

基于SVR算法的燃气轮机功率预测研究

王文超 苗夺谦 陈骥远

(同济大学计算机科学与技术系 上海 201804)

摘要 对燃气轮机未来几小时的功率预测是跳闸等故障预警的关键,而国内该方面研究尚少。采用支持向量回归模型,并融合多变量预测,以提高预测的准确性。以某电厂燃气轮机运转的实际数据为例,设计多组对比实验,详细阐述了实验流程以及重要参数的选取方法,最终验证了该方法的有效性。

关键词 燃气轮机,功率预测,支持向量回归,多变量回归

中图分类号 TP227 **文献标识码** A

Gas Turbine Power Prediction Based on Support Vector Regression

WANG Wen-chao MIAO Duo-qian CHEN Ji-yuan

(Department of Computer Science and Technology, Tongji University, Shanghai 201804, China)

Abstract Forecasting for the next few hours of the gas turbine power is the key to predict trips, while the domestic related researches are still few. Though using support vector regression(SVR) model, and integration of multivariate forecasting, it improves the accuracy of the forecasts. These experimental data from the real data of a power plant, through some comparative experiments, describes the experimental procedure and selection methods of important parameters in detail. The results verify the effectiveness of the support vector regression techniques applied to practical power prediction.

Keywords Gas turbine, Power prediction, SVR, Multivariate regression

1 引言

随着科学技术的不断发展,燃气轮机的应用也越来越广泛,且燃气轮机的复杂性也在逐渐增大。然而,燃气轮机也因为各种不同程度的故障严重影响到机器的安全运转,甚至造成巨大的经济损失和伤亡事故。据相关文献^[1]介绍,在机器设备上应用故障预测技术,获利投资比极高,目前各工业国家都在向预测性维修方向发展,而其中重要一环便是对各种运行状态进行准确预测。现阶段对燃气轮机以及旋转机械的预测研究主要分为两大方向:基于统计学基础的参数回归模型和基于机器学习理论的非参数回归模型。其中,文献^[2]采用参数估计的蒙特卡洛(MC)模型实现了对燃气轮机的状态预示,此外还包括二叉树(BT)方法、有限差分(FD)等。非参数方法包括人工神经网络(ANN)、自回归滑动平均(ARMA)和支持向量回归(SVR)等,其中,文献^[3]就是通过 ARMA 模型实现了对燃气轮机涡轮和压缩机清洁度的预测。非参数回归模型由于完全由数据驱动,无需先验指导,因此适应能力强、模型精度高,并且对于非线性、非齐次问题有着非常好的效果。近几年预测模型的研究热点也转向了非参回归法。

针对燃气轮机或旋转机械的非参数回归方法,目前主要对燃气轮机转子振动的预测^[4]研究较多,而对功率的预测研

究尚少,主要是由于功率不仅受燃气供给量的影响,还受各种因素的直接或间接影响,预测模型很难建立。然而,功率又是体现机器运行状态的最直接因素,任何机器故障都会直接导致功率的突然变化。因此,研究燃气轮机功率未来几小时内的变化走势也显得格外重要。

本文基于 SVR 算法建立多变量回归预测模型,再将测试样本进行测试,用最近 5 点的观测值,通过模型计算出下一点的预测值,从而实现功率的短期预测(4 小时以内),并且经过一系列参数调整、算法比较等实验,实现了燃气轮机功率预测的有效性,较为全面地阐述了该实验具体的过程,为今后燃气轮机故障预测等方面的研究提供一些借鉴和指导。

2 支持向量回归(SVR)基本原理

2.1 支持向量机(SVM)

支持向量机(Support Vector Machine, SVM)是由 Vapnik 等人在 20 世纪 90 年代提出的一种基于统计学理论(Statistical Learning Theory)的机器学习(Machine Learning)算法^[5]。支持向量机算法基于结构风险的最小化,不同于传统的经验风险最小化原则,因此,适合小样本的学习训练,并且具有学习速度快、泛化能力强等特点。因此,支持向量机在机械性能监测诊断领域具有良好的发展前景,也是近几年的研

