

利用增量式非线性流形学习的状态监测方法

张熠卓^{1,2}, 徐光华^{1,2}, 梁霖^{1,2}, 张锋^{1,2}, 李淑智¹

(1. 西安交通大学机械工程学院, 710049, 西安; 2. 西安交通大学机械制造系统工程国家重点实验室, 710049, 西安)

摘要: 针对传统流形学习方法难以处理大批量设备运行数据的特点, 提出了一种采用增量式流形学习方法的机械设备状态监测方法. 该方法首先利用局部切空间排列算法对训练样本集进行非线性维数约简, 得到初始的低维流形结构, 然后通过增量式学习机制对新增的时序样本点进行动态聚类. 通过对压缩机喘振试验数据及滚动轴承故障数据的分析表明, 该方法的计算复杂度低, 可以有效地识别出隐藏在高维特征集中的非线性故障特征, 因此具有良好的工程应用前景.

关键词: 流形学习; 增量式学习; 状态监测; 局部切空间排列

中图分类号: TP206 **文献标志码:** A **文章编号:** 0253-987X(2011)01-0064-05

Condition Monitoring Method for Mechanical Equipments Based on Incremental Nonlinear Manifold Learning

ZHANG Yizhuo^{1,2}, XU Guanghua^{1,2}, LIANG Lin^{1,2}, ZHANG Feng^{1,2}, LI Shuzhi¹

(1. School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China; 2. State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: Most nonlinear manifold learning methods can not be efficiently operated in a 'batch' mode when data are collected sequentially. An incremental learning scheme is proposed for condition monitoring of mechanical equipments. A nonlinear dimensionality reduction algorithm called local tangent space alignment (LTSA) is first utilized to train the low-dimensional manifold structure from the original feature space. To process the long term recordings, an unsupervised learning scheme based on incremental local tangent space alignment (ILTSA) is chosen to cluster the new coming samples dynamically. The proposed method is evaluated by vibration signals measured on defective bearings with different fault types and pressure signals collected from compressor with surge fault, respectively. The results show that the scheme enables to achieve a high accuracy for condition classification with potential in identifying novel patterns from high dimensional feature space.

Keywords: manifold learning; incremental learning; condition monitoring; local tangent space alignment

机械系统在运行过程中一旦发生故障, 其特征往往表现出非线性特性, 如压缩机在运行过程中因失稳而产生的喘振现象^[1]. 由于难以建立真实工况下的数学模型, 因此对反映设备状态的时

间序列进行分析就成为实现故障诊断与状态监测的关键环节. 随着设备结构日益复杂, 监测诊断方法趋多元化, 诊断过程获得信息的维数也同时增大, 因此研究如何从“高维数、非线性”的数据中准

收稿日期: 2010-05-06. 作者简介: 张熠卓(1982—), 男, 博士生; 徐光华(联系人), 男, 教授, 博士生导师. 基金项目: 国家自然科学基金资助项目(51075323); 北京交通大学轨道交通车辆结构可靠性与运用检测技术教育部工程研究中心资助项目(SROM RGV(BJTU)2010-002); 中央高校基本科研业务费专项资金资助项目.

网络出版时间: 2010-10-22

网络出版地址: http://www.cnki.net/kcms/detail/61-1069_t.20101022.1521.002.html

确地提取出设备的运行状态,已经成为众多学者的关注热点^[2].

近年来,非线性流形学习作为一种机器学习的新方法,为有效地从高维数据中提取出机械设备运行状态的特征提供了良好的思路.如阳建宏等人^[3]提出了一种基于主流形识别的非线性齿轮振动信号降维方法;蒋全胜等人^[4]运用拉普拉斯特征映射的非线性降维算法来保留振动信号中内含的整体几何结构信息;梁霖等人^[5]利用局部切空间排列算法提取出了滚动轴承的冲击特征波.这些研究都充分证明了流形识别故障特征方法的有效性.

本文提出了一种基于增量式流形学习的状态监测方法,该方法首先利用局部切空间排列算法对训练样本集进行非线性维数约简,得到初始的低维流形结构,然后利用增量式学习机制更新初始的流形结构,并且对新增的时序样本点进行动态聚类,实现在线状态识别.通过实际案例的应用分析,表明增量式流形学习方法适用于大批量、长周期的设备运行数据,并且可以有效地提取出隐含在高维数据中的设备运行状态.

