

doi: 10.7690/bgzdh.2016.08.001

基于自适应核主元分析的航空发动机异常监测

文 莹, 闫雅慧

(空军空降兵学院高射炮兵系, 广西 桂林 541002)

摘要: 针对航空发动机工作参数时变特性导致虚警率高的问题, 采用自适应策略改进核主元分析模型进行异常监测。给出核主元分析异常监测框架, 先用标准化的正常数据离线建模, 后对实时数据计算统计量进行在线监测; 为适应监测数据的时变特性, 采用移动窗口自适应模型改进框架, 利用实时采集的正常数据不断更新模型和监测数据。以航空发动机异常监测仿真实验进行验证。仿真结果表明: 该方法能有效适应工作参数的变化, 具有较高的故障检测率和较低的虚警率。

关键词: 异常监测; 自适应; 核主元分析; 航空发动机

中图分类号: TP206.3 **文献标志码:** A

Aero Engine Anomaly Monitoring Based on Self-adaptive Kernel Principal Component Analysis

Wen Ying, Yan Yahui

(Department of Anti-Aircraft Artillery, Air Force Airborne Academy, Guilin 541002, China)

Abstract: Aiming at the high wrong alarming rate caused by aero-engine work parameter time-varying feature, use self-adaptive strategy to improve kernel principal component analysis model for anomaly monitoring. The framework of anomaly monitoring based on kernel principal component analysis (KPCA) was presented. At first, use standardized normal data to establish off-line model, then carry on on-line monitoring for real time calculation statics. Because of time-varying feature of monitoring data, use moving window self-adaptive model to improve framework, and adopt real time normal data to renew model and monitoring data. Validate it by aero-engine anomaly monitoring simulation. Results of aero-engine simulation indicate that the proposed method can meet requirements of work parameter change, has higher fault detection rate and lower false alarm rate.

Keywords: anomaly monitoring; self-adaptive; KPCA; aero-engine

0 引言

异常监测作为预测与健康管理的关键步骤之一, 其目的是生成一个可靠的早期预警信号, 来区分关键部件当前行为和正常行为之间的关系。现代航空发动机大都配备了数据采集系统, 在运行过程中通过各种传感器采集过程数据, 为采用核主元分析方法进行异常监测提供了数据条件。核主元统计方法适用于处理非线性问题, 不需要深入了解运行机制, 也不需要大量的故障样本数据, 只需要足够的、合适的正常历史运行数据集就可以建立统计模型, 通过比较该模型与实际运行过程数据的统计行为实现异常监测^[1-4]。

航空发动机工作条件复杂, 必然存在多种不同工作模式或工作参数发生正常漂移的情况^[5]。传统的核主元分析方法(kernel principal component analysis, KPCA)在建模监测时, 假设历史运行数据来自工作参数恒定的单一工作模式, 不能根据实时采集的数据更新模型, 虚警率或故障漏报率较高, 难以满足实际工程应用的需要。与传统单一工作模

式下的静态异常监测方法相比, 工作参数缓变条件下的异常监测需要处理更加复杂的数据对象, 对实时性有更高要求, 因而具有重要的研究价值^[6-7]。

针对多模时变条件下的异常监测问题, 目前常用的解决方案是多模型策略^[8-10]和递归策略^[11-12]。多模型策略是对多个工作模式下历史数据分别建立对应的子模型, 监测对应工作模式下的异常行为, 能处理有限离散情况, 但对工作参数时变的情况缺乏处理能力; 递归策略是在原始样本的基础上增加新采集的正常数据作为训练样本, 不断更新原模型, 随着采集样本的增加, 核矩阵的维数与计算量也相应增加, 难以满足实时性要求。笔者采用自适应策略来解决这一问题, 先建立核主元分析异常监测框架, 在不增加计算复杂度的条件下, 采用移动窗口自适应模型改进框架, 利用实时采集的正常数据对监测模型进行更新, 以提高异常监测的准确性。

