

基于 EMD-PCA-LSSVM 方法的滚动轴承安全域估计和状态辨识^①张 媛^{②*} 秦 勇* 邢宗义** 贾利民* 廖贵玲*

(* 北京交通大学交通运输学院 北京 100044)

(** 南京理工大学自动化学院 南京 210014)

摘 要 将安全域的思想引入滚动轴承的状态监测中,综合利用经验模式分解(EMD)、主成分分析(PCA)和最小二乘支持向量机(LSSVM),进行了滚动轴承运行状态的安全域估计以及正常和各种故障状态的辨识。首先,按一定的时间间隔将采集的振动数据分段,每段数据进行 EMD 后获得各本征模函数(IMF)分量;其次,基于各段数据的本征模函数分量,利用主成分分析方法提取出每段数据的 T^2 统计量和平方预估误差(SPE)统计量控制限值作为滚动轴承的状态特征量;最后,利用二分类的 LSSVM 进行滚动轴承运行状态的安全域估计,利用多分类的 LSSVM 进行滚动轴承的正常以及滚动体故障、内圈故障、外圈故障四种状态的辨识。试验结果显示安全域估计准确率和多种状态辨识正确率均大于 95%,验证了上述方法的有效性。

关键词 滚动轴承, 状态监测, 安全域, 经验模式分解(EMD), 主成分分析(PCA), 最小二乘支持向量机(LSSVM)

0 引 言

在轨道车辆、汽车、工程机械行业中,滚动轴承应用十分广泛,但其故障率较高,据统计,仅有 10%~20% 可达到设计寿命^[1],因此需对滚动轴承运行状态进行准确有效的监测和识别,以确保设备运行安全。特征的提取和状态的辨识是滚动轴承状态监测中需解决的关键问题,国内外学者已对此进行了深入研究。在特征提取方面,经验模式分解(empirical mode decomposition, EMD)是一种较新的信号处理方法,具有自适应和高信噪比的特点,十分适合于机械振动信号等非平稳、非线性信号的分析处理^[2],且通过 EMD 分解出的本征模函数(intrinsic mode function, IMF)分量可用于提取故障特征信息。目前,已有学者采用计算各 IMF 的能量矩^[1]、能量熵^[3]、Renyi 熵^[4]、Shannon 熵^[5]以及计算 IMF 矩阵奇异值^[5]等方法进行了故障特征提取的研究。但上述研究均未考虑振动信号的统计特征,而振动信号的各种统计量往往包含了丰富的对象运行状态

信息,可检测出其运行状态的变化。主成分分析(principal component analysis, PCA)是一种典型的多元统计模式识别方法,参数限制少,计算简单^[6]。迄今,基于 PCA 的多元统计性能监控方法在质量控制、过程监测和故障诊断等领域已有诸多应用^[7,8]。在状态监测中,通过 PCA 对当前状态的采样数据进行最佳综合简化后,可求得不同子空间的统计变量值及其控制限,这些统计变量及其控制限可以表征当前状态的统计特性。在状态辨识方面,最小二乘支持向量机(least square support vector machine, LSSVM)是一种基于经典支持向量机(SVM)的扩展和改进的智能分类方法,其不仅具有经典 SVM 的小样本学习能力强、泛化性能好、有效避免局部极小的特点^[9],而且运算速度快,无需指定逼近精度^[10],因此 LSSVM 在模式识别、预测控制等领域有广泛应用^[11,12]。针对滚动轴承状态监测问题,本文从域的角度出发,提出了基于 EMD-PCA-LSSVM 方法估计滚动轴承运行状态安全域以及辨识正常和各种故障状态的方案,并通过试验验证了其有效性。

① 863 计划(2011AA110501),国家科技支撑计划(2011BAG01B05)和国家重点实验室自主课题(RCS2010ZZ002)资助项目。

② 女,1985 年生,博士生;研究方向:轨道车辆故障诊断与状态评价;联系人,E-mail: zhangyuan111@gmail.com

(收稿日期:2012-07-04)

