD 值计算, 确定出异常点, 并将其剔除出样本实验库, 提高预测精度。异常点诊断结果如表 3 所示。

因已判断出输入变量间存在严重共线性影响, 为提高预测速率和准确度, 本次实验对共线性的

变量采用变量筛选法进行处理并起到降维功能。

在进行降维的过程中, 主要是通过多元回归原理建立回归方程, 先找出输入变量集的主分量, 然后通过 F 统计量计算各变量的显著性水平, 最终决定出删除的输入变量, 确定出主分量回归方程, 实现降维。

表 3 异常点诊断结果

Table 3 Some diagnosis of outlier

样本点	学生化残差	COOKD
331	2.345	0.003
332	2.753	0.005
333	2.555	0.006
364	-2.102	0.007
390	-2.203	0.004
402	2.068	0.005
419	-2.485	0.008
461	-2.903	0.005
465	-2.108	0.006

表 2 条件指数 k 与方差膨胀因子 vif

Table 2 Condition indices and variance inflation factors

变量	vif	k
X_1	1.60384	1.03120
X_2	3.54732	2.16532
X_3	12177659	198730
X_4	495630	4.01231
X_5	1.46302	469889
X_6	583721	7.50321
X_7	39911730	2910532
X_8	40353721	8611223
X_9		70563

3.3 改进的 GA-SVR 预测仿真

经分析研究发现基于 SVR 进行风电机组噪声预测模型的建立时, 采用高斯核函数更适应于噪声特性。而 GA 能够快速有效地对模型参数进行寻优, 并得到很好的寻优结果。首先仅用改进后的 GA-SVR 对风电机组噪声信号进行预测仿真, 预测结果如图 5(a) 所示, 蓝色曲线表示预测结果。其相对应的相对误差百分比及其平均值为 1.2103%, 如图 6(a) 所示。

