全相关核偏最小二乘故障诊断方法及在抽油机上应用

汪 波[1], 夏钦锋[2], 钱 龙[1], 彭 军[1], 周 伟[1†]

(1. 重庆科技学院 智能技术与工程学院, 重庆 401331; 2. 中石化重庆涪陵页岩气勘探开发有限公司, 重庆 408000)

摘要: 针对油田抽油机生产数据存在强非线性和强耦合性, 导致故障诊断困难的问题, 本文提出一种全相关动态核偏最小二乘(FCDKPLS)故障诊断方法. 首先, 构建抽油机生产数据自回归模型, 反映数据变量间的动态特性; 其次, 分析了KPLS算法中输出变量与输入变量残差子空间的相关性, 为此, 在输出模型上构建一个辅助矩阵, 从而表征输入变量与输出变量的全相关性, 建立输入变量和输出变量之间更直接的联系. 最后, 将提出的全相关动态偏最小二乘方法应用于抽油机过程故障诊断, 实验结果表明本文提出方法的有效性.

关键词: 故障诊断; 核偏最小二乘; 全相关; 抽油机

引用格式: 汪波, 夏钦锋, 钱龙, 等. 全相关核偏最小二乘故障诊断方法及在抽油机上应用. 控制理论与应用, 2020, 37(9): 2039 – 2046

DOI: 10.7641/CTA.2020.90531

Fault diagnosis based on fully-correlated kernel partial least squares for pumping unit

WANG Bo[1], XIA Qin-feng[2], QIAN Long[1], PENG Jun[1], ZHOU Wei[1†]

(1. College of Intelligent Technology and Engineering, Chongqing University of Science and Technology, Chongqing 401331, China;
2. Sinopec Chongqing Fuling Shale Gas Exploration and Production Corporation, Chongqing 408000, China)

Abstract: Fault diagnosis of pumping unit system is a challenging issue owing to the system that exhibits strong nonlinearity and strong coupling of the production parameters. In this paper, a fault diagnosis method based on fully-correlated dynamic kernel partial least squares (FCDKPLS) is developed for pumping unit system. First, auto regressive model of the production data is constructed to obtain the dynamic performance between the production data of pumping unit. Then, the correlation between the output variable and the input residual subspace is studied by the KPLS method. To address this issue, an auxiliary matrix based on the output model is developed to represent the fully-correlated between the input variable and the output variable. In particular, a more direct link between the input variable and the output variable can be obtained. The proposed FCDKPLS algorithm is applied to the pumping unit system, and experimental results show the effectiveness of the proposed approach.

Key words: fault diagnosis; kernel partial least squares (KPLS); fully-correlated; pumping unit

Citation: WANG Bo, XIA Qinfeng, QIAN Long, et al. Fault diagnosis based on fully-correlated kernel partial least squares for pumping unit. *Control Theory & Applications*, 2020, 37(9): 2039 – 2046

1 引言

游梁式抽油机系统是石油生产中最常用的人工举升方法. 抽油机的采油过程是一个复杂的工业系统, 在抽油杆向地下移动的过程中存在许多未知因素. 在采油过程中, 由于抽油机系统的不稳定性, 容易导致故障发生. 一旦发生故障会导致抽油机采油量减少, 停产甚至损坏抽油设备. 然而, 抽油机复杂的工作环境和多变的井下条件使得诊断抽油机故障变得非常困难. 因此, 为抽油机系统建立精确有效的故障诊断模型是一个具有挑战性和重要意义的工作.

在实际的采油生产过程中, 示功图作为一条闭合曲线, 反映了抽油杆行程中载荷和位移的关系, 是监

收稿日期: 2019−07−05; 录用日期: 2020−05−12.

†通信作者. E-mail: zhouw@cqust.edu.cn; Tel.: +86 23-65023137.

本文责任编委: 张化光.

国家科技重大专项(2016ZX05060), 重庆市自然科学基金项目(cstc2019jcyj–msxmX0080), 重庆市教委科学技术研究基金项目(KJQN201801506)资助.

Supported by the National Science and Technology Major Project of China (2016ZX05060), the Natural Science Foundation of Chongqing (cstc2019jcyj–msxmX0080) and the Science and Technology Research Program of Chongqing Municipal Education Commission (KJQN201801506).

视抽油机井下运行状况的主要工具[1]. 在传统的抽油机故障诊断中, 工程师通常通过分析示功图的形状来判断井下工作状况, 诊断的准确性取决于工程师的经验判断.

近年来, 人工智能方法已广泛用于抽油机故障识别与诊断. 这些方法包括粗糙集理论[2]、人工神经网络[3–4]、支持向量机[5–6]等. 但是, 这些方法主要使用载荷和位移数据来诊断井下工作条件. 如今, 随着油田信息化建设的发展, 油田企业已经积累了大量的生产数据, 包括电压、电流、油压、采出液量、功率因数和泵效率等. 在抽油机采油过程生产中, 这些大量的生产数据为抽油机故障诊断提供了可能.

近年来，数据驱动的故障诊断方法广泛应用于工业系统诊断中[7–8]. 其中, 多变量统计过程监控(multivariate statistical process monitoring, MSPM)具有较强的变量间相关性解释能力, 被广泛用于复杂工业过程的故障诊断[9–10]. MSPM方法通常限于显著利用的偏最小二乘(partial least squares, PLS)的方法[11–12]. 基于PLS模型, 已经开发了许多成功的监视方法[13–14]. 这些方法从大量输入和输出数据中提取潜在特征, 从而可以有效地消除回归中的无效噪声, 用以诊断工业过程中的故障. 同时, PLS可以提取更少的组件来解释更多与变量相关的问题. 近年来, 李刚等人[15]揭示了用于过程监控的PLS的几何性质. 根据李刚的结果, 周东华等人[16]首先分析了用于质量相关故障检测的PLS的固有缺陷, 并提出了具有更详细分解过程变量矩阵的总PLS(total PLS, T-PLS)模型. 但是油井的生产过程表现出非线性、参数耦合和时变等特点, 因此很难使用传统PLS算法建立准确的抽油机故障诊断模型[17].