1 核主元分析异常监测框架

核主元分析方法是一种非线性多元统计分析方法^[13]。假设输入空间中的一组 M 维 n 个样本数据为

收稿日期: 2016-04-21; 修回日期: 2016-06-11

作者简介: 文 莹(1987—), 男, 湖南人, 博士, 从事装备综合保障、故障诊断与预测、指挥自动化与智能化研究。

$x_k(k=1,2,\dots,n)$, 核主元分析首先通过非线性映射 $\Phi: R^M \rightarrow F$, 将输入空间映射到特征空间 $F: \Phi(x_k)$ 中, 然后在特征空间 F 中做线性主元分析。

1.1 选择核函数

通过引入核函数 $k(\bullet)$ 替代非线性映射函数的内积运算:

$$k(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle. \quad (1)$$

可以避免复杂的非线性计算, 并且能比较容易计算得到核矩阵 $\mathbf{K} = [k(x_i, x_j)]$, $i, j = 1, 2, \dots, n$ 。计算核矩阵 \mathbf{K} 的特征值 λ 和特征向量 α :

$$n\lambda\alpha = \mathbf{K}\alpha. \quad (2)$$

并标准化特征向量 $\bar{\alpha}$, 最终可以计算得到特征空间 F 的核主元:

$$t_j = \sum_{i=1}^n \bar{\alpha}_i^j k(x_i, x), j = 1, 2, \dots, n. \quad (3)$$

目前常用的核函数主要有 3 类^[14]: 高斯径向基函数、多项式核函数和神经网络核函数。常选用高斯径向基函数

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right). \quad (4)$$

其中 σ 为需选定的参数。

1.2 确定有效核主元数目

有效核主元数目直接影响到异常监测的效果。通常采用主元贡献率法确定有效核主元数目^[15]。设核矩阵 \mathbf{K} 的特征值由大到小依次为 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$, 则第 j 个核主元贡献率定义为:

$$\text{Contr}(\lambda_j) = \lambda_j / \sum_{i=1}^n \lambda_i. \quad (5)$$

它表明第 j 个核主元所包含的信息占全部信息的百分比。为了降低数据维数的同时保留足够的信息, 在确定有效核主元数目时, 要求 p 个核主元数目的累积贡献率必须大于一个临界值, 即

$$\text{CPV} = \sum_{i=1}^p \lambda_i / \sum_{i=1}^n \lambda_i \times 100\% \geq \text{CL}. \quad (6)$$

其中 CL 为所设定的临界值, 通常取为 85%。

1.3 确定统计指标及其控制限

KPCA 完成主元变换后, 需在主元空间和残差子空间设置 2 个统计量及其控制限, 最常用的统计量指标是霍特林 T^2 统计量(简称 T^2 统计量)和平方预测误差统计量(squared prediction error, SPE), 也称 Q 统计量^[15]。

T^2 统计量反映了输入向量 \mathbf{x} 偏离正常状态的程度, 定义为

$$T^2 = (t_1, t_2, \dots, t_p) \mathbf{A}^{-1} (t_1, t_2, \dots, t_p)^T. \quad (7)$$

其中: t_j 为输入向量 \mathbf{x} 在特征空间中第 j 个核主元; \mathbf{A} 为与前 p 个核主元对应的特征值组成的对角阵。

在过程服从正态分布的前提下, T^2 统计量服从自由度为 $(p, n-p)$ 的 F 分布。当检验水平为 γ 时, T^2 统计量的控制限 T_γ^2 可利用 F 分布按下式计算:

$$T_\gamma^2 = \frac{p(n-1)}{n-p} F_{p, n-p, \gamma}. \quad (8)$$

其中 $F_{p, n-p, \gamma}$ 是对应于置信水平为 γ 、自由度为 p 时, 在 $n-p$ 条件下的 F 分布临界值。若 $T^2 < T_\gamma^2$, 则表明输入向量 \mathbf{x} 的 T^2 统计正常。

SPE 统计量反映的是某一时刻模型和测量值之间的偏离程度。在 KPCA 模型中, 可由下式计算 SPE 统计量:

$$\text{SPE} = \sum_{i=1}^l t_i^2 - \sum_{i=1}^p t_i^2. \quad (9)$$

其中 l 为 n 个特征值中非零特征值的个数。

当置信水平为 γ 时, SPE 统计量的控制限可按如下式计算:

$$\text{SPE}_\gamma = \theta_1 \left[\frac{C_\gamma \sqrt{2\theta_2 h^2}}{\theta_1} + 1 + \frac{\theta_2 h(h-1)}{\theta_1^2} \right]^{\frac{1}{h}}. \quad (10)$$