本文受国家自然科学基金项目(60970061,61075056,61103067),中央高校基本科研业务费专项资金资助项目资助。

王文超(1988—),男,硕士生,CCF 学生会员,主要研究方向为嵌入式、数据挖掘,Email:jhbrown@foxmail.com;苗夺谦(1964—),男,博士,教授,博士生导师,CCF 高级会员,主要研究方向为粗糙集理论、粒计算、Web 智能与模式识别、云计算;陈骥远(1990—),男,硕士生,主要研究方向为数据挖掘。

究热门^[6]。支持向量机的基本思想是将原有的数据集投射到高维空间中,在此基础上构建一个超平面来解决分类、回归以及其他问题。支持向量机利用核函数方法,将原先线性学习器进行扩展,再利用拉格朗日二乘法及二次规划解决其在高维空间下的最优化问题。

2.2 支持向量回归(SVR)

将支持向量机引入回归问题即是支持向量回归,基本思想就是利用支持向量机方法寻找一个与训练集中所有点拟合度最高的超平面,通过把样本集因变量进行上下平移,将回归问题转化为分类问题^[7]。因此,类似于支持向量机,支持向量回归的基本形式如下:

$$\begin{aligned} & \text{minimize } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ & \begin{cases} y_i - w \cdot x_i - b \leq \epsilon + \xi_i \\ w \cdot x_i + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned} \quad (1)$$

式中, $\epsilon \geq 0$, 表示 SVR 模型预测值与实际值的最大差距。每组训练实例都有对应的 ξ_i, ξ_i^* , ξ_i 表示的是松弛变量,表示可以违反约束条件的范围,用来保证不会落在 ϵ 之外,惩罚系数 C 决定了函数拟合精度。

为了便于计算,引入 Lagrange 方程:

$$\begin{aligned} L = & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) - \sum_{i=1}^l (\eta_i \xi_i + \eta_i^* \xi_i^*) - \sum_{i=1}^l \alpha_i \\ & (\epsilon + \xi_i - y_i + w \cdot x_i + b) - \sum_{i=1}^l \alpha_i^* (\epsilon + \xi_i^* + y_i - w \cdot x_i - b) \end{aligned} \quad (2)$$

为求得 L 最小值,需对所有变量求偏微分,分别对 w, ξ 和 α_i 求得偏导数为 0,之后代入式(2)。于是,最优化问题可以表示成如下形式:

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i, x_j) + b \quad (3)$$

$$b = y_i - \sum_{i=1}^l (\alpha_i^* - \alpha_i)(x_i, x_j) \quad (4)$$

式中, $K(x_i, x_j)$ 为核函数,作用是将数据集从低维非线性投影至高维线性,这也是支持向量算法的一个显著特点。常用的核函数有:

高斯径向基核函数(RBF):

$$K(x_i, x_j) = \exp\left(-\frac{x_i - x_j}{\sigma^2}\right) \quad (5)$$

高斯径向基具有局部性强的特点,外推能力随着参数的增大而减弱。

多项式核函数:

$$K(x_i, x_j) = [(x_i \cdot x_j) + 1]^d \quad (6)$$

式中, d 为特征调节参数,多项式核函数具有全局特性,因此是一种全局多项式,即使是相对距离很远的点,也会对多项式核函数的值产生影响。 d 值越大,映射的维数越高,计算量也越大。当 d 过大时,函数集的 VC 维升高,学习机器的复杂性也提高, SVM 的推广能力降低,容易出现“过拟合”现象。

Puk 核函数:

$$K(x_i, x_j) = \frac{1}{[1 + (\frac{2\sqrt{\|x_i - x_j\|^2}}{\sigma} \sqrt{2^{(1/\omega)} - 1})^2]^\omega} \quad (7)$$

