

基于动态频率算法的远程监控系统数据采集优化策略

张 斌, 朱建涛, 徐 墨

(桂林电子科技大学 计算机科学与工程学院, 广西 桂林 541004)

摘 要: 提出一种基于动态频率算法的数据采集优化策略, 以 ARMA 预测为基础, 根据数据变化趋势, 动态改变采集频率, 从而以较小频率采集到报警信息, 在保证关键信息准确采集的同时, 提高了数据的实时性与数据挖掘效率. 最后给出具体算法与实验, 以及对比说明该算法的效果.

关键词: 数据采集; 故障诊断; ARMA; 动态频率

中图分类号: TP31

文献标识码: A

文章编号: 1000-7180(2016)08-0086-06

Data Gathering Optimization Strategy Based on Dynamic Frequency Algorithm for Remote Monitoring System

ZHANG Bin, ZHU Jian-tao, XU Zhao

(School of Computer Science and Engineering, Guilin University of Electronic Technology, Guilin 541004, China)

Abstract: The paper proposes an algorithm based dynamic frequency to solve the problem. Base the prediction of ARMA and data changing trend, the algorithm changes the frequency of data collection. Through the algorithm, the key information could be collected in time and the efficiency could be improved. At last, the algorithm and experiment will be displayed and the benefits of the algorithm in comparison with polling have been shown.

Key words: data gathering; fault diagnosis; ARMA; dynamic frequency

1 引言

对关键设备进行状态监测和故障预测可以提高设备的可靠性. 目前, 机械设备正朝着大型化、连续化、高速化、精密化、系统化和自动化方向发展^[1], 因此对监控实时性提出了更高的要求. 而监控的目的是及时预测和发现报警信息, 报警信息附近运行数据的可靠性和完备性是故障诊断结果正确与否的前提和关键^[2]. 使用软件定时轮询的方式进行数据采集^[3], 时间设置太长, 丢失重要数据的可能性增大, 设置时间太短, 又会产生巨大的网络资源消耗和数据冗余, 造成服务器巨大的计算压力. 例如以固定间隔 1 s 进行采集, 一个小时需要采集 3 600 次, 其中多数数据可能线性回归于正常值, 波动较小没有故

障诊断和数据挖掘价值. 以固定间隔 10 s 进行采集, 10 s 内设备运行数据的值可能已经超出报警阈, 监控系统不能及时发现数据变化, 从而失去监控的实际意义. 机械运行状态符合平稳时态序列, 因而可以使用自回归滑动平均模型 (ARMA) 对其进行预测. 而机械运行状态从正常值到报警阈值是具有时间过渡的, 因而可以在预测机械状态基础上, 提出一种根据数据运行趋势动态改变数据采集频率的算法, 从而可以以平均较小的频率采集到较多的报警及其附近数据, 为故障诊断提供数据依据. 本文介绍了远程故障诊断数据采集方法及 ARMA 的动态建模方案, 最后提出动态频率采集算法并对比分析该算法效果.

收稿日期: 2015-10-20; 修回日期: 2015-12-07

基金项目: 广西科学研究与技术开发计划项目 (桂科攻 1348014-6); 广西工业创新发展项目 (桂工信投资[2014]276 号)

2 远程故障诊断数据采集方法

2.1 数据采集测试方案

在设计远程机械数据采集方案之前,首先是对机械控制系统进行全面的分析,将控制系统进行功能分块,从而筛选出重要监测点,了解机械的各个状态特征,从而可以在分析运行状态的同时对下一状态进行在线预测。

远程数据监控系统总体框架如图 1 所示,主要分为四部分:数据采集现场、通信服务器、web 服务器及浏览器客户端。数据采集现场主要包括机械设备、机械控制系统及智能网络模块。通信服务器通过控制系统支持的网络协议采集控制器 PLC 中的寄存器数据,同时根据历史数据使用 ARMA 模型实时预测下一时刻数据,然后使用动态频率算法通过改变数据采集线程睡眠时间 t ,从而改变采集频率 f 。采集到的数据存储到数据库后,web 服务器读取数据库中数据。通过 B/S 模式,可以在浏览器随时查看数据运行情况 & 报警信息。

