

最小二乘支持向量机预测绝缘子等值附盐密度

舒服华¹, 张望祥²

(1. 武汉理工大学机电工程学院, 湖北 武汉 430070; 2. 武汉大学电气工程学院, 湖北 武汉 430072)

摘要: 考虑到气象因子条件对绝缘子的等值附盐密度影响复杂,难以建立精确数学模型等问题,提出了一种最小二乘支持向量机的绝缘子在一定的气象因子条件下的等值附盐密度预测新模型。以温度、湿度、风速等主要气象因子为输入,绝缘子等值附盐密度为输出,通过最小二乘支持向量机模型,拟合输入与输出之间的复杂非线性函数关系。以现场采集的气候数据为样本对模型进行学习训练,用训练好模型预测绝缘子在一定气候条件下的等值附盐密度。实践表明该方法具有建模速度快、预测精度高、操作简便等优点,不仅克服了常规的BP预测模型的不足,而且性能优于标准支持向量机预测模型。

关键词: 绝缘子; 等值附盐密度; 预测模型; 最小二乘支持向量机

中图分类号: TM852

文献标志码: A

文章编号: 1001-1609(2008)05-0420-04

A Prediction Model for Insulator's ESDD Based on Least Square Support Vector Machine

SHU Fu-hua¹, ZHANG Wang-xiang²

(1. School of Mechanical and Electronic Engineering, Wuhan University of Technology, Wuhan 430070, China;

2. School of Electric Engineering University of Wuhan, Wuhan 430072, China)

Abstract: In consideration of the problem that equal salt deposit density of insulator is complexly influenced by climate factors and accurate model is difficult to construct, a novel prediction model of insulator's ESDD under different climate conditions is proposed based on least square support vector machine (LS-SVM). With five main climate factors including temperature, humidity, wind velocity, and so on as inputs, ESDD of insulator as output, the nonlinear mapping between input and output is fitted through LS-SVM. Collecting and processing field data as learning samples to train the model, the insulator's ESDD under certain climate condition is hence predicted by the trained model. Experimental results demonstrate that the model based on LS-SVM is constructed more rapidly than the standard SVM-based model, and its prediction error is smaller. Moreover, the present model gains better prediction accuracy and speed compared with the BP model.

Key words: insulator; equal salt deposit density(ESDD); prediction model; least square support vector machine(LS-SVM)

0 引言

目前,在超高压和特高压的交流系统中,外绝缘污秽问题已成为外绝缘选取的主要考虑因素。测量等值附盐密度(ESDD)是衡量高压绝缘子自然污秽的主要手段,其值反映了绝缘子污秽物种类和密度,并受测量用水量的影响。从试验数据来看,当前还无法推断被测绝缘子产生污秽的主要自然因素,也没有关于自然因素与污秽物密度之间的标准或经验公式。相对来说,气象数据的获得比等值密度的数据提取要更加节省人力和物力,由气象数据推测ESDD可实现绝缘子抗污闪能力的在线监测,避免传统检测下停电带来的经济损失和带电测量造成的危害^[1,2]。

近年来,许多学者将人工神经网络(ANN)技术运用于预测基于气象因子的绝缘子ESDD,取得了一定的突破^[3,4]。它具有大规模的并行处理和分布式的信息存储能力、良好的自适应性、自组织性及很强的学习、联想、容错和抗干扰能力,能以任何精度逼近复杂非线性系统,可获得较高的预测精度和预测效率,但也存在先天性不足。如建立模型时存在着网络内部单元意义不明确、训练时间长、易陷入局部极值点、外插能力弱、泛化能力低等缺点。支持向量机(SVM)是由统计学习理论发展起来的一种新型学习机器,它以结构风险最小化原理为理论基础,具有较强的学习泛化能力,特别是对小样本数据的模式识别和函数估计具有出色的学习推广性能。克服了ANN结构依赖设计者经验的缺点,较好解决了高维

收稿日期:2007-08-23; 修回日期:2008-03-27

作者简介:舒服华(1964-),男,湖北省武汉市人,博士研究生,副教授,主要从事优化设计和计算机应用研究。

数、局部极小、小样本等机器学习问题。最小二乘支持向量机(LS-SVM)是标准SVM的一种新扩展,它用等式约束代替标准SVM的不等式约束,将二次规划问题转化为线性方程组求解,降低了计算复杂性,具有更快的求解速度和更好的鲁棒性。笔者尝试用最小二乘支持向量机估计和预测基于气象因子的绝缘子ESDD值,取得了较好的效果。