3 数据处理及建模

3.1 样本选取

本次实验样本来自于某电厂 20 个相同型号的燃气轮机一年的数据,采集的频率为 20 分钟,总共采集 18 个属性。燃气轮机功率变化范围较大,且受燃气进气量影响最大,而燃气进气量又是受人为控制。因此本次实验从所有样本中截取了 8 段数据,每段数据进气量几乎保持不变,从而排除了人为因素,每段数据有 100 行测量值,其中,前 80 行用于训练,后 20 行用于测试。

3.2 预处理及建模

原始数据总共包含 18 个属性,若将所有属性均用于多变量预测,时间耗费巨大,而且效果也未必最佳,因此,本文采用的是 SPSS 工具对功率做相关性(correlation)分析。

如图 1 所示,最终选取 4 个正相关度最高和 1 个负相关度最高的属性,总共 5 个属性。

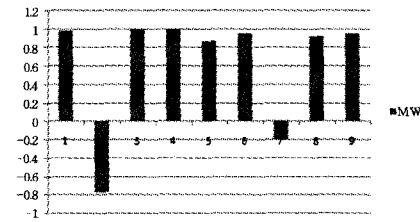


图1 各变量与功率(MW)比较

属性选择完成之后,构建 SVR 预测模型,模型的输入、输出量分别是:

$$X = [x(k-5), x(k-3), \dots, x(k-1)] \quad (8)$$

$$Y = [x(k)] \quad (9)$$

式中, $x(k-5), x(k-3), \dots, x(k-1)$ 表示的是前 5 个时刻功率以及之前所取的 5 个相关属性的变化值; $Y = [x(k)]$ 是下一时刻的预测值,这是单步预测值,之后的实验还有多步预测实验。

建模以及实验的过程如图 2 所示,经过测试样本测试后,还要再次通过比较修改参数,直至预测效果达到较好水平。

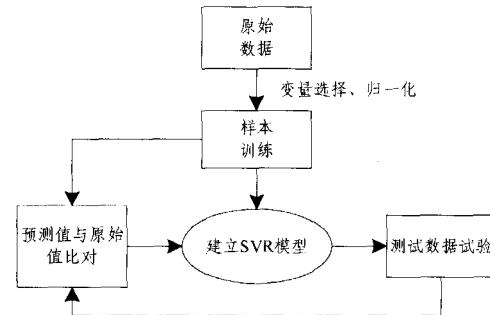


图2 SVR模型训练过程

检验预测数据准确度用的是均方根误差(RMSE),见式(10):

$$\sigma = \sqrt{\frac{\sum d_i}{n-1}} \quad (10)$$

式中, d_i 表示一组测量值与平均值的误差, σ 越小时,预测效果越好。

4 实验

4.1 核函数及参数选取实验

为了构造出一个具有良好性能的支持向量机模型,核函数的选择是十分重要的。在选择核函数解决实际问题时,通常采用的方法有:一是利用专家的先验知识预先选定核函数;二是采用 Cmess-Validation 方法,其即在进行核函数选取时,分别试用不同的核函数,归纳误差最小的核函数就是最好的核函数,同时核函数的参数也用同样的方法进行选定。本文通过实验对比了几个常用核函数的预测效果,如表 1 所列,Puk 核函数的效果要远远好于其余核函数,因此针对本次实验数据选择 Puk 核函数。

表 1 不同核函数实验结果

核函数	标准化多项式	径向基	多项式	Puk
均方根误差	16.660	18.2174	19.2917	0.3895

惩罚系数 C 用于控制拟合精度和算法的复杂度。 C 过大,往往产生过拟合,并且训练时间也会大大增加,而 C 值较小,又易导致欠拟合,因此,选择适当的 C 值对预测精度的影响至关重要。