为了解决过程数据的非线性问题, Rosipal等[18]人提出了一种非线性核PLS(kernel PLS, KPLS)方法, 将非线性输入数据映射到高维特征空间中. 彭开香等[19]人提出了一种基于KPLS模型的与质量有关的非线性故障检测方法TKPLS(total KPLS, TKPLS). 文献[20]提出了一种改进的KPLS算法, 以提高质量相关故障的检测精度, 并降低质量无关故障的误报率. 文献[21–22]提出了一种基于最优偏好矩阵的改进KPLS方法来解决主成分误解的问题. 最优偏好矩阵用于调整过程变量的分布和协方差矩阵的特征值. 文献[23]中, 提出了一种基于多块KPLS的分布式故障诊断方法来监视大型工业过程. 文献[24]开发了一种时间片批处理监视方法来解决线性和非线性变量的问题.

尽管KPLS和相应的扩展方法已广泛用于过程监视和故障诊断, 仍亟需解决以下问题: 1) 在KPLS模型中, 通过离线数据得到的静态数学模型不能及时反映数据之间的相关性, 且不能准确描述与变量相关联的潜在特征的变化. 2) 由于时变特性和工业变量参数的多重相关性, 传统KPLS不能完全提取变量之间的隐藏特性, 导致诊断精度低.

受到多元统计方法的启发, 本文提出了一种新颖的全相关动态核偏最小二乘(fully-correlated dynamic kernel partial least squares, FCDKPLS)故障诊断算法, 并应用于抽油机系统的故障诊断. 首先, 基于自回归模型(auto regressive, AR)建立输入和输出变量的强动态相关特性, 从而反映变量之间隐藏的动态关系; 其次, 通过分析证明KPLS模型的输出变量影响到输入残差子空间, 为此构建输出变量辅助矩阵, 表征输出与输入向量全相关, 从而直接反映输入与输出变量的关联性. 实验结果表明, 本文提出的方法在抽油机故障诊断上具有良好的诊断效果.

2 自回归模型

在抽油机系统中, 一个因变量总是和多个自变量有关, 且数据中输入变量与输出向量之间存在着时序相关性. 为表征抽油机数据动态特性, 对数据建立AR模型:

$$x(\vartheta) = \alpha_1 x(\vartheta-1) + \alpha_2 x(\vartheta-2) + \cdots + \alpha_h x(\vartheta-h) + \varepsilon_\vartheta, \tag{1}$$

其中: $x(\vartheta)$为抽油机数据变量, $\alpha_1, \alpha_2, \cdots, \alpha_n$为模型回归系数, ε_ϑ为模型随机误差, h为模型阶次. 令

$$\alpha = [\alpha_1\ \alpha_2\ \cdots\ \alpha_h]^{\mathrm{T}},$$

$$Y = [x(h+1)\ \ x(h+1)\ \ \cdots\ \ x(n)]^{\mathrm{T}},$$

$$X = \begin{bmatrix} x(h) & x(h-1)\cdots & x(1) \\ x(h+1) & x(h) \quad \cdots & x(2) \\ \vdots & \vdots & \vdots \\ x(n-1) & x(n-2)\cdots x(n-h) \end{bmatrix},$$

则AR模型表示为

$$Y = X\alpha + \varepsilon_\vartheta, \tag{2}$$

从而可以求得模型系数矩阵.

AR模型解决自相关问题在于确定模型阶次h, 可以根据贝叶斯信息准则来确定最合适的阶数. 若要得到更为精确的时滞阶次, 可利用统计假设检验[25]判断各个变量是否具有自相关性.

通过构建AR模型, 反映数据变量间的动态关系, 然后对动态扩展数据进行KPLS分析.

3 FCDKPLS算法

3.1 KPLS算法

KPLS通过非线性映射函数$\Phi(\cdot)$将其转换到高维特征空间F中, 在特征空间中构建PLS回归模型. 设$\{(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)\} \subset \mathbb{R}^{L_1} \times \mathbb{R}^{L_2}$, N为训练样本个数, L_1为输入变量个数, L_2为输出变量个数. 输入与输出向量分别建立回归模型如下:

$$\begin{cases} \Phi(x)=\sum_{i=1}^{n}t_i p_i^{\mathrm{T}}+\varphi_i(x), \\ Y=\sum_{i=1}^{n}t_i q_i^{\mathrm{T}}+\varepsilon_i, \end{cases} \tag{3}$$

其中: i表示KPLS中保留的隐变量个数, 通过交叉检验得到; t_i为输入向量的得分矩阵, p_i为输入向量的负载矩阵, q_i为输出向量负载矩阵, φ_i和ε_i分别为输入和输出向量残差矩阵. KPLS算法流程如算法1所示[26].

算法 1　KPLS算法.

1) 初始化u_i;

2) 计算$\Phi(x)$的得分矩阵:

$$t_i=\Phi(x)\Phi^{\mathrm{T}}(x)u_i=Ku_i;$$

3) 计算输出负载矩阵: $P_i=Y^{\mathrm{T}}t_i/\|t_i\|$;

4) 计算负载主元u_i: $u_i=YP_i$;

5) 重复2)–4)步直至t_i收敛;

6) 计算残差矩阵K和Y:

$$K_{i+1}=(I-t_it_i^{\mathrm{T}})K(I-t_it_i^{\mathrm{T}}),$$
$$Y=Y-(I-t_it_i^{\mathrm{T}})Y;$$

7) $i=i+1$, 返回步骤2).

3.2　KPLS相关性分析

在KPLS中, 输入残差子空间$\varphi_i(x)$与输出向量Y存在着相关性:

$$\varphi_i^{\mathrm{T}}(x)Y=\varphi_i^{\mathrm{T}}(x)(\sum_{i=1}^{n}t_iq_i^{\mathrm{T}}+\varepsilon_{i-1}), \tag{4}$$

其中: t_i为KPLS中的主元空间部分, $\varphi_i(x)$为KPLS中的残差部分. 当$i\geqslant l$时

$$\varphi_i^{\mathrm{T}}(x)t_l=0. \tag{5}$$

结合式(5), 式(4)可以表示为

$$\varphi_i^{\mathrm{T}}(x)Y=\varphi_i^{\mathrm{T}}(x)\varepsilon_{i-1}=(\varphi_{i-1}(x)-t_ip_i^{\mathrm{T}})^{\mathrm{T}}\varepsilon_{i-1}. \tag{6}$$

在KPLS中输入和输出向量的负载向量p_i和q_i分别可以写为

$$p_i=\varphi_{i-1}^{\mathrm{T}}(x)t_i/t_i^{\mathrm{T}}t_i, \tag{7}$$
$$q_i=\varepsilon_{i-1}^{\mathrm{T}}t_i/t_i^{\mathrm{T}}t_i. \tag{8}$$