1 安全域理论

安全域分析及估计理论最早由 Wu^[13] 等提出, 先应用于电力系统安全状态评价, 之后扩展到网络控制、公路交通、电子政务等方面。近年来, 本文作者所在团队将安全域基本思想引入轨道交通系统及关键设备的运行安全状态监测及评价中^[14], 提出了基于安全域估计的机械设备服役状态的安全评估方法。

从滚动轴承状态监测和识别研究的角度看, 安全域是一个在从域的角度全面描述滚动轴承可正常(无故障)运行区域的定量模型。直观地说, 在状态特征变量所确定的空间中划定一个区域, 估计此区域的边界, 当滚动轴承的状态特征变量值所确定的状态点落在此区域内时, 则认定此时滚动轴承运行状态正常(无故障), 当状态点落在此区域外时, 则认为此时滚动轴承运行状态非正常(故障)。当状态特征变量数为 2, 即所确定的空间为二维平面空间时, 滚动轴承的运行状态安全域如图 1 所示, 其中 v_1 、 v_2 为两个状态特征变量, P_1 、 P_2 分别表示滚动轴承运行状态正常和故障时的两个状态点。安全域估计的主要工作是获得安全域边界, 即一个能够区分正常和故障两种状态的分类决策函数。

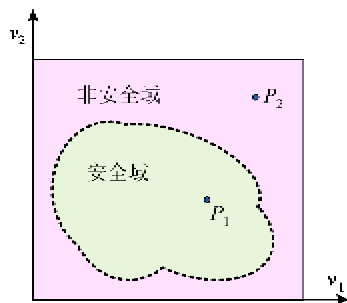


图1 安全域示意图

2 EMD-PCA-LSSVM 方法

2.1 EMD 基本原理

经验模式分解(EMD)是由 Huang 等^[15]提出的一种信号时频分析方法, 该方法利用局部特征时间尺度, 从原始信号中提取出若干个本征模函数(intrinsic mode function, IMF)分量和一个残余量。IMF 分量突出了数据的局部特征, 残余分量体现了信号中缓慢变化量, 对其进行分析, 可以准确有效地把握原信号的特征信息^[16]。

每个 IMF 分量必须满足以下两个条件: 过零点的数量与极值点的数量相等或至多相差一个; 在任一时间点, 局部最大值确定的上包络线和局部最小值确定的下包络线的均值为零, 即信号关于时间轴局部对称。EMD 可通过“筛选”过程来获得各 IMF 分量, 详细计算步骤见文献[17]。

通过“筛选”过程, 原始信号最终可分解为若干个平稳的 IMF 分量和一个残余量的线性和。将各 IMF 分量的频率成分从大到小排列, 可表明各个 IMF 分量被分解到不同频段, 有利于信号特征的提取。

2.2 基于 PCA 的统计特征提取

作为一种成熟的过程监测与故障诊断方法, 主成分分析(PCA)的基本思想是根据状态或过程变量的历史数据, 利用多元投影的方法将多变量样本空间分解成由主元变量张成的较低维的特征子空间和一个相应的残差子空间, 并分别在这两个子空间中构造能够反映空间变化的统计量, 然后将观测向量分别向两个子空间进行投影, 并计算相应的统计指标用于状态和过程的监控^[18]。特征空间的统计量可以反映状态的主要变化, 常用的有 T^2 统计量等; 残差空间的统计量可以反映影响状态变化的干扰和噪声, 常用的有平方预估误差(squared prediction error, SPE)统计量、Hawkins 统计量等^[19]。

本文基于正常和故障状态下的滚动轴承振动数据的各 IMF 分量, 选取最常用的 T^2 和 SPE 两统计量, 分别计算不同状态下的两统计量值及其控制限, 并提取其控制限的值作为滚动轴承的状态特征量用于安全域估计和状态辨识。PCA 基本算法以及 T^2 和 SPE 统计量的控制限的详细计算见文献[19]和文献[20]。在计算过程中, 取常用的置信度水平 $\alpha = 0.95$ 。