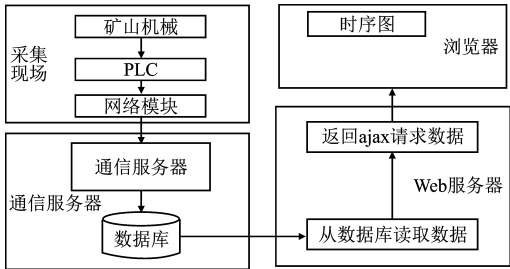


图 1 监控系统总体设计

2.2 ARMA 预测模型

时间序列预测的基本原理是,根据大量的历史数据推导出一个描述所研究时间序列的数学模型,利用建立的数学模型就可以用过去及现在的数据推导出未来数据,很适合实时一步预测^[4]。典型的时间序列预测模型 ARMA 及其相关算法已经在包括工业在内的各个领域取得了成功的预测应用^[5]。ARMA 是一种能够根据平稳时间序列预测未来短时间内数据的方法,对于平稳、正态、零均值的时序 $\{x_t\}$,若 x_t 的取值不仅与前 n 步的各个取值 $x_{t-1}, x_{t-2}, \dots, x_{t-n}$ 有关,而且还与前 m 步的各个干扰 $a_{t-1}, a_{t-2}, \dots, a_{t-m}$ 有关 $(n, m = 1, 2, \dots)$,则按照多元线性回归思想,可以得到最一般的 ARMA 模型:

$$\begin{aligned} x_t &= \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \dots + \varphi_n x_{t-n} - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_m a_{t-m} + a_t \\ a_t &\sim \text{NID}(0, \delta_a^2) \end{aligned}$$

上式表示 n 阶自回归 m 阶滑动平均模型,记为 $\text{ARMA}(n, m)$; n, m 分别表示 AR 部分和 MA 部分的阶次; $\varphi_i (i = 1, 2, \dots, n), \theta_j (j = 1, 2, \dots, m)$ 分别表示各部分的模型参数。若 x_t 的取值仅与其前 n 步的各个取值 $x_{t-1}, x_{t-2}, \dots, x_{t-n}$ 及 a_t 有关,则称为 n 阶自回归模型 $\text{AR}(n)$ 。如果 x_t 的取值仅与各个干扰 $a_{t-1}, a_{t-2}, \dots, a_{t-m}$ 及 a_t 有关,则称为 m 阶滑动平均模型 $\text{MA}(m)$ 。其中 a_t 是均值为 0,方差不为 0 的随机干扰项,即白噪声。对模型的选择及其阶数的确定可以根据平稳时间序列 $\{x_t\}$ 的自相关函数和偏自相关函数的拖尾和截尾进行判定。如表 1 所示。

表 1 模型判断

模型	AR 模型	MA 模型	ARMA 模型
自相关函数	拖尾	截尾	拖尾
偏自相关函数	截尾	拖尾	拖尾

AIC 准则比较适用于 ARMA 模型检验^[6],该准则从提取出观测时最大信息量出发,适用于 ARMA(包括 AR, MA)模型的检验。在一般情况下可以表示为^[7]:

$$\text{AIC}(p) = -2\ln L + 2p \tag{1}$$

$$\ln L = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} - \left(\frac{n}{2}\right) \ln\left(\frac{\delta_a^2}{n}\right) \tag{2}$$

将式(2) 去掉常数项即前两项后带入式(1) 后该准则最终形式可以表示为

$$\text{AIC}(p) = n \ln \delta_a^2 + 2p \tag{3}$$

式中, p 是模型参数的数量同; L 是时序 $\{x_t\}$ 的似然函数; n 是观测值长度, δ_a^2 为残差平方和。显然,在模型阐述估计方法给定的条件下, $\text{AIC}(p)$ 是模型阶数 p 的函数,当 p 增大时, $\ln \delta_a^2$ 下降,但后一项 $2p$ 增大,因此,取 $\text{AIC}(p)$ 值最小时的模型阶次 p 为适用模型阶次。 P 小意味着模型的简洁, L 大意味着模型的精确,因此评价模型时兼顾了间接性和精确性^[8]。