将p_i和q_i代入式(6)得到

$$\varphi_i^{\mathrm{T}}(x)Y=\varphi_i^{\mathrm{T}}(x)\varepsilon_{i-1}=\varphi_{i-1}^{\mathrm{T}}(x)\varepsilon_{i-1}-\varphi_{i-1}^{\mathrm{T}}(x)t_iq_i^{\mathrm{T}}=\varphi_{i-1}^{\mathrm{T}}(x)(\varepsilon_{i-1}-t_iq_i^{\mathrm{T}}). \tag{9}$$

通过式(9)可知输入残差与输出变量之间具有一定的关系. 具体关系如图1中的步骤2所示, $\varphi_i(x)$与ε_i呈垂直状态. 假定残差向量子空间与输出向量之间没有关联, 即$\varepsilon_i-t_iq_i^{\mathrm{T}}=0$. 由于$t_i$与$\varepsilon_i$为垂直状态, 说明在模型进行第$i-1$次迭代时$t$与$\varepsilon$正交, $\varepsilon_{i-1}=t_{i-1}q_{i-1}^{\mathrm{T}}$; 同理, 第$i$次迭代时, $\varepsilon_i=t_iq_i^{\mathrm{T}}$. 由于模型的迭代次数不同, 模型内部第$i$次迭代时的主元空间与第$i-1$次迭代时的残差空间之间的联系不清楚, 从而无法确保t_i与ε_{i-1}是否满足正交原则. 实际上, $\varepsilon_{i-1}\neq t_iq_i^{\mathrm{T}}$. 因此, KPLS中输入变量残差子空间与输出变量之间存在着相关性.

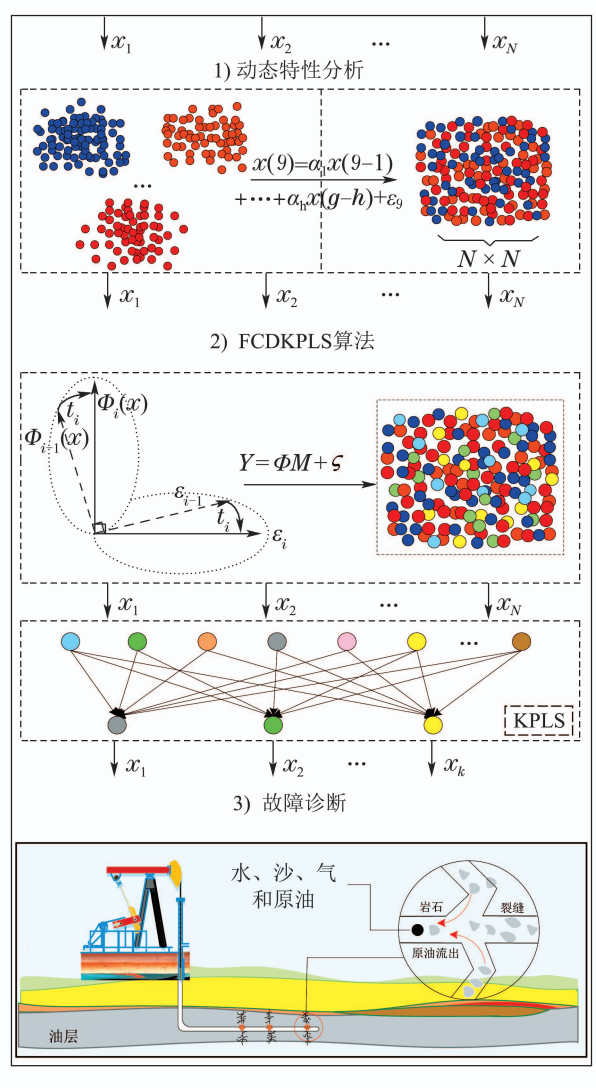


图 1　基于FCDKPLS的故障检测过程

Fig. 1　Fault detection process based on FCDKPLS

3.3　输出变量全相关辅助矩阵

把生产数据分为输入$\Phi(x)\in\mathbb{R}^{n\times L_1}$和输出$Y\in\mathbb{R}^{n\times L_2}$2个矩阵, 并在输出变量中构建辅助矩阵$M$, 包含满$\Phi(x)$与$Y$的相关性, 建立与输出变量相关的全相关矩阵, 构建如下回归模型:

$$Y=\Phi(x)M+\varsigma, \tag{10}$$

其中: M为构建的辅助分解矩阵, 将原始输出向量映射到特征空间中重新进行输入与输出向量的非线性分解; ς为$\Phi(x)$完全不相关的噪音或干扰.

对于M的构建, 有

$$\mathrm{cov}(\varsigma,\varphi(x))=\xi\{\varsigma\varphi^{\mathrm{T}}(x)\}=0. \tag{11}$$

因为ς与$\Phi(x)$完全无关, 有

$$\frac{1}{N}Y^{\mathrm{T}}\Phi(x)=\frac{1}{N}M^{\mathrm{T}}\Phi^{\mathrm{T}}(x)\Phi(x)\approx M^{\mathrm{T}}\frac{\Phi^{\mathrm{T}}(x)\Phi(x)}{N}. \tag{12}$$

因此, M具体可表示为

$$M=(\Phi^{\mathrm{T}}(x)\Phi(x))^{\dagger}\Phi^{\mathrm{T}}(x)Y, \tag{13}$$

其中$(\Phi^{\mathrm{T}}(x)\Phi(x))^{\dagger}$是$\Phi^{\mathrm{T}}(x)\Phi(x)$的伪逆矩阵. 要计算其伪逆矩阵, 需对$\Phi^{\mathrm{T}}(x)\Phi(x)$进行奇异值分解(singular value decomposition, SVD)[27]运算, 得到以下方程式:

$$\Phi^{\mathrm{T}}(x)\Phi(x)=[A_1\ A_2]\begin{bmatrix}\Lambda_1&0&0\\0&\Lambda_2&0\\0&0&\Lambda_3\end{bmatrix}\begin{bmatrix}A_1^{\mathrm{T}}\\*\end{bmatrix}=A_1\Lambda A_1^{\mathrm{T}}, \tag{14}$$

式中:

$$A_1=[a_1\ \cdots\ a_{pc}],$$
$$A_2=[a_{pc+1}\ \cdots\ a_N],$$
$$\Lambda=\mathrm{diag}\{\lambda_1,\cdots,\lambda_{pc}\},$$

pc是非零的奇异值的数量.