本文通过多次实验进行比较,如图 3 所示,在 $C=100$ 处时产生拐点,变化趋于稳定,因此 C 值取 100 较为合适。

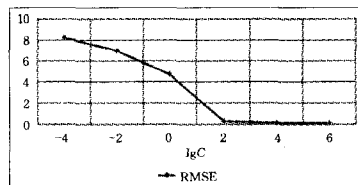


图 3 惩罚系数实验结果

4.2 多变量预测实验

由前面提到的 SVR 算法特点可知,它是将低维不可分数组通过核函数投影至高维可分的过程,因此具有多变量分析特点,这里选取之前已经提到的 5 个相关属性对功率做预测分析。而多变量是否能够提高预测准确度必须通过实验验证,这里将 8 组样本分别进行单变量与多变量实验,核函数选择 Puk, $C=100$, $\epsilon=0.01$, 回归步长为 12, 预测步长 $m=10$, 即预测之后第 10 步(3 小时 20 分钟)的值。表 2 列出了多变量与单变量的均方根误差(RMSE)比较结果。

表 2 多变量实验结果

样本序号	单变量误差	多变量误差
1	2.80	0.94
2	2.09	1.97
3	0.63	0.47
4	1.53	1.20
5	2.01	1.67
6	0.38	0.21
7	0.91	0.75
8	1.54	1.40

从表 2 可以看出,8 个样本里仅有 1 个多变量误差高于单变量误差。不难得出,多变量预测的整体效果明显优于单变量预测,不过,每组多变量预测所花时间约为 3 秒,而单变量预测不到 1 秒,因此并非变量越多越好,只有尽可能地选出合适变量,才能取得最佳结果。

4.3 SVR 和其他算法比较

本文还将其与其他常用算法模型进行比较,图 4 是支持

向量回归(SVR)多变量预测模型,图 5 是多层感知器(MLP)神经网络^[8]模型,图 6 是时序自回归滑动平均(ARMA)模型^[5],均为提前 10 步预测。方点为实际值,圆点为预测值,若肉眼看不到圆点表示预测值与实际值非常吻合,误差极小。

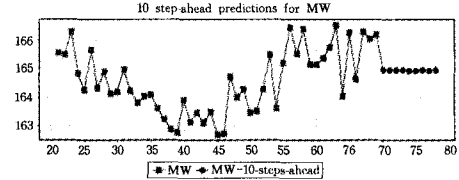


图 4 支持向量回归预测

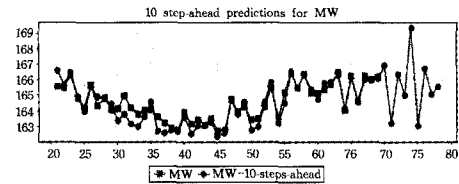


图 5 多层感知器模型(MLP)预测

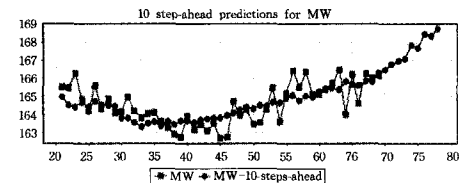


图 6 自回归滑动平均模型(ARMA)预测

对比 3 幅图不难发现,SVR 模型的准确性较 MLP 和 ARMA 模型好很多。图 6 的 ARMA 模型效果较差,可能是由于 ARMA 模型只在相对平稳的预测中发挥较好效果,而燃气轮机在相对平稳的运行中,功率波动还是比较大,因此可能更加适合对波动适应更强的 SVR 预测模型。同时,SVR 模型还克服了神经网络模型中收敛较慢、需要大数据样本支持等缺陷。