2.3 ARMA 动态建模方案

在监控系统中使用动态 ARMA 模型^[9] 进行实时预测,需要模型识别,阶数判定及参数估计。监控系统对 ARMA 建模综合使用 Box 和 $(2n, 2n-1)$ 建模方案^[10]。建模过程首先需要进行模型识别和阶数判定,然后进行参数估计。模型识别及阶数判定步骤如下所示。

步骤 1:采用最小二乘法进行参数估计,从 $n = 1$ 开始对 $\{x_t\}$ 拟合 $\text{AR}(n)$ 模型,检查最后一个自回归系数 φ_n ,如果 $\varphi_n \approx 0$,则确定适用模型为 $\text{AR}(n-1)$ 。否则令 $n = n+1$,继续拟合 $\text{AR}(n)$ 模型,如达到

给定的 n 的上限仍无 $\varphi_n \approx 0$, 则表示 AR 模型对 $\{x_t\}$ 不适合, 转下一步.

步骤 2: 从 $n = 1$ 开始对 $\{x_t\}$ 计算自协方差函数, 检查 R_n 的取值大小, 如果 $R_n \approx 0$, 则确定 MA($n - 1$) 为适用模型; 否则令 $n = n + 1$ 继续计算, 重复这一过程, 直至搜索到给定的 n 的上限. 如果搜索完毕仍无 $R_n \approx 0$, 则表示 MA 模型对 $\{x_t\}$ 不适合, 转下一步.

步骤 3: 该方法从 $n = 1$ 开始, 首先采用最小二乘法拟合 ARMA(2, 1) 模型, 使用 AIC 准则进行适用性检验; 若不适用, 再令 $n = n + 1$ 拟合 ARMA(4, 3) 模型; 如此循环, 直至确定出适用的 ARMA($2n$, $2n - 1$) 模型.

上述确定模型步骤具有以下优点: 第一, 如果 $\{x_t\}$ 是 AR 或者 MA 序列, 则能够很快得到适用模型; 第二, 确定 ARMA 模型搜索步长为 2, 只需少数几次就能得到适用模型, 计算量较小^[11].

动态建模过程如图 2 所示, 确定模型包括模型识别与阶数判定, 更新样本数据为删除预测样本数列末尾数据, 将采集到的最新数据添加到样本数列队首. 为减少建模计算时间, 根据预测误差 Δu 是否大于 u 决定是否重新确定模型, 其中 Δu 为数据预测值与真实值之间差值的绝对值.

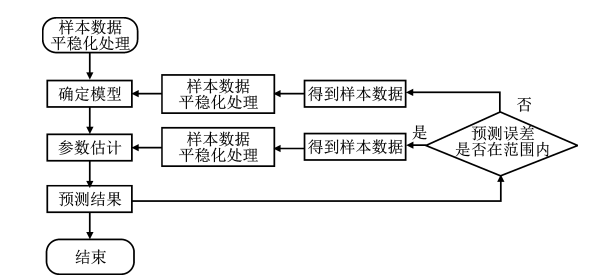


图 2 ARMA 动态建模流程

3 动态数据采集频率算法

3.1 算法思想

设备状态从正常发展到不正常, 导致故障发生的过程称为劣化过程. 不管故障原因如何, 劣化过程中特征曲线的共同点就是信号远离正常值, 变化梯度大, 且多以单调上升或者下降形式出现^[12]. 根据设备劣化过程数据变化趋势提出动态频率采集算法, 该算法思想如下: 按照最佳固定周期 T 采集, 首先判断预测到的数据是否大于报警值 alarm, 如果大于则加大采集频率; 采集时长 $2T$ 后恢复到最佳采集频率, 其次, 判断最新采集到的数据 now 是否

大于上次采集到的数据 last, 如果大于, 则判断预测到的下次数据 prediction, 是否大于 now 及 prediction, 是否大于算法初始化所有正常数据平均值 ave, 如果大于则说明数据呈上升趋势, 应该加大采集频率, 采集时长 T 后恢复到最佳采集频率.