根据SVD的特性, 可以得到

$$A_1^{\mathrm{T}}A_1=I_{pc}. \tag{15}$$

假定

$$\begin{aligned}\Psi&=\bar{\Phi}^{\mathrm{T}}(x)\bar{\Phi}(x),\\ \Pi&=A_1\Lambda^{-1}A_1^{\mathrm{T}},\end{aligned} \tag{16}$$

则有

$$\begin{aligned}\Psi\Pi\Psi&=A_1\Lambda A_1^{\mathrm{T}}A_1\Lambda^{-1}A_1^{\mathrm{T}}A_1\Lambda A_1^{\mathrm{T}},\\ \Pi\Psi\Pi&=A_1\Lambda^{-1}A_1^{\mathrm{T}}A_1\Lambda A_1^{\mathrm{T}}A_1\Lambda^{-1}A_1^{\mathrm{T}}.\end{aligned} \tag{17}$$

按照伪逆的定义, Π是Ψ的伪逆. 因此,

$$(\Phi^{\mathrm{T}}(x)\Phi(x))^{\dagger}=A_1\Lambda^{-1}A_1^{\mathrm{T}}. \tag{18}$$

实际上, 由于$\Phi(x)$可以任意大甚至是无限大, 因此上述公式不能直接用于计算. 为了避免使用$\Phi(x)$, 定义以下内核矩阵K是一种常见的方法, 高斯核函数[27]进行运算:

$$K=\Phi(x)\Phi^{\mathrm{T}}(x)=\exp(-\frac{\|x-y\|^2}{c}). \tag{19}$$

为说明构建的辅助分解矩阵能表征输入和输出变量之间的相关性, 这里采用拉格朗日乘子法进行分析证明.

首先, 分别标准化$\Phi(x)$和Y为X_0和Y_0, 拉格朗日乘子法如下:

$$f=\mu_1^{\mathrm{T}}X_0^{\mathrm{T}}Y_0\varpi_1-\nu(\mu_1^{\mathrm{T}}\mu_1-1)-\tau(\varpi_1^{\mathrm{T}}\varpi_1-1), \tag{20}$$

其中: μ_1和ϖ_1分别是输入变量X_0和输出变量Y_0的轴向量, 对f分别求关于μ_1,ϖ_1,ν,τ的偏导且置0, 记

$$\theta=2\nu=2\tau.$$

在KPLS推导中, θ是优化问题的目标函数且使θ达到最大必须有

$$\begin{cases}\varphi_{i-1}^{\mathrm{T}}Y_0\varpi_i=\theta\mu_i,\\ Y_0^{\mathrm{T}}\varphi_{i-1}\mu_i=\theta\varpi_i.\end{cases} \tag{21}$$

由上式得出

$$\varphi_{i-1}(x)\varphi_{i-1}^{\mathrm{T}}(x)Y_0\varpi_i=\theta\varphi_{i-1}(x)\mu_i, \tag{22}$$
$$Y_0Y_0^{\mathrm{T}}\varphi_{i-1}(x)\mu_i=\theta Y_0\varpi_i. \tag{23}$$

由式(22)得到

$$\varphi_{i-1}(x)\mu_i=\varphi_{i-1}(x)\varphi_{i-1}^{\mathrm{T}}(x)Y_0\varpi_i/\theta. \tag{24}$$

将式(24)代入式(23), 结合$Y_0\varpi_i{=}t_i$, $p_i{=}\varphi_{i-1}^{\mathrm{T}}t_i/t_i^{\mathrm{T}}t_i$, $\theta^2=t_i^{\mathrm{T}}t_i$得

$$Y_0Y_0^{\mathrm{T}}\varphi_{i-1}(x)p_i=t_i. \tag{25}$$

假定输入与输出变量全相关, KPLS 残差子空间$\varphi_i(x)$可以表示为

$$\varphi_i^{\mathrm{T}}(x)Y=(\Phi(x)-YY^{\mathrm{T}}\Phi(x)/Y^{\mathrm{T}}Y)^{\mathrm{T}}(\varphi+\varepsilon). \tag{26}$$

由于$\varphi=\Phi(x)M$, 且ε为$\Phi(x)$完全不相关的噪音或干扰, 式(26)可以写成

$$\begin{aligned}&\varphi_i^{\mathrm{T}}(x)Y=\\ &(\Phi(x)-YY^{\mathrm{T}}\Phi(x)/Y^{\mathrm{T}}Y)^{\mathrm{T}}\Phi(x)M=\\ &\Phi^{\mathrm{T}}(x)Y-\Phi^{\mathrm{T}}(x)Y=0.\end{aligned} \tag{27}$$

因为$\Phi^{\mathrm{T}}(x)\Phi(x)$是一个满秩方阵, 所以

$$(\Phi^{\mathrm{T}}(x)\Phi(x))^{\dagger}\Phi^{\mathrm{T}}(x)\Phi(x)=I,$$

其中I为单位矩阵. 因此, 得到输入残差与输出变量的结果收敛, 使构造的辅助矩阵可以表征输入向量与输出变量之间全相关性.

3.4 FCDKPLS算法

上述理论推导了输入变量与输出变量存在全相关性, 依据AR动态特性, 构造辅助分解矩阵模型, 使数据间输入和输出变量之间的相关性被充分挖掘. 故障诊断步骤包括离线主元模型、在线故障诊断, FCDKPLS算法流程如算法2所示.

算法 2 FCDKPLS算法.

1) 构建AR模型系数矩阵α;

2) 计算全相关辅助矩阵M:

$$M=(\Phi^{\mathrm{T}}(x)\Phi(x))^{\dagger}\Phi^{\mathrm{T}}(x)Y;$$

3) 初始化u_i;

4) 计算输入向量得分矩阵:

$$t_i = \Phi(x)\Phi^{\mathrm{T}}(x)u_i = Ku_i;$$

5) 计算输出向量负载矩阵: $P_i = M^{\mathrm{T}}t_i/\|t_i\|$;

6) 计算负载主元u_i: $u_i = MP_i$;

7) 重复3)–6)步直至t_i收敛;

8) 计算残差矩阵K和Y:

$$K_{i+1} = (I - t_i t_i^{\mathrm{T}})K(I - t_i t_i^{\mathrm{T}}),$$
$$M = M - t_i t_i^{\mathrm{T}}M;$$

9) $i = i + 1$, 返回步骤4).