1 最小二乘支持向量机原理

支持向量机的基本思想是通过用内积函数定义的非线性变换将输入空间变换到一个高维空间,在这个高维空间中寻找输入变量和输出变量之间的一种非线性关系,即通过非线性映射 $\varphi(\cdot)$ 把样本空间映射到一个高维特征空间(Hilbert空间),把寻找最优线性回归超平面的算法归结为求解一个凸约束条件下的一个凸规划问题,并可以求得全局最优解,使在特征空间中可以应用线性学习机的方法解决样本空间中的高度非线性回归问题。支持向量机非线性回归原理见图1^[5]。

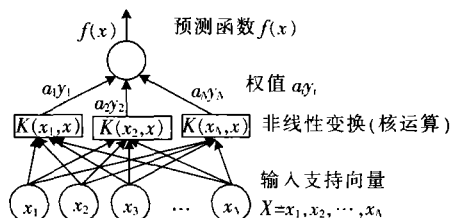


图1 支持向量机原理

LS-SVM通过最小化经验风险与结构风险之和来求解SVM。设某一待回归系统的输入、输出样本数据集为 $\{x_k, y_k\} (k=1, 2, \dots, N)$ 。其中, x_k 为 n 维系统输入向量, $x_k \in R^n$; y_k 为系统输出, $y_k \in R$ 。LS-SVM回归模型可以表示为以下形式

$$f(x) = \omega^T \varphi(x) + b \quad (1)$$

式中, $\varphi(\cdot)$ 为 $R \rightarrow R^m$ 将输入空间映射为高维特征空间的映射函数; ω 为超平面的权值向量, $\omega \in R^n$; b 为偏置量。

LS-SVM回归算法选择误差 e_k 的2-范数为损失函数,其优化问题为

$$\begin{cases} \min J(\omega, e) = \frac{1}{2} \omega^T \omega + \frac{1}{2} \gamma \sum_{k=1}^N e_k^2 \\ \text{s.t. } y_k = \omega^T \varphi(x_k) + b + e_k, k=1, 2, \dots, N \end{cases} \quad (2)$$

式(2)中, $\gamma > 0$ 为惩罚系数,用于调节误差所起的作用,它能够使训练误差和模型复杂度之间取一个折衷,以便使所求的函数具有较好的泛化能力,并且 γ 值越大,模型的回归误差越小。

引入Lagrange函数进行求解

$$L(\omega, b, e, a) = J(\omega, e) - \sum_{k=1}^N a_k (\omega^T \varphi(x_k) + b + e_k - y_k) \quad (3)$$

式(3)中, $a_k (k=1, 2, \dots, N)$ 为Lagrange乘子。

最优的 a 和 b 可以通过KKT条件获得^[6],即

$$\begin{cases} \frac{\partial L}{\partial \omega} = 0 \\ \frac{\partial L}{\partial b} = 0 \\ \frac{\partial L}{\partial e_k} = 0 \\ \frac{\partial L}{\partial a_k} = 0 \end{cases} \Rightarrow \begin{cases} \omega = \sum_{k=1}^N a_k \varphi(x_k) \\ \sum_{k=1}^N a_k = 0 \\ a_k = \gamma \cdot e_k \\ \omega^T \varphi(x_k) + b + e_k - y_k = 0 \end{cases} \quad (4)$$

通过消去式(4)中的变量 ω 和 e_k ,优化问题转化为求解如下线性方程组

$$\begin{pmatrix} 0 & \bar{1}^T \\ \bar{1} & \Omega + \gamma^{-1} I \end{pmatrix} \begin{pmatrix} b \\ a \end{pmatrix} = \begin{pmatrix} 0 \\ y \end{pmatrix} \quad (5)$$

式(5)中, $y = [y_1, y_2, \dots, y_N]^T$; $\bar{1} = [1, 1, \dots, 1]$; $a = [a_1, a_2, \dots, a_N]^T$; I 为单位矩阵; Ω 为方阵。

根据Mercer条件可知,存在映射 $\varphi(\cdot)$ 和核函数 $K(\cdot, \cdot)$ 使得

$$\Omega_{ij} = \varphi(x_i)^T \varphi(x_j) = K(x_i, x_j), i, j=1, 2, \dots, N \quad (6)$$