2.3 基于 LSSVM 的状态辨识

支持向量机(SVM)的基本思想是通过定义适当的核函数实现非线性变换, 将输入空间变换到一个高维空间, 然后在此新空间中求取最优线性分类面^[21]。最小二乘支持向量机(LSSVM)在 SVM 的基础上进行了扩展和改进, 将 SVM 中的不等式约束替代为等式约束, 且将误差平方和损失函数作为训练集的经验损失, 从而将 SVM 中的二次规划问题转化为线性方程组求解, 在保证精度的前提下大幅降低了计算复杂性^[22]。关于 LSSVM 的详细描述可参考文献[23]。鉴于高斯径向基核在诸多工程实际问题中表现优异, 故本文选其作为核函数。经试凑, 径向基函数宽度取为 0.5。

针对多分类问题,需要组合多个二分类 SVM 来构造多分类 SVM 分类器,常用的方法有一对一 SVM、一对多 SVM、决策导向无环图 SVM (directed acyclic graph SVM, DAGSVM) 等,其中一对一 SVM 和 DAGSVM 更适用于实际应用,但一对一 SVM 存在误分和拒分现象^[24],故本文采用 DAGSVM 多分类方法进行滚动轴承的正常及滚动体故障、内圈故障、外圈故障 4 种状态的辨识。

对于一个有 M 类的数据样本分类问题, DAGSVM 需要构造每两类间的分类面,即 $M(M-1)/2$ 个完成二分类的子分类器,并将所有子分类器构成一个两向有向无环图,包括 $M(M-1)/2$ 个节点和 M 个叶。其中每个节点为一个子分类器,并与下一层的两个节点(或叶)相连。当对一个未知样本进行分类时,首先从顶部的根节点(包含两类)开始,据根节点的分类结果用下一层的左节点或右节点继续分类,直到达到底层某个叶为止,该叶所表示类别即为未知样本的类别^[25]。本文用于滚动轴承状态辨识的 DAGSVM 如图 2 所示。

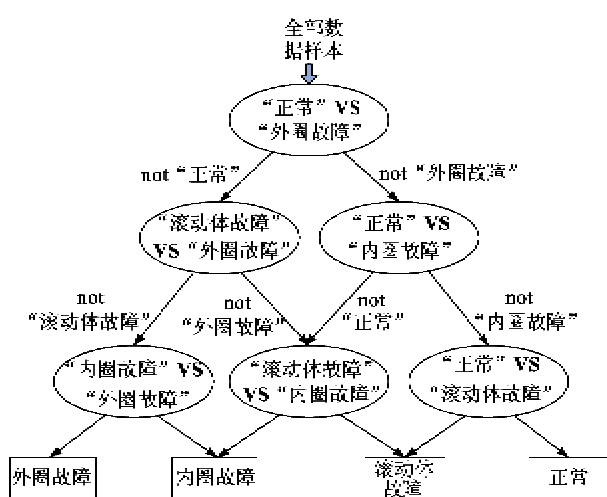


图 2 DAGSVM 多分类方法示意图

2.4 EMD-PCA-LSSVM 方法实施步骤

基于 EMD-PCA-LSSVM 方法的滚动轴承安全域估计方法的实施步骤如下：

步骤 1: 分别采集滚动轴承的运行状态处于正常和故障(包括滚动体故障、内圈故障和外圈故障)两种情况下的振动加速度数据。

步骤 2: 确定合适的时间间隔, 将所采集的各状态下数据分别划分为若干段(如采集时间为 T_s , 时间间隔为 t_i , 则所采集数据分为 T_s/t_i 段), 每段数据将对应一个特征向量。