首先, 对归一化后的变量进行动态特性分析, 以获得变量的隐藏动态关系. 其次, 构建FCDKPLS模型, 使数据间在输入残差与输出变量之间提取更多地潜在变量. 最后, 将FCDKPLS模型用于实时数据分析, 并将获得的主元用于故障诊断. FCDKPLS算法的故障诊断与监测过程流程图的具体步骤如图1所示.

根据图1所得FCDKPLS故障检测过程流程图, 分别给出离线建模和在线监测步骤:

A 根据正常情况建模.

① 获得观测数据子集;

② 用每个变量的均值和标准偏差来规范化数据子集;

③ 根据规范化的数据计算核矩阵, 并计算中心化核矩阵K;

⑤ 进行时滞阶次分析, 提取动态非线性关系;

⑥ 建立全相关辅助矩阵M, 结合KPLS方法得到主元;

⑦ 确定控制界限.

B 在线监测.

① 在线采集观测变量数据;

② 用每个变量的均值和标准差来规范化数据子集;

③ 计算规范化后的在线观测数据的核向量, 计算中心化核向量;

④ 时滞阶次分析, 提取动态非线性关系;

⑤ 得到的统计量与模型得到的控制限进行比较, 当超出控制限后, 进行故障源的追溯与判别.

4 基于FCDKPLS算法的故障检测方法

通过上面算法求出m个主元的隐变量的得分向量$T = [t_1\ t_2\ \cdots\ t_m]$, $U = [u_1\ u_2\ \cdots u_m]$以及负载向量$P = [p_1\ p_2\ \cdots\ p_m]$, $Q = [q_1\ q_2\ \cdots q_m]$, 再选取了其中$m$个主元后得到$\Phi(x)$和输出$Y$的重构数据$\hat{\Phi}$和$\hat{Y}$及相应的重构误差矩阵.

定义在第i时刻的平方预测误差 (squared prediction error, SPE)可以写为

$$\mathrm{SPE}(i) = \sum_{j=1}^{m}(\bar{\Phi}_{i,j} - \hat{\Phi}_{i,j})^2, \tag{28}$$

其中: $\bar{\Phi}_{i,j}$为$\bar{\Phi}_{n\times n}$中i时刻第j个变量的测量值, $\hat{\Phi}_{i,j}$为第i时刻第j个变量的重构值, i=1, 2, $\cdots$, n, j=1, 2, $\cdots$, n. 当检验水平为α时, SPE第i时刻的SPE(i)的控制限可由下面权重χ^2分布来计算:

$$\begin{cases} \mathrm{SPE}(i)_\alpha = g\chi^2, \\ g = b/2a, \\ h = 2a^2/b, \end{cases} \tag{29}$$

其中: g是一个加权参数, h是自由度, a和b分别是SPE(i)的估计均值和方差.

对过程监测数据矩阵标准化后记为$\bar{\Phi}_{n\times n}$, 对于$\bar{\Phi}_{n\times n}$中第i时刻过程变量向量T^2可表示为

$$T_i^2 = t_i\Lambda^{-1}t_i^{\mathrm{T}}, \tag{30}$$

Λ是得分矩阵的协方差, 则T^2统计量控制限计算公式为

$$T_{k,i,\alpha}^2 = \frac{k(n-1)}{n-k}F_{k,n-1,\alpha}, \tag{31}$$

式中: α为显著性水平, n为数据采样次数.

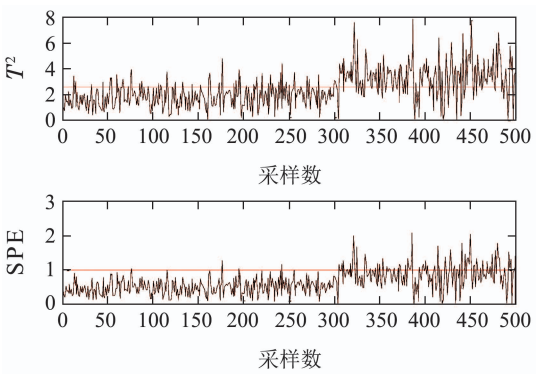
5 实验研究

在本节中, 通过大港油田实际生产数据来验证本文提出的FCDKPLS算法的有效性. 在石油开采过程中, 当抽油机中某个阀门或者某处电压不稳定时, 都会引起连锁故障, 从而降低采油效率, 甚至损坏设备, 带来严重的经济损失和人员伤亡. 因此, 需要对抽油机采油过程中的各个参数进行实时监控, 及时监测可能发生的故障, 从而有效地减少故障, 实现高效采油.

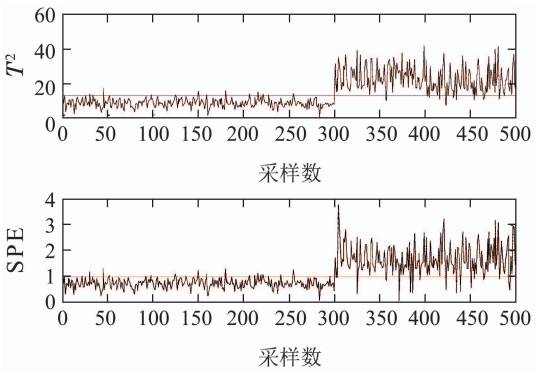
本文主要以抽油机生产过程中的3种典型故障为例进行实验分析: 故障1是连抽带喷故障, 故障2是疑是杆脱落故障, 故障3是气体影响故障. 连抽带喷故障是由于油井的自喷能力, 在抽油过程中, 柱塞基本不受载荷作用, 示功图载荷和位移的大小取决于油井喷势的强弱和原油的粘度. 在此过程中主要由出口阀漏油影响, 故障发生时影响电压、电流、有效冲程、喷射和产生流体量等参数. 疑是杆脱落故障是由于抽油杆弹性疲劳、抽油杆丝扣没有上紧、抽油泵遇卡等原因使得抽油杆超过其拉伸屈服极限导致的. 由于摩擦力的作用, 使得上、下载荷线不重合. 在此过程中主要由载荷影响, 故障发生时影响电压、电流、有效冲程、泵输出和最大、最小载荷等参数. 气体影响故障是当抽油泵中的油液混入较多气体时, 在上冲程过程中, 由于油液内混有气体的因素, 使泵腔内压力在该下降的时候不能正常下降, 导致固定凡尔开启时间延后, 使得抽油杆加载速度变慢; 在下冲程过程中, 泵腔内的压力变化和上冲程过程恰好相反. 在此过程中主要由压力影响, 故障发生时影响电流、冲程、泵效率和有功功率等参数.