其中: $\theta_i = \sum_{j=p+1}^n \lambda_j^i, i=1, 2, 3$; $h = 1 - \frac{2\theta_3}{\theta_2}$; C_γ 为标准正态分布检验水平为 γ 的临界值。若 $\text{SPE} < \text{SPE}_\gamma$, 则表明输入向量 \mathbf{x} 的 SPE 统计正常。

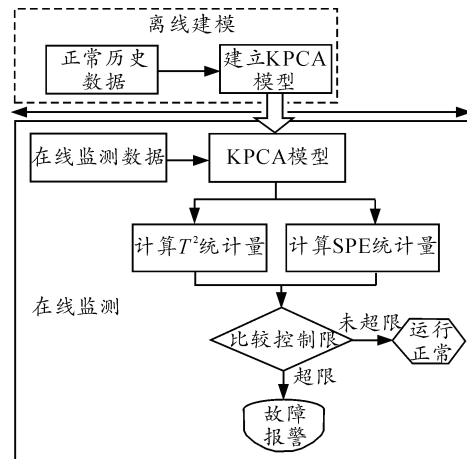


图 1 核主元分析异常监测框架

基于 KPCA 模型的异常监测就是通过判断实时采样数据的 T^2 统计量和 SPE 统计量是否超过正常数据计算得到的控制限来确定是否故障。KPCA 异常监测框架如图 1 所示, 包括离线建模阶段和在线监测阶段。离线建模阶段利用正常历史数据计算核主元及统计量, 并确定统计量的控制限; 在线监测

阶段计算实时采样数据的统计量,并与控制限比较,从而判断正常还是故障。

2 基于移动窗口的自适应 KPCA 模型

静态 KPCA 的缺点在于根据输入数据建立模型后,模型固定不变。当用固定的 KPCA 模型监测时变系统的性能时,必然会引起偏差。为使 KPCA 模型适应被监测系统运行的时变特性,提高异常监测的准确率,需要根据实时采集的正常数据,对模型自动更新。笔者提出一种基于移动窗口的自适应方法,其基本思想:利用新采集的正常数据替换相应数量的旧数据,重新形成正常样本集,使新样本集的样本个数保持不变,然后利用新的正常样本集重新建立 KPCA 模型,更新核主元数目、监测统计量及其控制限,以更新后的 KPCA 模型进行异常监测。

设移动窗口的宽度为 w , 移动步长为 h , 基于移动窗口的自适应 KPCA 异常监测步骤如下:

1) 收集正常历史数据,初始化移动窗口数据 $x_k (k=1,2,\dots,w)$, 并置累积数 $i=0$ 。

2) 对移动窗口数据进行标准化处理,根据标准化数据建立 KPCA 模型,确定核主元数目,计算核主元、统计量及其控制限。

3) 对于一个新采集的运行监测数据,首先进行标准化处理,然后根据当前的 KPCA 模型计算其 T^2 统计量和 SPE 统计量,并与当前统计量的控制限进行比较。如果 T^2 和 SPE 没有超标,则认为新采集的数据正常,并进行累加操作,令 $i=i+1$; 否则认为新采集的数据为异常样本,不进行累加操作。

4) 如果累积数 $i < h$, 则移动窗口数据不变,模型不更新,转到 3), 用原模型继续监测; 如果累积数 $i = h$, 则将新采集的 h 个正常数据加入到移动窗口数据中,为保持窗口宽度不变,去掉 h 个旧数据,得到新的移动窗口数据 $x_k (k=h+1, h+2, \dots, h+w)$, 然后对累积数清零,令 $i=0$, 转到 2) 更新模型。

上述方法中,移动窗口以一定步长向前推进,不断更新 KPCA 模型,以适应被监测系统运行过程中的时变特性。移动窗口宽度及移动步长的选择至关重要。移动窗口宽度太小,则不能形成有效的统计量,从而影响监测结果的准确性;移动窗口太大,核矩阵的维数会很大,导致计算量也很大,影响监测的实时性。移动步长太小,则会由于频繁更新而增加计算量,不利于实时监测;移动步长太大,则对参数漂移不敏感,从而造成监测结果不准确。总之,移动窗口宽度及步长的选择应考虑被监测系统运行的实际情况,通过不断调整以达到比较理想的效果。