表 3 是 3 种模型具体的均方根误差比较,从 8 组样本随机挑选 3 组进行比较。SVR 选择的是多变量预测,均为提前 10 步预测。

表 3 3 种算法误差(RMSE)结果

样本序号	SVR	MLP	ARMA
1	0.940	2.77	5.53
2	0.470	0.455	0.758
3	0.210	0.907	4.989

结束语 本文提出了一种基于支持向量回归的燃气轮机功率的预测方法,其通过准确预测功率的变化趋势,来实现对异常状态的提前预知,具体结论如下:

由于 SVR 算法是支持多变量的预测,因此如何从大量的属性中寻找最相关属性尤为重要。利用相关性分析,挑选出与功率正相关的前 4 个属性和负相关的 1 个属性进行多变量分析效果最佳。

Puk 核函数其预测效果最好,尤其是对异常情况(如跳闸)的预测准确度最高,因此可以作为今后主要考虑的核函数选择。

由于步长、变量的选择等对准确性影响较大,以及在特殊情况下(如机器运行状态过于平稳、启动、关机时发生跳闸),预测的准确性还不是很高,因此,如何进一步提高预测的

准确性也将是今后长期研究的课题。

参考文献

- [1] 王然风. 基于支持向量回归技术的大型复杂机电设备故障诊断研究与应用[D]. 太原: 太原理工大学, 2005
- [2] Puggina N, Venturini M. Development of a Statistical Methodology for Gas Turbine Prognostics [J]. Journal of Engineering for Gas Turbines and Power, 2012, 134
- [3] Luo Hua-geng, Ghanime G, Wang Li-ping. Arma Model for Turbine and Compressor Clearance Forecasting[C]//Proceedings of ASME Turbo Expo 2010: Power for Land, Sea and Air GT2010. Glasgow, UK, June 2010; 14-18

(上接第 336 页)

以及 hash 函数, 重建秘密 S 和信誉值。

1. 首先, DD 利用其私钥 (x_1, x_2, y_1, y_2, z) 解密密文 $E(f(i) \parallel \epsilon(2^i))$, 恢复出信誉反馈值 $\epsilon(2^i)$ 和所有的子秘密 $f(i)$ (注意到在 Cramer-Shoup 加密方案中, 子秘密 $f(i)$ 的完整性得到保证)。

2. DD 通过计算和比较每个子秘密的 hash 码来验证其正确性。

3. 若有 t 或超过 t 个子秘密可用, 则 DD 可利用 Shamir 门限秘密共享方法计算并恢复出多项式 $f(u)$ 和秘密 $S = a_0$ 。

4. DD 通过计算和比较秘密 S 的 hash 值来验证其正确性。然后 DD 可确信信誉反馈密文 $\epsilon(2^i) (i \in [1, n])$ 来自于诚实的秘密共享者。

4.5 信誉计算阶段

DD 在收到信誉反馈密文 $\epsilon(2^i) (i \in [1, n])$ 后, 运用 ElGamal 密码系统的乘法同态性做如下计算:

首先 DD 将所有的信誉反馈密文相乘得到

$$\epsilon(2^1) \cdot \epsilon(2^2) \cdot \dots \cdot \epsilon(2^n)$$

利用 ElGamal 密码系统的乘法同态性可计算

$$\epsilon(2^1) \cdot \epsilon(2^2) \cdot \dots \cdot \epsilon(2^n) = \epsilon(2^1 \cdot 2^2 \cdot \dots \cdot 2^n) = \epsilon(2^{1+2+\dots+n})$$

结果中的指数 $2^1+2^2+\dots+2^n$ 是总的信誉反馈值。DD 解密得到这个总信誉反馈值, 然后发送给信誉引擎。

信誉引擎保存此信誉反馈的记录, 并设置一个过期的时间节点, 然后公布此信誉反馈值以及发布时间。任何需要信誉系统服务的人都可以从信誉引擎中获得信誉反馈值、发布时间和过期时间。