该算法初始化需要确定最佳固定采集周期 T , 划分动态采集频率等级及级数 c , 根据 c 对最佳固定周期 T 内数据增量 Δy 进行等级划分. 在进行算法初始化时, 需要以最小固定采集周期 $T_{\min}$ 采集 n 次报警数据, 然后得到所有的劣化过程时长的最小值 t . 如图 3 所示, 将最短劣化时长 t 的一半作为最佳固定采集周期 T . 任何一次劣化时长均大于或者等于 t , 因此在劣化过程中可以采集到两次或者两次以上数据, 并且可以预测到下次数据, 然后根据动态频率算法决定是否加大采集频率.

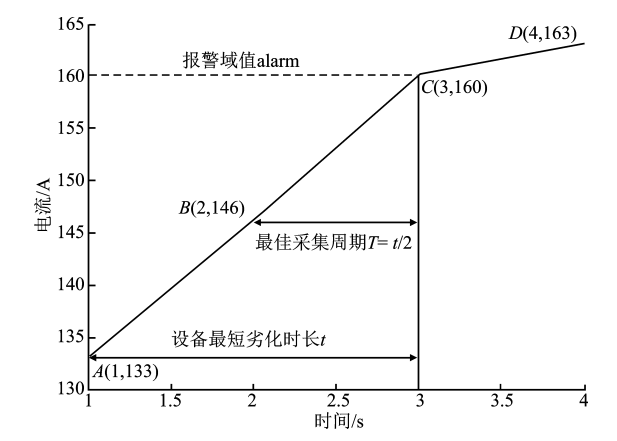


图 3 报警发生过程

划分动态采集频率等级 f , 即从最小采集周期 $T_{\min}$ 到最佳固定采集周期 T 之间等级的划分, 将最佳固定采集周期 T 除以 2, 直到结果小于 $T_{\min}$ 时, 将 $T_{\min}$ 及所有大于 $T_{\min}$ 的结果作为动态频率各个等级采集周期, 即 $f \in \{2/T, 4/T, \dots, 1/T_{\min}\}$, 分为 c 个

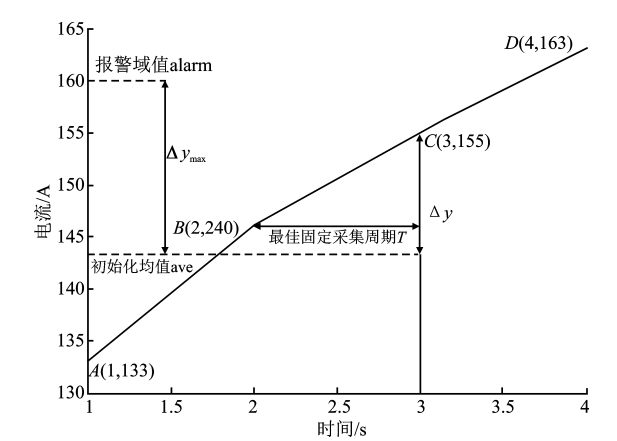


图 4 动态采集频率等级划分

等级. 如图 4 所示,对最佳固定采集周期 T 内相对于 ave 的数据增长大小 Δy 划分等级,得到报警阈值 $alarm$ 与正常数据平均值 ave 的差值 $\Delta y_{\max}$ 除以 c 得到平均值 u ,划分为 c 个等级 U ,即 $U \in \{u, 2u, \dots, cu\}$. 其中 U 与 f 对应关系为: $2/T$ 对应 u , $4/T$ 对应 $2u, \dots, 1/T_{\min}$ 对应 cu ,式(4)表示其函数关系. 当预测到的数据 $prediction$ 大于报警阈值 $alarm$ 时,则按照最小采集周期 $T_{\min}$ 进行采集;当时,根据 c 对应的频率等级 f 进行采集.

$$U = \begin{cases} u \log_2 fT, fT \in [2, 2C - 1] \\ CU, f = T_{\min} \end{cases} \quad (4)$$

动态频率采集算法初始化流程如图 5 所示,确定最佳固定采集周期 T ,动态频率等级 f ,数据平均值 ave ,数据上升等级 U . 算法采集报警次数 n 初始化为 3.