3 航空发动机异常监测仿真实验分析

NASA 将“商用模块化航空推进系统仿真”(commercial modular aero-propulsion system simulation, C-MAPSS)用于仿真大型涡轮风扇发动机^[16]。C-MAPSS 具有一定开放性和实用性,许多学者用其模拟发动机在不同工作条件下的飞行数据或故障数据;因此,笔者采用 C-MAPSS 仿真实验来验证文中方法的有效性。

C-MAPSS 在 Matlab 和 Simulink 环境下开发,提供图形界面以便于设置仿真参数。C-MAPSS 中的发动机模型如图 2,主要包括风扇(fan)、低压压缩机(low pressure compressor, LPC)、高压压缩机(high pressure compressor, HPC)、低压涡轮机(low pressure turbine, LPT)和高压涡轮机(high pressure turbine, HPT)5 部分,图中 N1 表示扇轴;N2 表示中心轴;Combustor 表示燃烧室;Nozzle 表示喷管。

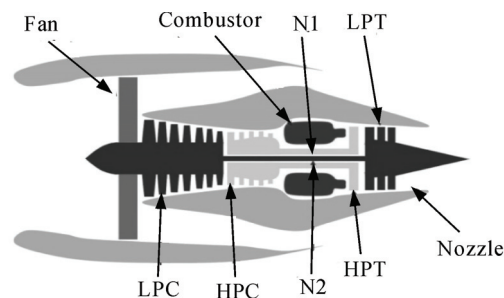


图 2 C-MAPSS 涡轮风扇发动机仿真模型

该发动机仿真模型包括 14 个输入参数和 27 个输出参数。输入参数为燃油流量及 13 个可调节的发动机各部件健康参数,通过设置输入参数可模拟发动机各部件的突发型故障和渐进型故障。输出参数模拟对发动机各个部件的传感器测量,包括温度、压力和速度等,通过记录输出参数对发动机健康状态进行监测。除了发动机模型以外,C-MAPSS 还提供了表征发动机工作条件的大气模型,该模型包括 3 个参数:高度,范围为海平面到 40 000 英尺;马赫数,范围为 0~0.9;海平面温度,范围为 -60~103 ℃。另外,发动机推力可由油门解算器角度(throttle resolver angle, TRA)参数控制,该参数也可作为工作条件参数。根据这些工作条件参数,C-MAPSS 提供了 14 组预先定义好的工作模式(FC01~FC14),供用户在运行发动机仿真模型之前选择,用户也可自定义工作模式。

笔者选择表 1 所示的 6 种工作模式并随机切换,运行 300 次仿真,采集每次稳态飞行时的 27 个输出参数值作为一个样本数据点,生成样本数据集。样本数据集中,前 200 组数据为发动机仿真模型正常运行时采集的数据,从第 201 次飞行开始在风扇模块引入风扇功率参数(fan efficiency modifier)的斜

坡变化，到第 300 次飞行累计下降 25%。

表 1 仿真实验工作模式

工作模式编号	高度/英尺	马赫数	海平面温度/T	TRA/(°)
FC04	1 000	0.25	59	100
FC05	10 000	0.25	59	100
FC06	20 000	0.70	59	100
FC07	25 000	0.62	59	60
FC08	35 000	0.84	59	100
FC09	42 000	0.84	59	100

采用自适应 KPCA 方法进行异常监测，移动窗口宽度 $w=100$ ，移动步长 $h=10$ ，在建立 KPCA 模

型时选用高斯径向基函数作为核函数。

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right)$$

其中 $\sigma=8$ ，核主元累积贡献率临界值 $CL=85\%$ ，统计量控制限的置信度为 99%，监测结果如图 3 所示。作为比较，采用静态 KPCA 方法进行异常监测，以前 200 组正常样本数据用于离线建模，核函数选择及相关参数设置与自适应方法一致，后 100 组数据用于在线监测，监测结果如图 4 所示。