1 基本理论

由于早期故障的样本数据通常较少,会导致数据的不完全现象,即数据集存在着“空洞”.局部切空间排列(LTSA)算法能够有效地处理“空洞”数据集,并可获得较好的嵌入结果^[6].因此,本文利用LTSA 算法对高维特征空间进行维数约简.对于一个给定的 m 维数据集, $\mathbf{X}=[x_1,x_2,\cdots,x_N]$, $x_i\in\mathbf{R}^m$.为了提取出一个 d 维($d<m$)的嵌入流形,LTSA 主要算法的步骤为:①对于高维空间的每个数据点 x_i ,基于欧氏距离测度函数确定出由 k 个邻近点组成的邻域 $\mathbf{X}_i=[x_{i1},x_{i2},x_{i3},\cdots,x_{ik}]$;②在 x_i 的邻域内选择一组正交基 $\mathbf{Q}_i$ 构成 x_i 的 d 维切空间,计算邻域内每个点 x_{ij} ($j=1,2,\cdots,k$)到切空间的正交投影 $\theta_j^{(i)}=\mathbf{Q}_i^T(x_{ij}-\bar{x}_i)$,其中 $\bar{x}_i$ 为邻域数据的均值, $\mathbf{Q}_i$ 为取 $\mathbf{X}_i(\mathbf{I}-\mathbf{e}\mathbf{e}^T/k)$ 的前 d 个最大的左奇异向量; $\mathbf{I}$ 为单位矩阵, $\mathbf{e}$ 为全 1 向量,由 x_i 的邻域在切

空间的投影所构成的局部坐标向量 $\boldsymbol{\Theta}_i=[\theta_1^{(i)},\theta_2^{(i)},\cdots,\theta_d^{(i)}]$ 描述了 x_i 邻域内的几何结构;③由 $\mathbf{B}=\sum_{i=1}^N\mathbf{S}_i\mathbf{W}_i\mathbf{W}_i^T\mathbf{S}_i^T$ 构建排列矩阵, $\mathbf{W}_i=(\mathbf{I}-\mathbf{e}\mathbf{e}^T/k)(\mathbf{I}-\boldsymbol{\Theta}_i^+\boldsymbol{\Theta}_i)$, $\mathbf{S}_i$ 是满足 $[x_1,\cdots,x_N]\mathbf{S}_i=[x_{i1},\cdots,x_{ik}]$ 的选择矩阵;④由于 $\mathbf{e}$ 是矩阵 $\mathbf{B}$ 的 0 特征值对应的特征向量,所以取 $\mathbf{B}$ 的第 2 到第 $d+1$ 个最小特征值对应的特征向量所组成的矩阵就是 $\mathbf{X}$ 中的正交低维全局坐标映射向量 $\mathbf{T}_i$, $i=1,2,3$.

在 LTSA 算法步骤中,选取邻域的计算复杂度为 $O(mN^2)$,计算局部坐标系统的复杂度为 $O(mk^2N)$,计算低维嵌入的计算复杂度为 $O(dN^2)$.在处理长周期的时间序列数据时,由于依靠传统的 LTSA 算法重复进行批量计算是一项非常耗时的工作,因此无法满足在线监测的需求.

2 采用增量式流形学习机制的状态识别方法

如图 1 所示,本文提出的状态监测方法包括学习和识别阶段.在学习阶段,该方法根据训练数据生成的初始流形结构,来确认已知的状态模式,而在识别阶段,该方法将新的输入数据动态聚类到现有的模式中,从而实现了状态监测.无论是在学习还是识别阶段,相同的特征生成策略都被用来构造信号的高维特征集.

2.1 特征生成策略

特征生成策略的主要目的是为了给特征选择或特征压缩方法提供一个特征集,在机械故障诊断与监测中,特征的生成往往通过输入样本的线性变化来实现,如时域统计分析、傅里叶变换、小波分析等.对于本身具有非线性特性的样本,滑动时间窗的处理方法给予了良好的解决方案.对于机械设备运行中所反映出的信号数据,选择在时频域中较为常见的特征参数来覆盖信号在时域和频域内的特征是比较通用的方法^[7].对原始信号进行滑动时间窗处理,分割成长度相同的数据段,将每段数据作为一个观测样本,分别进行时频统计,构造出高维特征向量.