抽油机采油过程中产生了大量的生产数据, 本文主要提取示功图数据、电流、电压等生产数据用于FCDKPLS算法建立故障诊断模型. 笔者从中石油大港油田2017年1月1日至2017年12月31日生产数据中提取构建了2000组数据集, 其中训练数据集由1500组数据组成, 测试数据集由500组数据组成. 由于示功图由144个载荷和位移数据构成, 直接用于建模会导致输入数据过大. 因此, 首先利用傅里叶描述子方法对示功图数据进行特征提取, 提取出21个特征数据表征示功图特征[28]. 每个数据集由41个变量组成，其中32个变量作为输入变量, 包括油压、套压、采出液量、出口阀漏油、喷射、流体量、有效冲程、泵输出、最大负荷、最小负荷和21个示功图特征数据. 9个变量作为输出变量, 包括三相电流、三相电压、功率、泵效率和载荷. 在训练数据集中, 前1000个数据集是正常的, 后500个数据集是故障样本. 在测试数据集中, 前300个数据集是正常的, 后200个数据集是故障样本.

对于故障1, 检测结果如图2所示.



(a) KPLS



(b) FCDKPLS

图 2 故障1的检测结果比较

Fig. 2 Comparison of detection result in fault 1

当泵效率发生变化时, 从图2(a)可以得出, KPLS中T^2 和SPE统计量均在第300个样本处发生跳变. 在T^2中, 正常样本大多数能低于控制限; 在故障样本中, 较多故障样本低于控制限, 故障诊断效果较差. 在SPE中, 几乎所有正常样本低于控制限. 但在故障样本中, 大多数故障样本低于控制限, 致使检测效果差. 图2(b)中, FCDKPLS中均能在第300个样本处检测到故障. 同时, T^2和SPE正常样本超过控制限均很少, 故障样本绝大多数超过控制限, 诊断效果较好.

图2(b)可以成功检测到故障, 证明输入向量残差子空间的变化确实受到输出变量的影响, 检测故障精度高. 但当检测到故障后无法对故障源进行定位. 因此, 本文引用贡献率来进行故障定位, 而传统的贡献图在进行多变量故障定位时, 故障变量会受到与之相关性较强的变量的干扰, 存在某些时刻正常变量贡献值大于故障变量的情况, 容易得到错误的结果. 在此基础上, 本文引用累积贡献率对故障源进行追溯定位, 以此来确定故障源. 第i时刻j个样本数据的累积残差贡献率[29]定义为

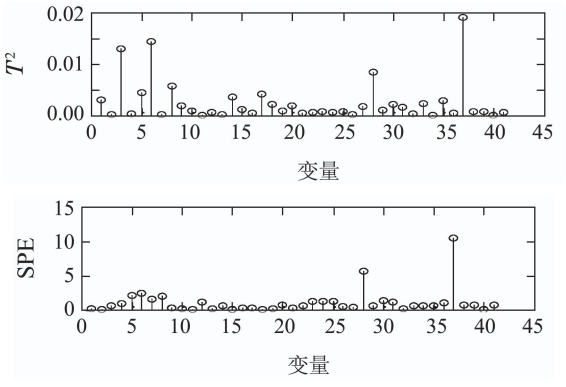
$$\operatorname{cont}\left(\mathrm{SPE}\right)_j=\frac{\sum\limits_{i=1}^{N}\operatorname{cont}\left(\mathrm{SPE}\right)_{i,j}}{N}. \tag{32}$$

基于T^2统计量的贡献率定义如下[30]:

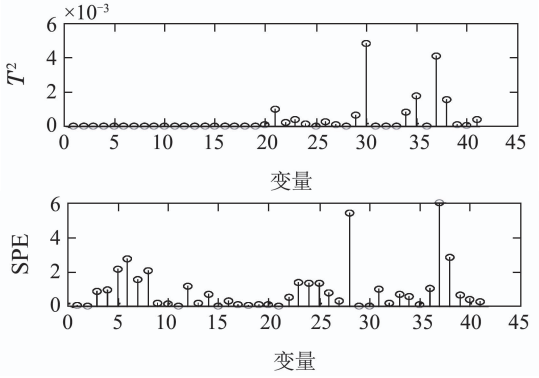
$$T^2=x^{\mathrm{T}}DI_mx, \tag{33}$$

式中$D=P^{\mathrm{T}}\Lambda^{-1}P$.

在故障1中, 引起连抽带喷故障的主要原因是由出口阀漏油引起的, 与其相关的故障变量有电压、电流、有效冲程、喷射和产生流体量等参数. 从图3中可以看出, 图3(a)的T^2和SPE统计量的故障源均为喷射(变量37)引起, 而图3(b)在T^2中故障由出口阀漏油(变量30)引起, SPE中由喷射(变量37)引起. 而在实际采油过程中, 故障1往往由出口阀漏油所致, 即变量30引起, 所以本文提出的方法可以准确定位故障源.