由式(5)求出 a, b 后,可得到LS-SVM的非线性

函数估计为: $y(x) = \sum_{k=1}^N a_k K(x, x_k) + b$ 。

选择不同形式的核函数 $K(\cdot, \cdot)$,可以生成不同的支持向量机。常用的核函数有:线性函数 $K(x_i, x_j) = x_i^T x_j$;多项式函数 $K(x_i, x_j) = (x_i^T x_j / c + 1)^d$;径向基(RBF)函数 $K(x_i, x_j) = \exp(-\frac{\|x_i - x_j\|^2}{2\sigma^2})$ 。

由此可见,LS-SVM辨识算法选用与标准SVM不同的损失函数,并用等式约束将求解二次规划问题转化为求解线性方程,显然可以极大地降低算法的复杂性,并提高辨识过程的抗干扰能力。LS-SVM模型的学习训练主要是对线性方程组(5)进行求解,而用LS-SVM模型进行预测只需计算各个训练样本与待测样本之间的核函数 $K(x_k, x)$,不涉及函数 $\varphi(\cdot)$ 的具体形式。

2 绝缘子等值附盐密度预测

2.1 物理模型与数据采集

等值附盐量是绝缘表面上的等值NaCl的量,当它和绝缘子表面上的实际沉积物分别溶解在一定量的水中时,它们将具有相等的体积电导率,NaCl量采用每平方米表面积的毫克数来表示。污秽直接受到工况下的自然条件影响,影响因素主要包括温度、湿度、降雨量、风速以及气压等。湿度对污秽生成影响最明显,湿度很小则表面形成干灰,不易凝结污物;湿度较大则污秽稀释性很大,较易冲刷掉;而当为小雾状天气时,湿度最适宜污秽物的生成。降雨量是个复杂的因素,一方面它能冲洗掉部分或全部污层中的电解质成分,使闪络不可能发生;另一方面它

又能使伞裙间的气隙桥接,从而引起闪络。风速对污秽影响是双重的,风速小,则表面污物沉淀少,随着空气流动的增大,ESDD 也加大,但超过一定限度后,ESDD 又降低,这是因为风力已足够吹刷一部分的凝结物。温度和气压的影响相对较小,往往要结合其他自然因素联合作用^[2]。

气象因素与绝缘子 ESDD 之间的这种复杂非线性映射关系,可以利用最小二乘支持向量机模型泛化,其优点在于:模拟真实系统的唯一限制条件是实验数据而不是系统理论的解析,不需要了解各个自然因素与最终盐值密度之间的数学关系公式。并且 LS-SVM 特别适合于小样本集的数据处理,可大大减少数据采集量,同时既能处理非线性数据,又能限制过拟合,比传统的模式识别和人工神经网络具有更高的预报能力。

测试数据来自湖北省清江地区 220 kV 线路的玻璃绝缘子以及自然污秽试验站样本,根据绝缘子耐污闪的设计标准,当 ESDD 超过 0.06 mg/cm²,即进行清洗。由于测量用水量对绝缘子的 ESDD 存在影响,每片绝缘子统一采用 300 mL 的蒸馏水。污秽物的种类繁多,该测试选取了主要自然污秽 CaSO₄ 等盐类。气象数据取自采样当天当地气象状况。

所有数据标本按照 5 种气象因素:温度、气压、湿度、降雨量和风速分类组合。每个绝缘子取了连续 40 天的气象条件,由于数据是分散的,为了使模型更好地收敛,通过方差分析剔除了超出范围的 8 个数据,剩余的 22 个用来建立和训练网络,10 个用来检测 LS-SVM 系统的输出均方误差。

2.2 建立 LS-SVM 模型

以雨量(mm³)、湿度(%)、风速(m/s)、气压(kPa)、温度(°C)5 个气象因子为输入量,绝缘子 ESDD (mg/cm²) 为输出量,建立 LS-SVM 系统,采用径向基核函数。模型性能评价指标采用平均误差计算公式

$$\epsilon = \frac{1}{m} \sum_{i=1}^m \frac{|Q_i - f_i|}{f_i} \quad (7)$$

式(7)中, Q_i 为实际的测量值; f_i 为预测值; m 为验证次数。

核宽度 σ、惩罚系数 γ 是建立 LS-SVM 模型的重要问题。σ 和 γ 的选择通常采用交叉验证方法,但

是交叉验证方法由于和的参数集参数有限的原因,经常出现不能满足达到误差精度的情况。为此,笔者采用一种自适应的 σ 和 γ 选择来建立 LS-SVM 模型,具体步骤见图 2^[5]。