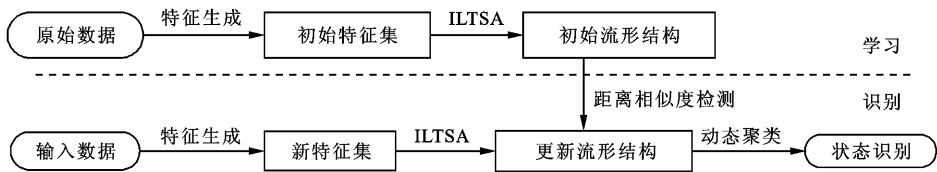


图 1 基于 ILTSA 的状态识别流程图

常用的时域指标包括均值、峰值、均方根、方差、峭度、峰值因子和形状因子等,频域指标采用不同频带内的频带能量.

2.2 基于增量式学习的状态识别方法

流形学习理论假定同状态的数据集属于一个或几个流形结构,不同状态的数据集属于不同的流形结构. 为了给聚类提供专家可判断的依据,用于训练的历史数据必须包括完整的状态数据(正常和故障). 由 LTSA 算法对训练样本集进行维数约简后,在低维嵌入空间中主要包含的信息包括初始低维流形的嵌入坐标,以及训练样本的类别数目. 由于不同状态的低维嵌入曲线在空间中不交叉,因此可以保存不同初始低维流形的中心点坐标,并作为状态辨识的依据. 当输入新增的时序样本点时,采用增量式学习对其进行动态聚类(见图 2),其流程为:①对于训练出的每一类低维嵌入坐标集,利用平均欧氏距离计算其聚类中心点的坐标;②对于每一个新增的样本点,使用 ILTSA 计算其所对应的低维坐标,并同时更新流形结构. ILTSA 算法的步骤如下^[8].

步骤 1 对于新增的样本点 x_{N+1} ,使用 k 近邻规则更新邻域矩阵(如果 x_{N+1} 到 x_i 的欧氏距离小于 x_i 邻域内的某一点到 x_i 的距离,则将 x_{N+1} 替换该点以更新 x_i 的邻域).

步骤 2 对更新后的邻域矩阵进行奇异值分解,得到新的局部几何坐标,并且按照传统的 LTSA 算法来构建新的局部排列矩阵.

步骤 3 利用 Rayleigh-Ritz 子空间迭代加速算法,计算更新后排列矩阵的特征向量,即为更新后的全局坐标映射.

计算更新后的各类中心点坐标与新增低维坐标

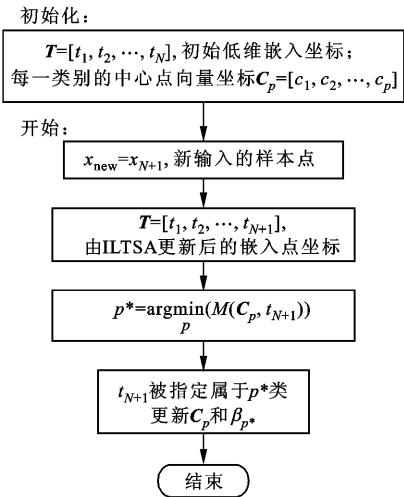


图 2 动态聚类流程图

之间的距离,根据距离相似度准则确定其所属类别. 定义相似度测度函数

$$M(C_p, t_{N+1}) = e^{d^2(C_p, t_{N+1})/\beta_p} \tag{1}$$

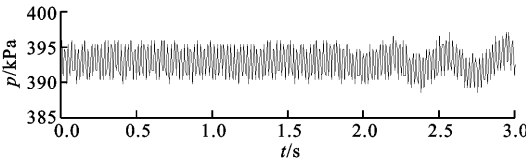
式中: C_p 为不同类的中心点向量坐标,其下标 p 代表训练样本的类别数; t_{N+1} 代表新增点的嵌入坐标; β_p 为不同类内所有样本点的平均欧氏距离.