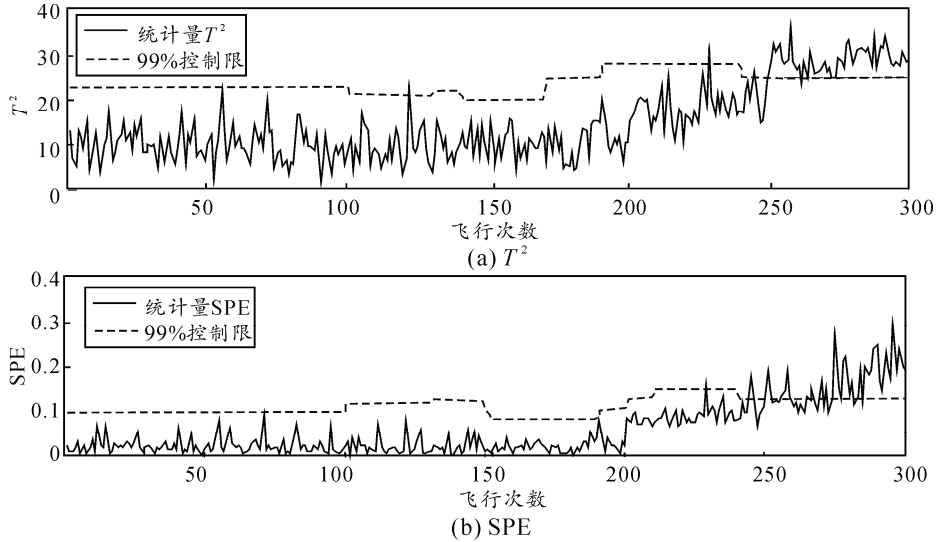


图 3 自适应 KPCA 异常监测结果

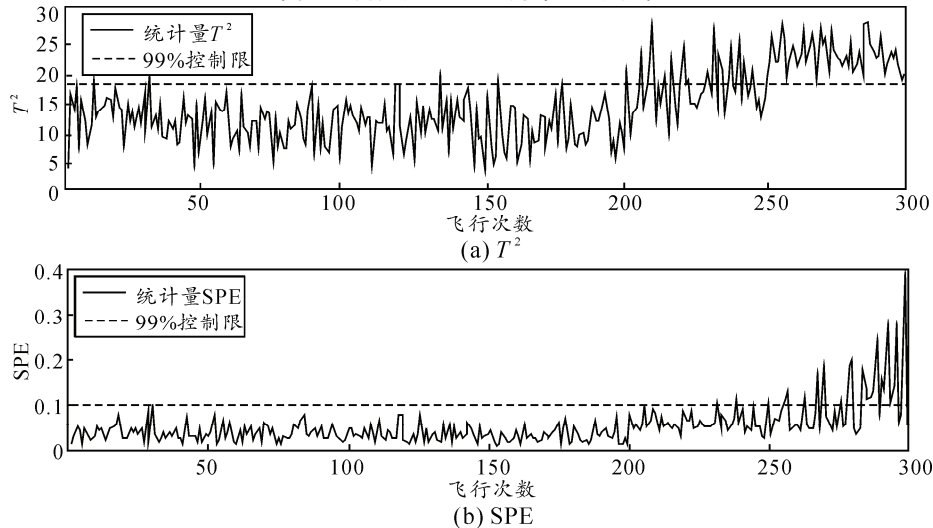


图 4 静态 KPCA 异常监测结果

比较图 3 和图 4 可以看出：静态 KPCA 方法不具备实时更新能力，当工作参数漂移时，无法跟踪正常工作状况，对前期微弱变化反应过于灵敏， T^2 统计量在 200 次飞行之后不断超限，虚警率为 45.5%，过高的虚警率降低了监测结果的可靠性。而自适应 KPCA 方法能够根据实时采集的新数据不断更新模型和控制限，跟踪工作状况的变化， T^2 和 SPE 统计量都能较好地地区分正常样本和异常样本， T^2 统计量的故障检测率为 95%，虚警率为 4%，SPE

统计量的故障检测率为 87%，虚警率为 6%。比较而言，自适应 KPCA 方法能够适应和跟踪工作参数在一定范围内的变化，给出更加准确的监测结果。

4 结论

针对航空发动机工作参数时变特性，笔者采用自适应策略改进核主元分析模型，并用仿真实验验证了该方法的有效性，具有一定的工程应用价值。