(a) KPLS

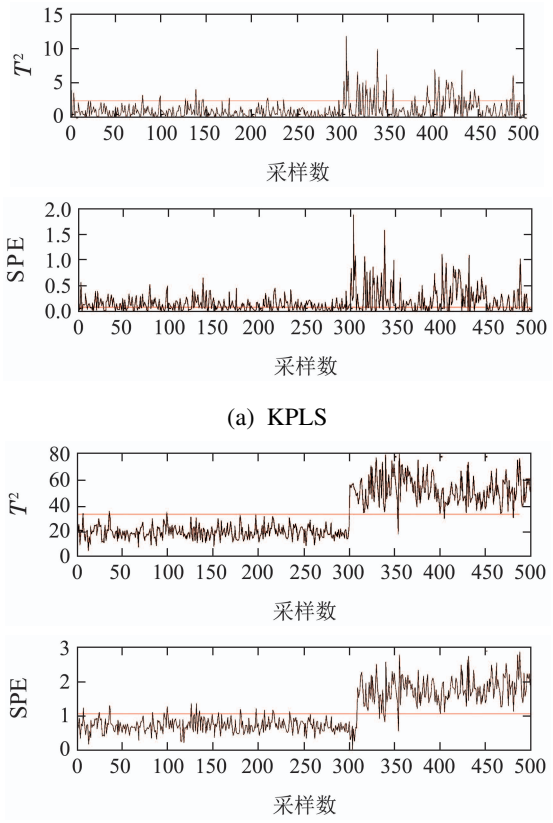


(b) FCDKPLS

图 3 故障1的变量贡献率比较

Fig. 3 Comparison of variables contribution rate in fault 1

对于故障2, 监测结果如图4所示. 当载荷发生变化时, 图4(a)中T^2和SPE统计量均能在故障点300处发生变化, 但T^2中绝大多数故障样本未超过控制限, 不能良好进行故障诊断. SPE中, 正常样本超过控制限较多, 同时故障样本未超过控制限也比较多, 在检测过程中不能良好的进行故障诊断. 不能正常进行故障检测. 如图4(b), T^2和SPE统计量均有良好的故障诊断效果, 与KPLS相比, FCDKPLS方法在监测故障2方面更有效. 同时, 也证明了在提取出所有相关特征后, 故障诊断的精度明显增加.



(a) KPLS

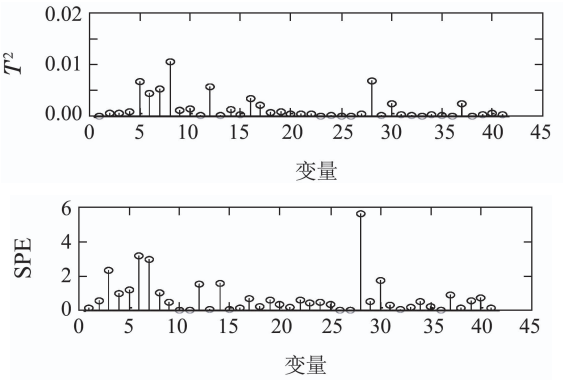
(b) FCDKPLS

图 4 故障2的检测结果比较

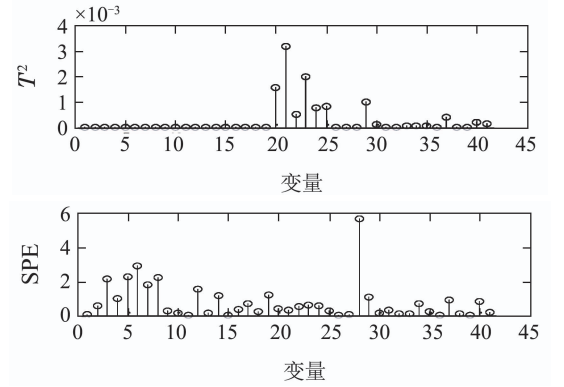
Fig. 4 Comparison of detection result in fault 2

在故障2中, 引起疑是杆脱落故障的主要原因是由载荷引起的, 与其相关的故障变量有电压、电流、有效冲程、泵输出和最大、最小负荷等参数. 从图5中可以看出, 图5(a)中T^2的故障源为有效冲程(变量8), 而SPE的故障源为最小载荷(变量28); 图5(b)在T^2中故障由最大载荷(变量21)引起, SPE中由最小载荷(变量28)引起. 在疑是杆脱落故障中, 故障源往往由最大载荷所致, 即变量21所引起, 从而验证所提出的方法可以对故障进行检测并帮助其对故障源进行准确定位.

通过对抽油机生产过程数据进行统计过程监控, 利用KPLS和FCDKPLS算法得到的各变量的贡献率, 最终用T^2和SPE统计量的误报率(FPR)和漏报率(FNR)作为衡量监控性能的指标, 结果如表1所示. 在多故障分类中, 计算故障的误报率和漏报率时, 其他故障样本被认为是“正常样本”. 所以, 误报警率是被识别为正常的“故障样本”的数量与正常样本的总数的比率; 漏报警率是故障的“正常样本”的数量与故障样本的总数的比率. 由表分析可知, 基于FCDKPLS方法的多元统计过程监控方法在抽油机生产过程故障检测与诊断方面具有良好的监控性能, 适用于复杂的非线性和动态工业过程.



(a) KPLS



(b) FCDKPLS

图 5 故障2的变量贡献率比较

Fig. 5 Comparison of variables contribution rate in fault 2

表 1 误报率、漏报率(%)比较

Table 1 Comparison of FPR and FNR

故障	KPLS				FCDKPLS			
	T^2		SPE		T^2		SPE	
	FPR	FNR	FPR	FNR	FPR	FNR	FPR	FNR
1	19.00	29.50	1.33	61.50	4.33	8.00	7.33	9.50
2	14.67	63.50	51.00	42.00	1.00	1.50	6.00	6.50
3	25.35	27.15	1.84	30.27	8.67	9.00	7.67	6.00

6 结论

本文提出一种全相关动态核偏最小二乘的抽油机故障诊断方法. 首先, 分析抽油机数据的动态特性, 建立了自回归模型, 有效挖掘数据间存在的潜在变量, 使数据间具有强动态性; 其次, 分析证明了KPLS输入

残差子空间与输出变量之间具有相关性, 在输出变量上构造辅助矩阵, 使输入变量残差矩阵与输出变量无关, 得到输入变量与输出变量全相关特性. 实验结果证明了FCDKPLS监测方法比传统KPLS监测方法表现出更好的监测性能, 表明所提出的FCDKPLS监测方法对于抽油机过程监测的有效性.

参考文献:

[1] TIAN Haifeng, YU Xianchuan. Design of the indicator diagram sensing system based on position sensing and displacement multiplexing. *Chinese Journal of Scientific Instrument*, 2019, 40(3): 172 – 180. (田海峰, 余先川. 基于位置感知和位移复用的示功图传感系统设计. 仪器仪表学报, 2019, 40(3): 172 – 180.)

[2] WANG J P, BAO Z F. Study of pump fault diagnosis based on rough sets theory. *The 3rd International Conference on Innovative Computing Information and Control (ICICIC)*. Dalian: IEEE, 2008, 6: 18 – 20.

[3] REN W J, TIAN Y C, ZHU Y B. Application of CS neural network in-pumping units' fault diagnosis. *Journal of Bionic Engineering*, 2017, 35(3): 324 – 332.