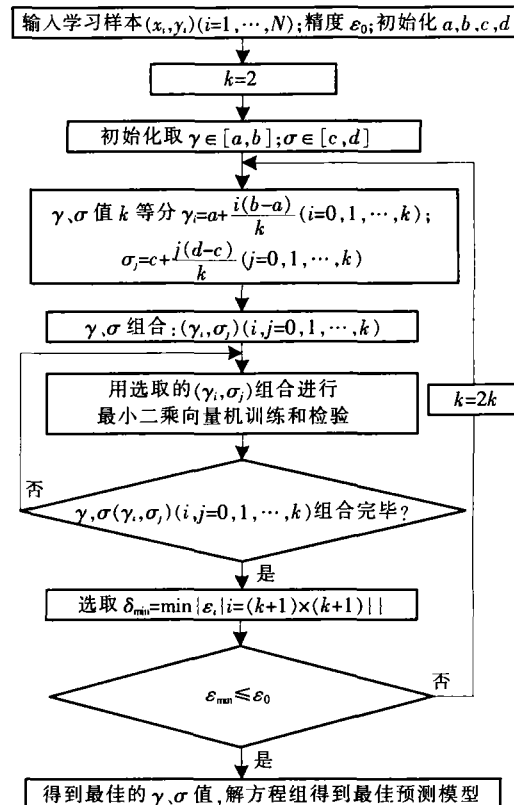


图 2 建立 LS-SVM 预测模型流程图

以训练样本对系统进行训练,完成训练后用测试样本对系统进行测试。步骤如下: ①输入训练样本; ②选定径向基核函数,初始化核宽度和惩罚因子; ③按自适应选择方法求解核宽度和惩罚因子; ④根据 LS-SVM 算法求解回归参数 a 和 b; ⑤将模型参数引入 LS-SVM 预测机,输入测试样本预测绝缘子的 ESDD。

预测结果见表 1,预测误差小于 3%,平均误差为 1.7。为了检验文中预测方法与神经网络预测方法的性能,设计一个 BP 神经网络,网络结构及参数见文[2]。ANN 模型预测的结果见表 1,最大预测误差

表 1 测试样本及结果

样本号	雨量/mm ³	湿度/%	风速/(m·s ⁻²)	气压/kPa	温度/°C	等值附盐密度/(mg·cm ⁻²)				
						实验值	ANN 预测值	误差/%	LS-SVM 预测值	误差/%
1	0	72	1	101.1	15	0.033 1	0.034 3	3.72	0.033 6	1.54
2	0	74	2	101.3	15	0.032 4	0.033 3	2.87	0.301 4	1.72
3	5	76	2	101.2	17	0.028 1	0.028 8	2.64	0.028 4	0.98
4	5	80	1	101.4	18	0.027 5	0.028 6	3.91	0.027 8	1.01
5	11	76	3	101.2	17	0.015 5	0.016 0	3.20	0.015 8	1.65
6	11	74	2	101.2	16	0.016 4	0.016 7	1.99	0.015 9	3.03
7	7	72	2	101.5	17	0.024 3	0.023 7	2.55	0.023 8	2.04
8	7	74	3	101.2	16	0.023 7	0.023 2	2.21	0.024 1	1.56
9	9	76	2	101.1	15	0.021 2	0.022 0	3.97	0.021 7	2.34
10	9	80	1	101.3	18	0.019 6	0.018 8	4.02	0.019 4	1.13

大于4%,平均误差为3.11,这表明LS-SVM绝缘子ESDD预测模型精度显著提高,同时,LS-SVM预测模型学习训练时间大大缩短,仅为ANN模型的千分之一。几种预测方法的结果比较见图3。

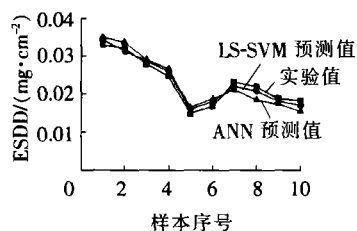


图3 几种预测方法结果比较

3 结语

由于气象因子与绝缘子ESDD存在非线性关系,人工智能技术预测绝缘子ESDD的方法,无须知道各个自然因素与最终盐值密度之间的数学关系公式,为高压系统管理提供了一个新的手段。最小二乘支持向量机由于将求解二次规划问题转化为求解线性方程,是最完善的预测和识别人工智能技术。它不仅适合小样本状态下的机器学习问题,而且性能优于标准SVM和ANN。在建模时间上一般比SVM方法缩短1~2个数量级,比BP方法可缩短2~3个数量级,而预测精度要比SVM标准模型高0.5~0.6