5 安全性分析

隐私保护的信誉协议里具有半可信参与者和恶意参与者。

半可信参与者能正确遵照协议执行, 其目标是获得其他参与者的信誉反馈值, 他们可能共谋以达到目标。在我们的协议中, 所有的信誉反馈都是密文传输的, 而且必须有 t 个或多个参与者才能恢复出秘密 S 和信誉反馈, 而这 t 个参与者是从 n 个参与者中随机选取的。这种随机选取法使得半可信参与者很难利用共谋攻击达到目标, 并且每个参与者的信誉反馈是加密的, 不知道 DD 的私钥, 则无法恢复出这些信誉反馈。因此, 对于半可信参与者, 我们的信誉协议能够保护每个参与者的隐私。

- [4] 吴庚申, 梁平, 龙新峰, 等. 基于 ARMA 的汽轮机转子振动故障序列的预测[J]. 华南理工大学学报: 自然科学版, 2005, 33(7): 67-73
- [5] Vapnik V. The Nature of Statistical Learning Theory [M]. New York: Springer, 1995
- [6] Song Zhao-qing, Cui He, Hu Yu-nan. Research and Development of Support Vector Machine Theory[J]. Journal of Naval Aeronautical and Astronautical University, 2008, 23: 143-148
- [7] 顾亚祥, 丁世飞. 支持向量机研究进展[J]. 计算机科学, 2011, 38(2): 14-17
- [8] 朱大奇, 史慧. 神经网络原理及应用[M]. 北京: 科学出版社, 2006

恶意参与者不一定遵守协议的执行, 可能根据自己的需要做一些偏离协议的行为, 他们可能设计复杂的攻击策略, 或者与其他参与者共谋以达到攻击目标。他们的目标包括获得其他参与者的本地信誉反馈, 获得对其他参与者的控制权。

在我们的协议中, 秘密 S 及其共享子秘密的 hash 值是可验证的。在计算信誉反馈前, DD 通过验证这些 hash 值的正确性来确保参与者遵守协议的执行。一旦发现有恶意的参与者, 协议就不会继续向下执行。因此也杜绝了恶意参与者参与协议。

因此我们的隐私保护信誉协议对于半可信参与者和恶意参与者都是安全的。

结束语 本文提出了一个能以隐私保护方式产生信誉反馈值的信誉协议。在我们的隐私保护信誉协议中, 信誉反馈请求以密文的方式发送。在信誉反馈时, 参与者的隐私得到保护, 不诚实的参与者也能被发现。因此诚实的参与者能够提供他们的真实信誉反馈, 最后得到一个真实的总信誉反馈值。通过分析得到, 我们的隐私保护信誉协议对于半可信参与者和恶意参与者都是安全的。

参考文献

- [1] Kamvar S D, Schlosser M T, GarciaMolina H. The enginetrust algorithm for reputation management in P2P networks[C]//Proc. of 12th Intl. Conf. on World Wide Web(WWW2003). New York: Springer-Verlag, 2003; 344-51
- [2] Xiong L, Liu L. Supporting reputation-based trust in peer-to-peer communities[J]. IEEE transaction on knowledge and data engineering, 2004, 12(7): 843-857
- [3] Pederson T. Non-interactive and information secure verifiable secret sharing[C]//Proc. of Advances in Cryptology-Crypto'91. New York: Springer-Verlag, 1991; 129-140
- [4] Resnick P, Zeckhauser R. Trust among strangers in internet transactions[J]. The economics of the internet and e-commerce, 2002, 11: 127-157
- [5] Shamir A. How to share a secret[C]//Communications of ACM 22, 1979. New York: Springer-Verlag, 1979; 612-613
- [6] Chor B, Goldwasser S, Micali S, et al. verifiable secret sharing and achieving simultaneity in the presence of faults[C]//Proc. of the 26th Annual symposium on foundation of computer science, 1985. New York: Springer-Verlag, 1985; 383-395
- [7] Mao Wen-bo. Modern cryptography: Theory and Practice[M]. Prentice Hall PTR, 2003; 514-590