步骤 3: 对每段数据分别进行 EMD 处理, 获得每段数据的 IMF 矩阵。

步骤 4: 对每个 IMF 矩阵进行 PCA 处理, 计算其 T^2 和 SPE 统计量的控制限, 两控制限的值构成每段数据的二维统计特征向量, 每个向量对应一个状态点。

步骤 5: 将获得的 T^2 和 SPE 统计量控制限数据分别进行归一化处理, 按照运行状态的不同分别标记为“正常”和“故障”。

步骤 6: 利用 LSSVM 将此二维统计特征量(状态点)进行分类, 得到最佳分类线, 此分类线即为滚动轴承运行状态的安全域边界。

进一步地, 辨识滚动轴承的正常、滚动体故障、内圈故障、外圈故障 4 种状态时, 分别采集此 4 种状态下的振动加速度数据, 重复步骤 2-5, 然后, 训练每两类间的 LSSVM 子分类器, 利用 DAGSVM 多分类方法即可得到状态辨识结果。

图 3 直观地给出了上述实施过程, 图中 $Q = T_s/t_i$, 为某种状态下数据段的个数。

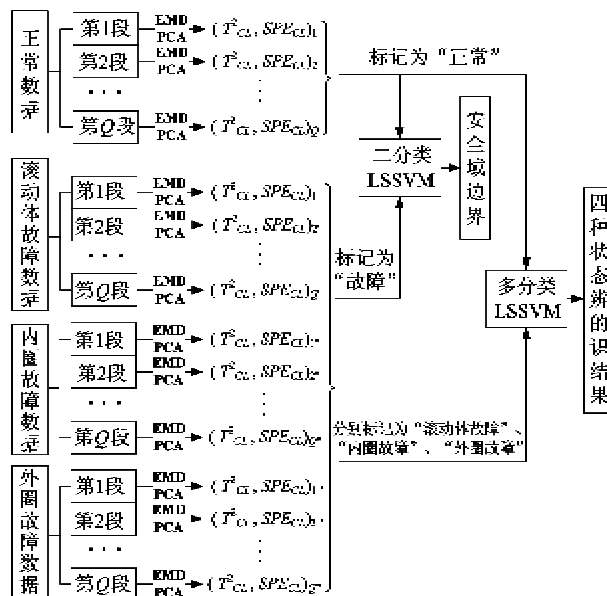


图 3 EMD-PCA-LSSVM 方法的实施过程

3 试验与结果

本文采用 Dr. Loparo^[26] 提供的滚动轴承实验数据, 轴承型号为 205-2RS JEM SKF 型深沟球轴承, 电机负载 3 马力, 转速 1730r/min, 滚动体及内、外圈的故障直径均 0.1778mm, 深度为 0.2794mm, 故障较轻, 采样时间 10s。为检验方法优劣, 利用以下两种情况数据进行了试验：

情况 1:采样频率 12kHz,驱动端数据;
情况 2:采样频率 48kHz,风扇端数据。

本文按滚动轴承的转速确定用于划分数据段的时间间隔,即按轴承每转一转所采集的数据点划为一个数据段。则情况 1 和情况 2 下的各状态的数据均划分为 288 段,其中情况 1 的每段数据包含 426 个数据点,情况 2 的每段数据包含个 1706 数据点。

此外,所有 LSSVM 训练时均将输入数据按 6 :4 的比例分为训练和测试两部分。

3.1 评价指标

为衡量本文提出基于 EMD-PCA-LSSVM 的安全域估计和状态辨识方法的性能,采用了如下所示的检出率 (detection rate, DR)、误报率 (false alarm rate, FAR) 和分类正确率 (classification rate, CR) 三个评价指标^[27]:

$$DR = \frac{\text{某类中被检出为该类的样本个数}}{\text{实际的该类的样本个数}} \quad (1)$$

$$FAR = \frac{\text{非该类样本中被检出为该类的样本个数}}{\text{实际的非该类的样本个数}} \quad (2)$$

$$CR = \frac{\text{所有被正确分类的样本个数}}{\text{样本总个数}} \quad (3)$$

其中, DR 和 FAR 是针对某一类数据样本的指标, DR 越接近于 1 且 FAR 越接近于 0 时表示这一类样本的识别准确性越高; CR 是针对所有数据样本的指标, CR 越接近于 1 样本的识别准确性越高。