倍,比BP模型高1~2倍。

利用气象数据获得的绝缘子等值密度可以满足工程需要,它将污物生成条件与抗污闪能力直接联系起来^[6],通过模型能够判断出影响等值密度的最主要气象因素,据此挑选合适类型的绝缘子,以提高整个线路的抗污闪能力。同时,检测方法更简便、直观,实现了绝缘子污秽物的在线评估,符合状态检修的发展要求。

参考文献:

- [1] 周龙,文远芳,詹琼华. 递归神经网络及其在MOA状态预测中的应用[J]. 高压电器, 1998, 34(1): 22-24.
- [2] 张寒,文习山,丁辉. 用神经网络预测基于气象因素的绝缘子等值附盐密度[J]. 高压电器, 2003, 39(6): 31-35.
- [3] 曹一波,谢小鹏. 基于最小二乘支持向量机的磨损预测[J]. 润滑与密封, 2007, 32(2): 138-141.
- [4] 吴德会. 基于最小二乘支持向量机的表面粗糙度预测模型铣削加工[J]. 中国机械工程, 2007, 18(7): 838-840.
- [5] KU C C, LEE K Y S. Identification and Control Using Diagonal Recurrent Neural Networks [C]// Proc American Control Conf. Chicago, USA, 1992: 545-549.
- [6] ZHU Han-xin, RAGHUVVEER M R. Influence of Representation Model and Voltage Harmonics on Metal Oxide Surge Arrester Diagnostics[J]. IEEE Trans. on Power Delivery, 2001, 16(4): 599-603.

(上接第419页)

表1 不同等级谐波测量准确度

等级/10%	输入/A	输出/A	误差/%
基波/50 Hz	822.92	823.17	0.030
3次	823.67	823.91	0.029
5次	818.46	817.45	-0.123
7次	824.52	824.68	0.019
9次	826.22	825.27	-0.115
11次	821.69	820.60	-0.133
13次	820.26	820.30	0.005
15次	819.28	819.76	0.059
25次	819.05	821.89	0.347
35次	818.89	822.19	0.403
41次	818.84	822.35	0.429
51次	818.83	822.55	0.454

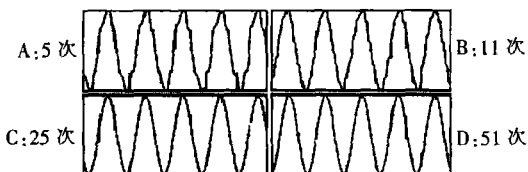


图5 谐波测量实验中互感器数字输出典型波形

4 结语

笔者介绍了闭环光纤电流互感器基本原理和结构组成,分析了光路主要误差源及相应的抑制措施。依据电子式电流互感器行业标准 IEC 60044-8,设计了实验室样机测量准确度实验。实验结果表明:室

温条件下,样机测量变比准确度达到0.1%,对工频电流响应相位差大约129.6';在-40~60℃范围内样机测量变比误差优于0.2%;互感器样机对51次以内谐波测量变比准确度优于0.5%。研制的光纤电流互感器样机测量准确度基本能够满足实用化要求。

参考文献:

- [1] 王政平,康崇,张雪原,等. 光学电流互感器的问题与对策[J]. 传感器技术, 2005, 24(5): 5-7.
- [2] BOHNERT K, GABUS P, BRÄNDLE H. Fiber Optic DC Current Sensor for the Electro-winning Industry [C]// Conference on Optical Fiber Sensors, 2005: 210-213.
- [3] 王夏霄,张春焱,张朝阳,等. 光纤电流互感器λ/4波片温度特性及其影响研究[J]. 激光与红外, 2006, 36(7): 596-599.
- [4] BLAKE J, TANTASWADI P, CARVALHO R T DE. In-line Sagnac Interferometer Current Sensor[J]. IEEE Transactions on Power Delivery, 1996, 11(1): 116-121.
- [5] 王夏霄. Sagnac型光纤干涉仪法拉第效应研究[D]. 北京:北京航空航天大学, 2006.
- [6] 廖延彪. 偏振光学[M]. 北京: 科学出版社, 2003.
- [7] 张唏. 光纤陀螺闭环检测与控制[D]. 北京:北京航空航天大学, 2001.
- [8] 王政平,李庆波,齐异,等. Verdet常数色散对光学电流传感器灵敏度影响的理论研究[J]. 哈尔滨工程大学学报, 2004, 25(2): 188-191.