3 压缩机喘振数据分析

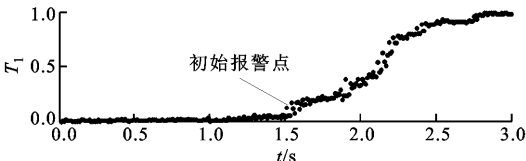
对于某种异常状态较为明确的问题,状态监测的目标主要是区分正常和异常状态,属于单状态识别问题,本文采用压缩机运行过程中出现的喘振故障进行验证. 试验数据采集某厂生产的 250 kW 离心压缩机,采样频率为 150 Hz,采用一组缓慢的节流数据作为训练样本,选取数据长度为 3 s, $k=20$. 图 3a 和 3b 分别为原始的压力信号时域波形和利用传统 LTSA 算法构造出的初始一维主流形曲线. 其中嵌入点坐标经过归一化处理得到

$$T(t_i) = \frac{T(t_i) - \min_i[T(t_i)]}{\max_i[T(t_i)] - \min_i[T(t_i)]} \tag{2}$$

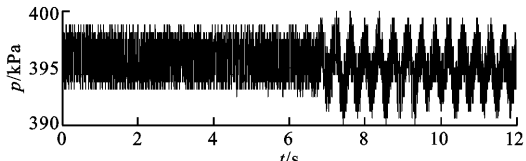
在压力信号出现低频波动前,一维主流形曲线



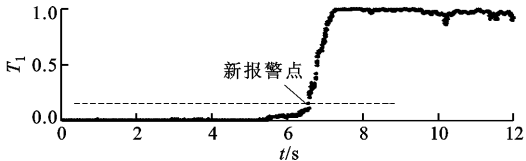
(a) 初始压力信号时域波形



(b) 初始一维主流形曲线



(c) 新增数据的时域示意波形



(d) 更新后的一维主流形曲线

图 3 压力信号时域波形及其对应的流形曲线

显示出明显的上升趋势,此时压缩机运行已由正常状态向喘振阶段过渡,由初始流形曲线设置喘振报警阈值为 0.2. 通过测试数据选择同次试验的快速节流信号,为了便于比较,图 3c 中只显示了快速节流在喘振发作前期 12 s 的时域信号,利用增量式学习方法计算得到更新后的一维主流形曲线如图 3d 所示. 从图 3 中可以看出,更新后的流形曲线在喘振发作前期与初始流形保持着相同的变化趋势.

随着样本点数的不断增多,批量式 LTSA 算法的耗时量逐渐增加,图 4 是在主频为 2.8 GHz、内存为 2 GB 容量的计算机上计算出的 LTSA 与 ILTSA 耗时量的对比. 当样本点数增加到 700 时,LTSA 一次批量计算所需要的时间已达 7.58 s,而 ILTSA 只需要 0.62 s. 由此可知,增量式算法的计算耗时量远低于批量式计算方法,在实际应用中,可以通过调整初始流形的报警点来控制报警级别,进行喘振在线监测.

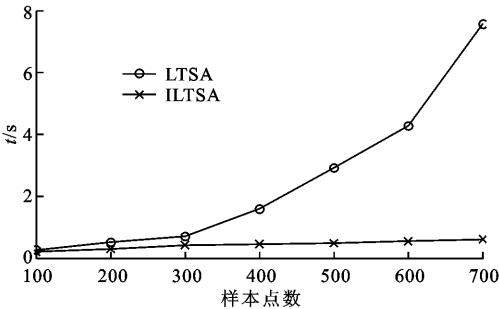


图 4 2 种算法计算时间的对比

4 滚动轴承的数据分析

对于多状态识别问题,仅依靠一维主流形曲线的识别是不够的,根据参考文献[9],有效嵌入维数应为已知的类别数减 1,因此多状态识别需要在二维或三维嵌入空间中实现.

本文采用滚动轴承数据来验证多状态识别方法,数据集来自美国 Cast Western Reserve University 电气工程试验室^[10],试验所用轴承为 SKF 公司生产的深沟球轴承. 选取驱动端的轴承数据作为试验分析对象(见表 1),其中 D07、D14、D21 分别代表不同故障程度的数据集.

表 1 滚动轴承试验数据类型

数据集	故障类型	故障损伤尺寸/mm
D07	内圈,滚动体,外圈	0.18
D14	内圈,滚动体,外圈	0.36
D21	内圈,滚动体,外圈	0.53

利用本文 2.1 节介绍的方法构造出高维特征集,对于每一类样本集,选择 116 个样本点作为训练样本集,选择 232 个样本点作为测试样本集. 图 5 为分别利用主分量分析(PCA)、等度规映射(ISO-MAP)和 LTSA 方法对 D07 数据集处理后得到的初始三维嵌入结果.