同时,为更加全面地衡量状态辨识效果,还采用了 Fleiss Kappa 统计量指标(简记为 FK 值, $FK \in [-1,1]$)定量评价状态辨识目标和分类器输出间的一致性^[28,29]。Fleiss Kappa 统计量是属性值测量系统中对定类数据进行一致性分析的常用指标,具体计算方法见文献[30],当其值大于 0.8 时可认为两组分类数据几乎完全吻合。

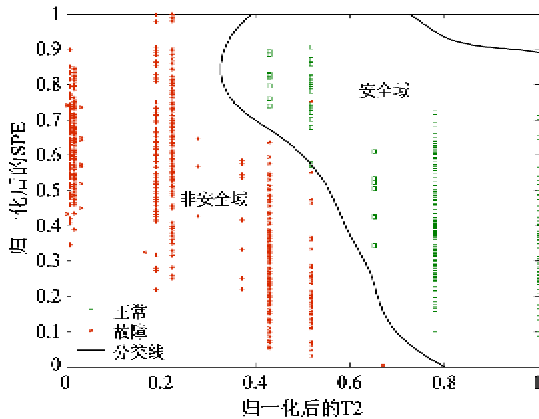
3.2 结果分析

表 1 所示为正常和故障两种状态时的测试数据分类结果,可知情况 1 和情况 2 的测试 CR 分别为 0.9979和0.9957,十分接近于 1;情况 1 和情况 2 的测试 FK 值分别为0.9943和0.9884,均大于 0.8 且接近于 1。图 4 给出了滚动轴承运行状态的安全域估计结果,其中图 4(a)和 4(b)分别为情况 1 和情况 2 的结果,可见表征正常和故障状态的安全域和非安全域之间分界线明显,分类性能优越。即,所获安全域边界的分类精度高,其输出结果与目标结果一致

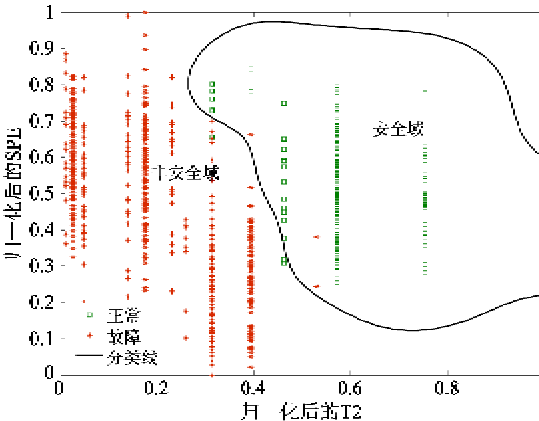
性很高,验证了本文提出安全域估计方法的有效性。

表 1 正常和故障两种状态的数据分类结果

	情况 1	情况 2
测试样本 CR	0.9979	0.9957
测试样本 FK 值	0.9943	0.9884



(a) 情况 1



(b) 情况 2

图 4 滚动轴承运行状态的安全域估计结果

表 2 给出了对正常、滚动体故障、内圈故障、外圈故障 4 种状态进行辨识时,用于多分类 LSSVM 的每两类间共 6 个子分类器的分类性能。可见,无论

表 2 多分类 LSSVM 的每两类间子分类器的 CR

	CR	
	情况 1	情况 2
正常 VS 外圈故障	1.0000	1.0000
滚动体故障 VS 外圈故障	0.9871	0.9828
内圈故障 VS 外圈故障	1.0000	1.0000
正常 VS 滚动体故障	1.0000	0.9914
正常 VS 内圈故障	1.0000	0.9914
滚动体故障 VS 内圈故障	0.9485	0.9267