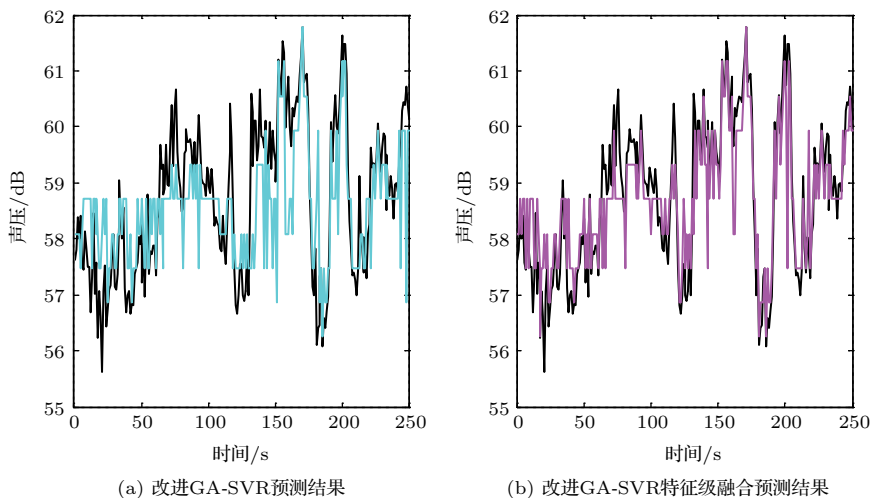


图 5 预测结果比较

Fig. 5 Comparison of predicted value

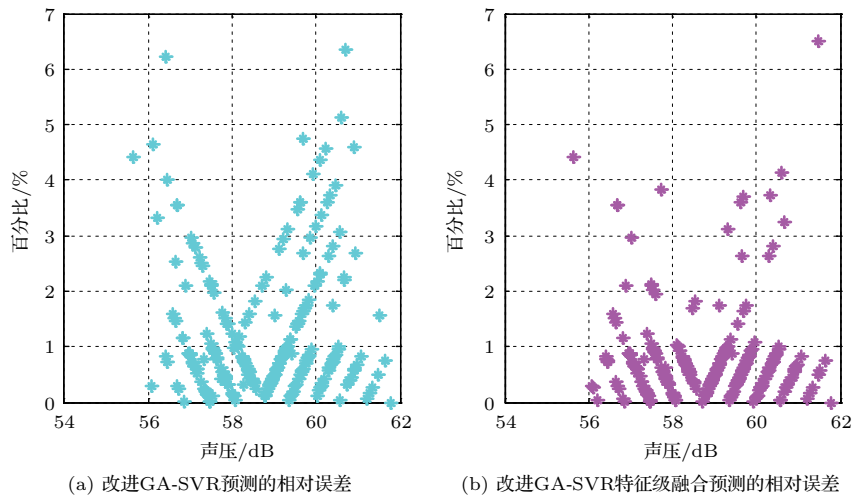


图6 相对误差百分比

Fig. 6 Relative error percentage

结果表明, 仅用改进的 GA-SVR 进行噪声预测可以得到较好的预测结果。

3.4 多源数据特征级融合与改进的 GA-SVR 相结合

基于改进的 GA-SVR 结合多源数据特征级融合理论建立预测模型, 用式 (2)、式 (3)、式 (7)、式 (8) 分别计算出样本数据点的奇异谱熵值、功率谱熵值、小波能谱熵和小波空间谱熵值, 计算结果如表 4 所

示, 并以此作为输入样本进行模型的建立和训练。然后, 用建立的模型对噪声声压进行仿真预测。最后, 比较最终预测值和预留的测试样本值, 如图 5(b) 所示, 粉色曲线表示预测结果。其相对应的相对误差百分比及其平均值为 0.7757%, 如图 6(b) 所示。

结果表明, 基于改进 GA-SVR 的多源数据特征级融合预测比单一应用改进 GA-SVR 的预测结果更精确, 其误差满足预测要求, 具有参考研究价值。

表4 特征量熵值的部分计算结果

Table 4 Partial results of entropy value of characteristic quantity

样本点	奇异谱熵 (HHT)	奇异谱熵表 (小波)	功率谱熵 (HHT)	功率谱熵 (小波)	小波能谱熵	小波空间谱熵
1	56.044352	89.66496	5.637492	7.324273	0.979079	99.71229
2	30.355616	89.50509	4.389894	7.383375	0.903201	99.84185
3	25.824474	88.73986	3.915231	7.322817	3.494736	99.55263
4	45.052191	89.77979	5.018310	7.388075	0.849476	99.72423
5	29.373089	89.73350	4.381837	7.354744	10.74266	99.53277
6	29.431296	89.72160	4.0892545	7.351689	0.912249	99.67992
7	45.210494	89.43568	5.096884	7.368043	0.777791	99.79305
8	23.342649	89.34303	3.861155	7.336220	1.850124	99.74064
9	15.473443	89.26888	3.615082	7.344845	2.689231	99.39189
10	27.035722	87.27765	3.961777	7.441320	1.059100	99.94540

4 结论

本文针对多输入变量的情况进行了非声学参数间共线性的诊断和处理, 并对输入变量进行了有效地降维和数据异常点的甄别处理。通过对风电机

组噪声特性的研究, 对 GA-SVR 提出了改进, 使其能更好地适用于风电机组噪声预测模型, 并提高模型的预测精度和速率。通过对基于信息熵的研究, 应用信息熵理论对风电机组噪声信号进行特征提取, 并将其与多源数据特征级融合理论相结合, 应用

于GA-SVR中建立风电机组噪声预测模型。实验结果表明,基于改进的GA-SVR结合多源数据特征级融合理论的思想,建立了预测模型并对风电机组噪声进行预测仿真,得到了较为精确的预测结果,其相对误差也较小,可以为风电机组的噪声信号检测及早期故障诊断研究提供一定的参考。