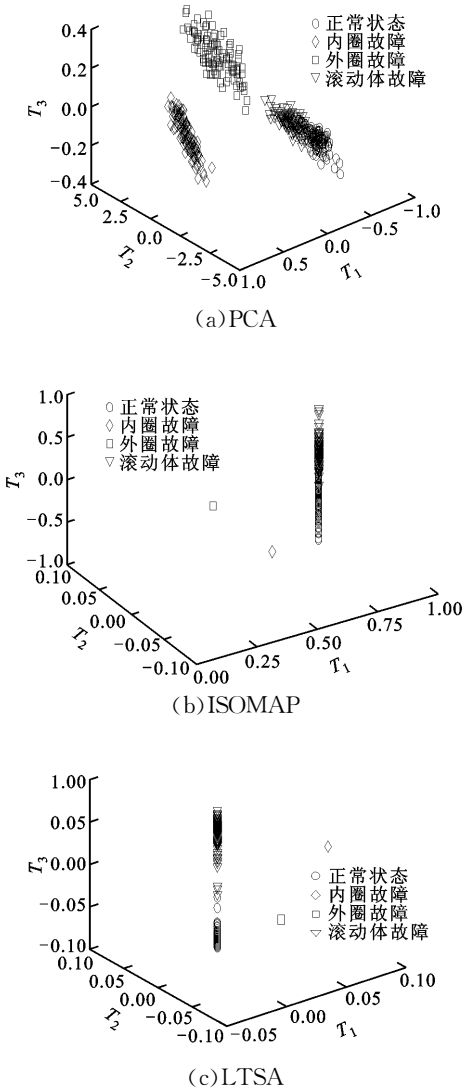


图 5 不同算法计算的初始流形结构

图 6 为利用增量式算法对 D14 数据集处理后所得到的三维嵌入坐标在一个方向上的投影. 图 6a 为初始结果,图 6b、6c 为增量式计算结果. 流形学习方法使得尽可能多的同类数据在低维空间中形成相同的拓扑结构(曲线或曲面),可以看出随着样本点数的增加,不同类别的低维嵌入结果所形成的曲线结构是相似的,并且曲线无交叉. 利用相似度准则函数对 ILTSA 结果进行动态聚类,表 2 中记录了各种不同数据集的平均识别精度. 为了对比其他线性和非线性增量式方法,增量式主分量投影方法(IPCA)

和增量式 ISOMAP(IISOMAP)方法^[11]的计算结果也同样被列出.可以看出,相比 IPCA 和 IISOMAP 方法,ILTSA 获得低维嵌入具有较高的识别精度.

5 结 论

本文提出了一种基于增量式非线性流形学习的机械系统状态监测方法.利用局部切空间排列算法对训练样本集进行非线性维数约简,得到初始的低维流形结构,然后通过增量式学习机制对新增的时序样本点进行动态聚类.增量式流形学习方法克服了传统流形学习计算复杂度高、无法处理大批量监测数据的缺点,能够准确地提取出不同类型的状态特征,因此具有较强的工程实用价值.

参考文献:

[1] 孙涛,徐光华,张春梅. 基于喘振频域特性的失稳预警技术研究 [J]. 西安交通大学学报,2007,41(11):1321-1325.
SUN Tao, XU Guanghua, ZHANG Chunmei. Instability early warning based on frequency-domain characteristic of surge [J]. Journal of Xi'an Jiaotong University,2007,41(11):1321-1325.

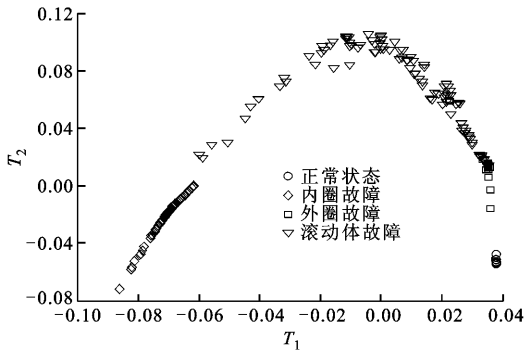
[2] 黎敏,徐金梧,阳建宏,等. 一种基于流形拓扑结构的轴承故障分类方法 [J]. 控制工程,2009,16(3):358-362.
LI Min, XU Jinwu, YANG Jianhong, et al. Classification method of bearing faults based on topological structure of manifold [J]. Control Engineering of China,2009,16(3):358-362.