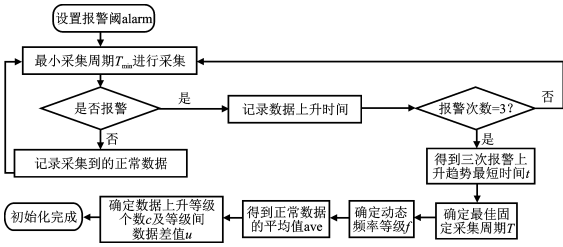


图 5 动态频率算法的初始化

表 2 最佳固定采集周期采集到的数据

129	146	163	169	158	140	127	121	125	135	147	156	156	146	137	129	131
141	154	162	155	143	132	128	131	137	147	154	156	148	136	124	123	132
145	158	163	156	138	124	116	126	145	165	171	161	141	122	114	119	136
151	159	155	145	132	129	135	142	147	147	149	144	139	136	139	146	153
162	164	156	144	133	135	149	163	171	165	146	127	116	123	139	154	166
168	158	145	132	122	123	135	151	159	154	140	124	117	125	136	147	155
151	144	134	130	134	138	140	144	143	142	137	131	132	134	141	146	145
143	142	137	130	126	127	135	144	150	149	147	145	146	142	140	142	143
145	146	141	139	140	143	148	153	156	160							

真实发生报警 15 次,预测到发生报警次数为 16 次. 预测到真实报警 13 次,将预测到的真实报警次数除以真实报警次数得到报警准确率 86.7%. 真实报警位置 TP 与预测到的报警位置 PP 表 3 所示.

表 3 真实报警位置与预测报警位置

TP	3	4	20	37	44	45	46	69	70	76	77	78	85	86	146	
PP	3	4	5	20	21	37	44	45	46	70	76	77	78	85	86	94

动态频率发生位置及对应位置采集频率等级分别如表 4 及表 5 所示.

表 4 动态频率发生位置

2	3	4	5	11	12	13	19	20	21	27	28
29	35	36	37	43	44	45	46	52	53	59	60
61	67	68	69	70	75	76	77	78	84	85	86
93	94	101	102	109	111	118	119	128	133	136	137
138	143	144	145								

动态频率采集算法如图 6 所示,图 6 中 $div1 = now - last$, $div2 = prediction - now$, $\Delta y = prediction - ave$.

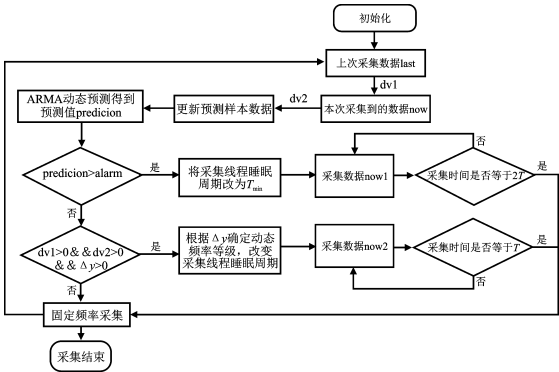


图 6 动态频率采集算法

3.2 算法效果分析

将该算法应用到上述部分 2.1 节所设计的远程监控系统,通过采集磨粉机中分级机电流数据,对该算法效果进行验证. 算法初始化将最小采集周期设置为 0.2 s. 将分级机电流报警阈值为 160,初始化得到电流平均值为 144. 得到最佳采集周期为 2 s,动态频率等级 $f = [1, 2, 4, 5]$. 以最佳固定采集周期采集到的 146 个数据如表 2 所示.