参 考 文 献

- [1] 吴宏钢. 风力发电机组振动检测及噪声评估研究[D]. 重庆: 重庆大学, 2010.
- [2] Kuwano S, Yano T, Kageyama T, et al. Social survey on wind turbine noise in Japan[J]. *Noise Control Engineering Journal*, 2014, 62(6): 503-520.
- [3] Son E, Kim H, Kim H, et al. Integrated numerical method for the prediction of wind turbine noise and the long range propagation[J]. *Current Applied Physics*, 2010, 10(2): S316-S319.
- [4] Tachibana H. Outcome of systematic research on wind turbine noise in Japan[C]. *Inter-Noise 2014*, Australian Acoustical Society, 2014.
- [5] 徐冠基, 柏林, 刘小峰, 等. 基于多元线性回归分析的风机噪声预测的研究[J]. *中国测试*, 2010, 36(5): 21-23.
Xu Guanji, Bo Lin, Liu Xiaofeng, et al. Research on noise prediction of wind turbine based on multiple linear regression analysis[J]. *China Measurement & Test*, 2010, 36(5): 21-23.
- [6] 程静, 王维庆, 何山. 基于回归分析与BP神经网络的风机噪声预测[J]. *噪声与振动控制*, 2013, 33(6): 49-52, 123.
Cheng Jing, Wang Weiqing, He Shan. Noise prediction of wind turbines based on regression analysis and BP neural network[J]. *Noise and Vibration Control*, 2013, 33(6): 49-52, 123.
- [7] Liu H, Mi X W, Li Y F. Comparison of two new intelligent wind speed forecasting approaches based on wavelet packet decomposition, complete ensemble empirical mode decomposition with adaptive noise and artificial neural networks[J]. *Energy Conversion and Management*, 2018, 155: 188-200.
- [8] 郭伟, 左曙光, 陈瑞峰, 等. 风机非稳态噪声信号分析[J]. *噪声与振动控制*, 2010, 30(1): 78-81.
Guo Wei, Zuo Shuguang, Chen Ruifeng, et al. Analysis of unsteady noise signal of fans[J]. *Noise and Vibration Control*, 2010, 30(1): 78-81.
- [9] 蒋伟康, 严莉. 基于神经网络的电机噪声性能在线检测技术研究[J]. *振动与冲击*, 2004, 23(4): 51-57.
Jiang Weikang, Yan Li. Resear on online detection of motor noise characteristics based on neural network[J]. *Journal of Vibration and Shock*, 2004, 23(4): 51-57.
- [10] 梁栋, 张凤琴, 陈大武, 等. 一种基于决策树和遗传算法-BP神经网络的组合预测模型[J]. *中国科技论文*, 2015, 10(2): 169-174.
Liang Dong, Zhang Fengqin, Chen Dawu, et al. A composite prediction model based on decision tree and GA-BPNN[J]. *China Sciencepaper*, 2015, 10(2): 169-174.
- [11] Xu L F, Zhao Q, Chen Y R, et al. Single spectrum end point predict of BOF with SVM[J]. *World Academy of Science, Engineering and Technology*, 2010, 4(2): 189-191.
- [12] 王晓景, 黎敏, 阳建宏, 等. 结合相空间和LS-SVM的风机状态预测方法[J]. *中国科技论文*, 2013, 8(8): 743-746.
Wang Xiaojing, Li Min, Yang Jianhong, et al. Trend prediction for condition of fans based on phase space and least squares support vector machine[J]. *China Sciencepaper*, 2013, 8(8): 743-746.
- [13] GB/T 22516-2008/IEC 61400-11:2002 风力发电机组噪声测量方法[S].
- [14] Atici U. Prediction of the strength of mineral admixture concrete using multivariable regression analysis and an artificial neural network[J]. *Expert Systems with Applications*, 2011, 38(8): 9609-9618.
- [15] Alin A. Multicollinearity[J]. *Wiley Interdisciplinary Review: Computational Statistics*, 2010, 2(3): 370-374.
- [16] 彭珊. 线性回归模型中关于异常点的若干问题的分析[D]. 哈尔滨: 东北林业大学, 2014.
- [17] 化柏林, 李广建. 大数据环境下多源信息融合的理论与应用探讨[J]. *图书情报工作*, 2015, 59(16): 5-10.
Hua Bolin, Li Guangjian. Discussion on theory and application of multi-source information fusion in big data environment[J]. *Library and Information Service*, 2015, 59(16): 5-10.
- [18] 刘同明, 夏祖勋, 解洪成. 数据融合技术及其应用[M]. 北京: 国防工业出版社, 1988.
- [19] Baydaroglu O, Kocak K. SVR-based prediction of evaporation combined with chaotic approach[J]. *Journal of Hydrology*, 2014, 508(508): 356-363.
- [20] Sanz-Garcia A, Fernandez-Ceniceros J, Antonanzas-Torres F, et al. GA-PARSIMONY: A GA-SVR approach with feature selection and parameter optimization to obtain parsimonious solutions for predicting temperature settings in a continuous annealing furnace[J]. *Applied Soft Computing*, 2015, 35: 13-28.
- [21] Gao R Z, Zhang J J, Shang Y M, et al. An improve genetic algorithm based on fixed point algorithms[J]. *Journal of Computers*, 2012, 7(5): 1109-1115.
- [22] 胡良平. 现代统计学与SAS应用[M]. 北京: 军事医学科学出版社, 2000.