[4] ZHANG L, TAN Z, LIU C, et al. Research on optimal operation method of pumping station based on machine learning. *2017 IEEE Conference on Energy Internet and Energy System Integration*. Beijing: IEEE, 2017: 1 – 6.

[5] LI K, GAO X W, TIAN Z, et al. Using the curve moment and the PSO-SVM method to diagnose downhole conditions of a sucker rod pumping unit. *Petroleum Science*, 2013, 10(1): 73 – 80.

[6] YU D L, ZHANG Y M, BIAN H M, et al. A new diagnostic method for identifying working conditions of submersible reciprocating pumping systems. *Petroleum Science*, 2013, 10(1): 81 – 90.

[7] HAO H, ZHANG K, DING S X, et al. A data-driven multiplicative fault diagnosis approach for automation processes. *ISA Transactions*, 2014, 53(5): 1436 – 1445.

[8] HU Z, CHEN Z, GUI W, et al. Adaptive PCA based fault diagnosis scheme in imperial smelting process. *ISA Transactions*, 2014, 53(5): 1446 – 1455.

[9] JING Q, YAN X, HUANG B. Performance-driven distributed pca process monitoring based on fault-relevant variable selection and bayesian inference. *IEEE Transactions on Industrial Electronics*, 2016, 63(1): 377 – 386.

[10] LI G, ALCALA C F, QIN S J, et al. Generalized reconstruction-based contributions for output-relevant fault diagnosis with application to the Tennessee Eastman process. *IEEE Transactions on Control Systems Technology*, 2010, 19(5): 1114 – 1127.

[11] MACGREGOR J F, JAECKLE C, KIPARISSIDES C, et al. Process monitoring and diagnosis by multiblock PLS methods. *AIChE Journal*, 1994, 40(5): 826 – 838.

[12] YI J, HUANG D, FU S, et al. Optimized relative transformation matrix using bacterial foraging algorithm for process fault detection. *IEEE Transactions on Industrial Electronics*, 2016, 63(4): 2595 – 2605.

[13] YIN S, DING S X, HAGHANI A, et al. A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark tennessee eastman process. *Journal of Process Control*, 2012, 22(9): 1567 – 1581.

[14] QIN S J. Survey on data-driven industrial process monitoring and diagnosis. *Annual Reviews in Control*, 2012, 36(2): 220 – 234.

[15] LI G, QIN S J, ZHOU D. Geometric properties of partial least squares for process monitoring. *Automatica*, 2010, 46(1): 204 – 210.

[16] ZHOU D, LI G, QIN S J. Total projection to latent structures for process monitoring. *AIChE Journal*, 2010, 56(1): 168 – 178.

[17] LIU W, QU P, SUI M, et al. Study of fault diagnosis based on T-S fuzzy neural network for pumping well. *System Simulation Technology*, 2013, 9(2): 141 – 146.

[18] ROSIPAL R, TREJO L J. Kernel partial least squares regression in reproducing kernel Hilbert space. *Journal of Machine Learning Research*, 2001, 2(12): 97 – 123.

[19] PENG K, ZHANG K, LI G. Quality-related process monitoring based on total kernel PLS model and its industrial application. *Mathematical Problems in Engineering*, 2013, 2013(1): 1 – 14.

[20] CHEN C, WANG Y J, ZHANG Y, et al. Indoor positioning algorithm based on nonlinear PLS integrated with RVM. *IEEE Sensors Journal*, 2018, 18(2): 660 – 668.

[21] YI J, HUANG D, HE H, et al. A novel framework for fault diagnosis using kernel partial least squares based on an optimal preference matrix. *IEEE Transactions on Industrial Electronics*, 2017, 64(5): 4315 – 4324.

[22] YI J, HUANG D, FU S, et al. Optimized relative transformation matrix using bacterial foraging algorithm for process fault detection. *IEEE Transactions on Industrial Electronics*, 2016, 63(4): 2595 – 2605.

[23] ZHANG Y, ZHOU H, QIN S J, et al. Decentralized fault diagnosis of large-scale processes using multiblock kernel partial least squares. *IEEE Transactions on Industrial Informatics*, 2010, 6(1): 3 – 10.

[24] JIANG Q, YAN X. Parallel PCA–KPCA for nonlinear process monitoring. *Control Engineering Practice*, 2018, 80(11): 17 – 25.

[25] ZHOU Donghua, HU Yanyan. Fault diagnosis techniques for dynamic systems. *Acta Automatica Sinica*, 2009, 35(6): 748 – 758. (周东华, 胡艳艳. 动态系统的故障诊断技术. 自动化学报, 2009, 35(6): 748 – 758.)

[26] ZHANG Y, HU Z. Multivariate process monitoring and analysis based on multi-scale KPLS. *Chemical Engineering Research and Design*, 2011, 89(12): 2667 – 2678.

[27] BIGLIERI E, YAO K. Some properties of singular value decomposition and their applications to digital signal processing. *Signal Processing*, 1989, 18(3): 277 – 289.

[28] ZHOU W, LI X L, YI J, et al. A novel UKF–RBF method based on adaptive noise factor for fault diagnosis in pumping unit. *IEEE Transactions on Industrial Informatics*, 2019, 15(3): 1415 – 1424.

[29] AN Xing, LIU Gang, ZHANG Liangliang, et al. Sensor fault location method based on cumulative residual contribution rate. *Journal of Civil and Environmental Engineering*, 2019, 41(2): 133 – 139. (安星, 刘纲, 张亮亮, 等. 基于累积残差贡献率的传感器故障定位方法. 土木建筑与环境工程, 2019, 41(2): 133 – 139.)

[30] PENG K X, MA L, ZHANG K. Review of quality-related fault detection and diagnosis techniques for complex industrial processes. *Acta Automatica Sinica*. 2017, 43(3): 349 – 365.

作者简介:

汪 波 硕士研究生, 目前研究方向为复杂工业过程监测与诊断, E-mail: www.wangbonihao@qq.com;

夏钦锋 工程师, 目前研究方向为油气田设备监测与故障诊断, E-mail: xiaqinfeng0617@qq.com;

钱 龙 硕士研究生, 目前研究方向为深度学习、故障诊断, E-mail: 283571087@qq.com;

彭 军 教授, 目前研究方向为人工智能、机器学习, E-mail: pengjun70@126.com;

周 伟 博士, 副教授, 目前研究方向为自适应学习与控制、故障诊断, E-mail: zhouw@cqust.edu.cn.