表 5 对应位置动态频率等级

1	5	5	5	2	4	4	3	5	5	1	3
3	1	3	5	1	5	5	5	3	4	1	2
2	1	2	4	5	1	5	5	5	4	5	5
2	5	2	2	1	1	1	1	2	1	1	1
1	1	2	3								

从上述发生频率改变的位置及动态频率等级可以发现,预测到真实发生报警位置的 13 次报警前后均以最小周期采集,详细记录了报警发生的过程.而错误预测报警的数据均接近于报警值,对故障发生过程分析具有对比参考意义.位置 69 报警没有预测到,但是通过动态频率算法以最小周期进行采集,而预测到位置 70 报警数据,因此其两次报警前后均使用最小周期采集到详细数据.动态频率等级与数据上升趋势成正比,因而数据采集密度随着劣化时间流逝加大,即劣化时间越长算法效果越好.

根据动态频率改变次数及级别可知,在两秒间隔采集的 146 个数据下,如果使用最小周期 0.2 s 采集以得到报警前后详细数据变化,需要采集 1 460 次,而通过动态频率算法则只需采集 330 次,即节省了 77.4% 的网络通信开销.

3.3 影响算法效果因素

对算法进行分析,得出影响算法的因素.分别以固定时间间隔 1 s、2 s、4 s 采集分级机电流样本数据,目的是分析不同采集频率对模型预测准度的影响大小.将采集到的数据进行预处理,包括提取趋势向、零化处理、标准化处理,使用 ARMA 模型并进行预测.预测数据与真实数据对比如图 7~9 所示.

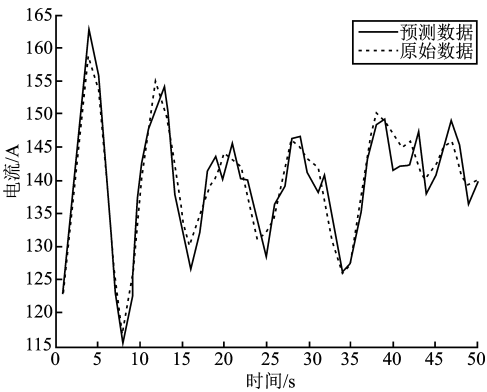


图 7 1 s 采集间隔预测

将采集到数据的真实值减去数据预测值得到绝对误差 Δu ,将绝对误差除以真实值从而得到误差比例.误差比例的均值和方差反应了预测效果的好坏,预测效果如表 6 所示.

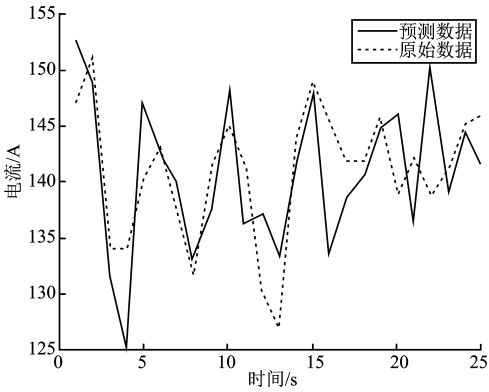


图 8 2 s 采集间隔预测

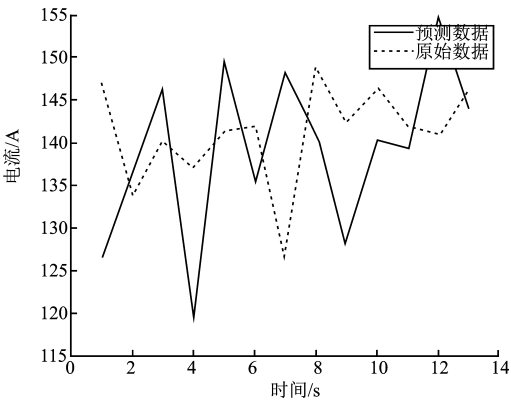


图 9 4 s 采集间隔预测

表 6 误差比例均值与标准差

采集间隔时间 $\Delta t/s$	1	2	4
误差比例均值	0.017 1	0.031 2	0.071 4
误差比例标准方差	0.010 6	0.012 6	0.048 2