[3] 阳建宏,徐金梧,杨德斌,等. 基于主流形识别的非线性时间序列降噪方法及其在故障诊断中的应用 [J]. 机械工程学报,2006,42(8):154-158.
YANG Jianhong, XU Jinwu, YANG Debin, et al. Noise reduction method for nonlinear time series based on principal manifold learning and its application to fault diagnosis [J]. Chinese Journal of Mechanical Engineering, 2006,42(8):154-158.

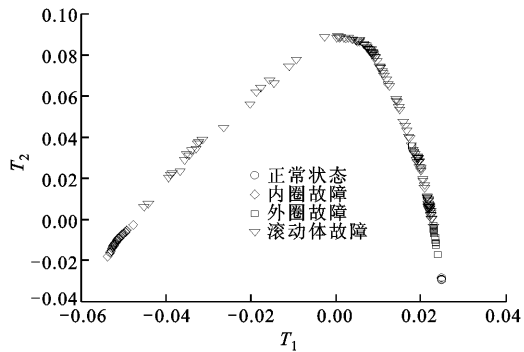
[4] 蒋全胜,贾民平,胡建中,等. 基于拉普拉斯特征映射的故障模式识别方法 [J]. 系统仿真学报,2008,20(20):5710-5713.
JIANG Quansheng, JIA Minping, HU Jianzhong, et al. Method of fault pattern recognition based on Laplacian eigenmaps [J]. Journal of System Simulation, 2008,20(20):5710-5713.

[5] 梁霖,徐光华,栗茂林,等. 基于非线性流形学习的冲击故障特征提取方法 [J]. 西安交通大学报,2009, 43(11):95-99.
LIANG Lin, XU Guanghua, LI Maolin, et al. Nonlinear manifold learning method of mechanical impact faults extraction [J]. Journal of Xi'an Jiaotong University, 2009,43(11):95-99.

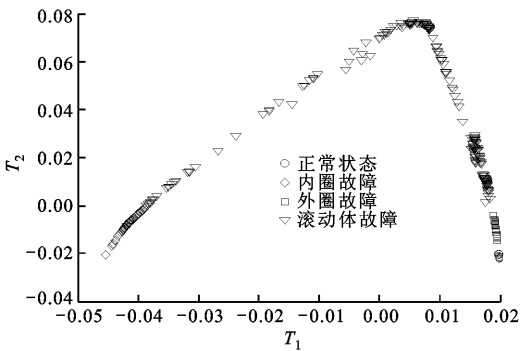
[6] ZHANG Zhenyue, ZHA Hongyuan. Principal mani-



(a)116 个样本点



(b)232 个样本点



(c)348 个样本点

图6 采用ILTSA计算的不同样本点数下的三维嵌入投影结果

表2 试验数据的识别精度统计

数据集	识别精度/%		
	IPCA	IISOMAP	ILTSA
D07	75	89	100
D14	75	75	85
D21	80	85	95

folds and nonlinear dimension reduction via local tangent space alignment [J]. SIAM Journal on Scientific Computing, 2004, 26(1):313-338.

[7] HE Qingbo, KONG Fanrang, YAN Ruqiang. Subspace-based gearbox condition monitoring by kernel principal component analysis [J]. Mechanical Systems and Signal Processing, 2007, 21(4):1755-1772.

[8] LIU Xiaoming, YIN Jianwei, FENG Zhilin, et al. Incremental manifold learning via tangent space alignment [J]. Artificial Neural Networks in Pattern Recognition, 2006, 4087:107-121.

[9] KOUROPTOVA O. Unsupervised learning with locally linear embedding algorithm: an experimental study [D]. Finland: University of Joensuu. Department of Computer Science and Statistics, 2001.

[10] LOPARO K A. Bearings vibration data set, case western reserve university [DB/OL]. [2008-12-05]. <http://www.eecs.cwru.edu/laboratory/bearing/download.html>.

[11] LAW H C, JAIN K. Incremental nonlinear dimensionality reduction by manifold learning [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(3):377-391.

(编辑 管咏梅)