从表 6 可以看出,误差比例均值及方差与采集间隔时间成正比,因此不适当的采集频率会对预测误差产生较大的影响.事实上,当 Δt 取得过大时,时间序列 $\{x_t\}$ 中将丧失观测信号中原有的相关关系,以致 ARMA 模型发生不应有的降阶^[7],从而造成预测误差显著增大.因此算法初始化最佳固定采集周期 T 时不能仅依据最小劣化时长,也要考虑到周期 T 下平均预测误差 Δu 的大小.当 Δu 大于 u 时会对数据上升等级 U 产生误判,从而选择错误的动态采集频率 f .此外,用来预测的样本长度过长,会增加建模计算量及造成建模时间过长,甚至参数求解精

度下降,而过短则不能反映统计特征,致使预测误差变大^[13],因此要考虑服务器实际计算性能选择合适样本长度。算法初始化过程,报警次数 n 取值越大,采集到最短劣化时长概率越大,算法效果越好。

4 结束语

本文使用动态 ARMA 预测模型实时预测,在设备劣化过程中根据数据变化动态改变数据采集频率,平均采集频率较小情况下可以预测报警并采集到报警附近详细的关键信息,对数据挖掘及故障诊断提供有效数据。实验结果证明,对渐变性故障,动态频率采集算法具有良好的效果。

参考文献:

[1] 蒋瑜等. 机械设备故障规律及运行趋势预测方法综述[J]. 机电一体化 2001,9(3): 14-17.

[2] 王衍学. 机械故障监测诊断的若干新方法及其应用研究[D]. 西安:西安交通大学 2009.

[3] 邵富杰. 远程故障诊断技术在数据采集系统中的应用[J]. 工业控制计算机,2007,20(7)15-20.

[4] 丁明. 基于时间序列分析的风电场风速预测模型[J]. 电力自动化设备,2005 25(8)32-34.

[5] Youqiang Sun. Prediction about time series based on updated prediction ARMA Model[C] // International Conference on Fuzzy System and Knowledge Discovery. Shenyang:IEEE,2013:680-684.

[6] Pappas S S, Ekonomou L, Karamousantas D Ch, et

al. Electricity demand loads modeling using Auto Regressive Moving Average (ARMA) models [J]. Energy, 2008, 33(9): 1353-1360.

[7] 杨叔子. 时间序列分析的工程应用[M]. 武汉:华中科技大学出版社,2007.

[8] 耿宏,刘晓萌,刘富学,等. 基于飞行计划的惯性导航误差模型分析[J]. 计算机测量与控制,2015, 23 (7) 2528-2531.

[9] 李向前. 复杂装备故障预测与健康管理关键技术研究[D]. 北京:北京理工大学,2014.

[10] 张树京. 时间序列分析简明教程[M]. 北京:清华大学出版社,2003.

[11] Chao Wu. Ambient Singals based Power System Oscillation Modes Identification considering Model Order Selection[C] // Power & Energy Society General Meeting. Calgary,AB: IEEE, 2009:1-7.

[12] 刘刚,郝敬敏,邝继双,等. 机械设备事故趋势与超限综合报警方法[J]河北农业大学学报,2000,23(4)88-92.

[13] 王可东,熊少锋. ARMA 建模及其在 Kalman 滤波中的应用[J]. 宇航学报,2012,33(8):1048-1055.

作者简介:

张 斌 男,(1970-),副教授. 研究方向为计算机网络与应用技术、计算机控制技术、网络安全.

朱建涛(通讯作者) 男,(1989-),硕士研究生. 研究方向为物联网应用.

E-mail:974813505@qq. com

徐 翌 男,(1991-),硕士研究生. 研究方向为工业监控.

(上接第 85 页)

作者简介:

陈远龙 男,(1992-),硕士研究生. 研究方向为模拟集成电路的研究. E-mail:1257569194@qq. com.

张 涛 男,(1992-),硕士研究生. 研究方向为模拟集成电路的研究.

张国俊 男,(1960-),教授. 研究方向为新型功率半导体器件与集成电路和系统设计.