是情况 1 还是情况 2, 6 个子分类器的 CR 值均大于 0.9, 且情况 1 中 6 个子分类器中的 4 个其 CR 值都达到了 1, 情况 2 中有 2 个子分类器的 CR 值达到 1, 2 个子分类器的 CR 值超过 0.99。图 5 和图 6 分别

给出了情况 1 和情况 2 的 6 个子分类器的分类效果, 可见各子分类器的效果均较理想。即, 各子分类器分类准确性高, 分类效果良好, 能够有效保障多分类 LSSVM 的分类性能。

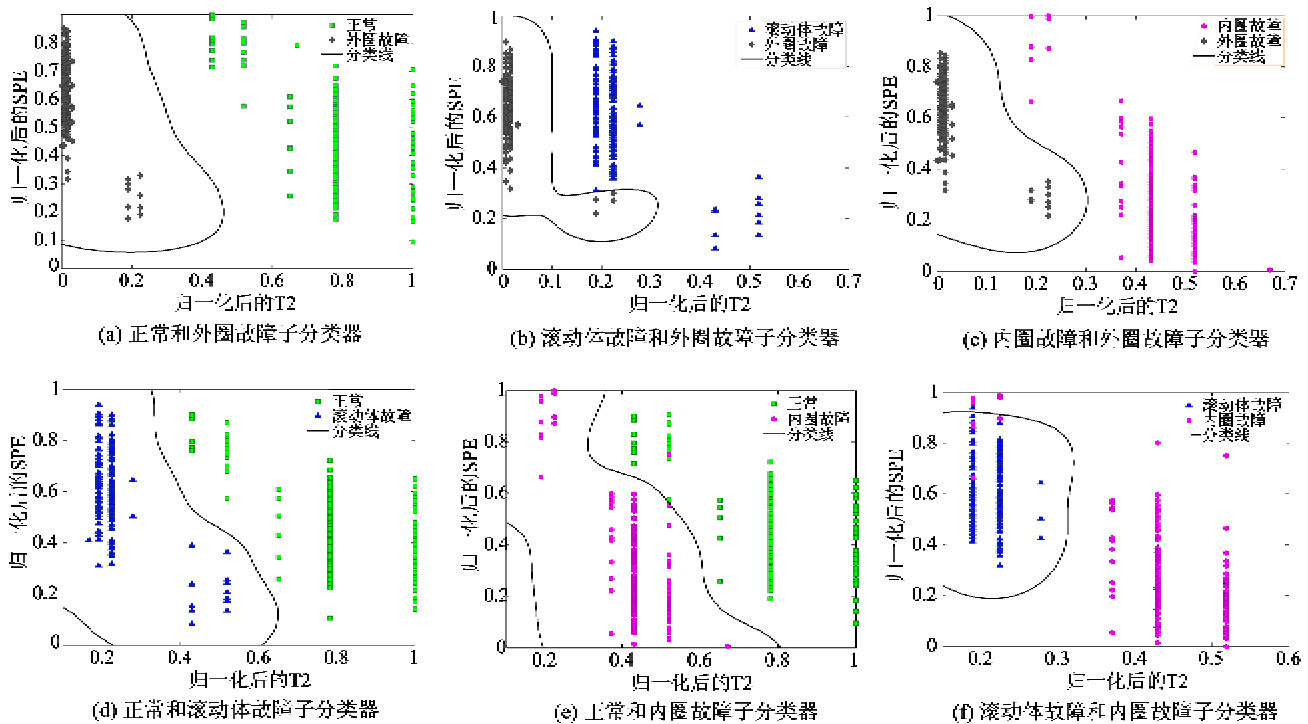


图 5 情况 1 的多分类 LSSVM 中各子分类器的分类效果

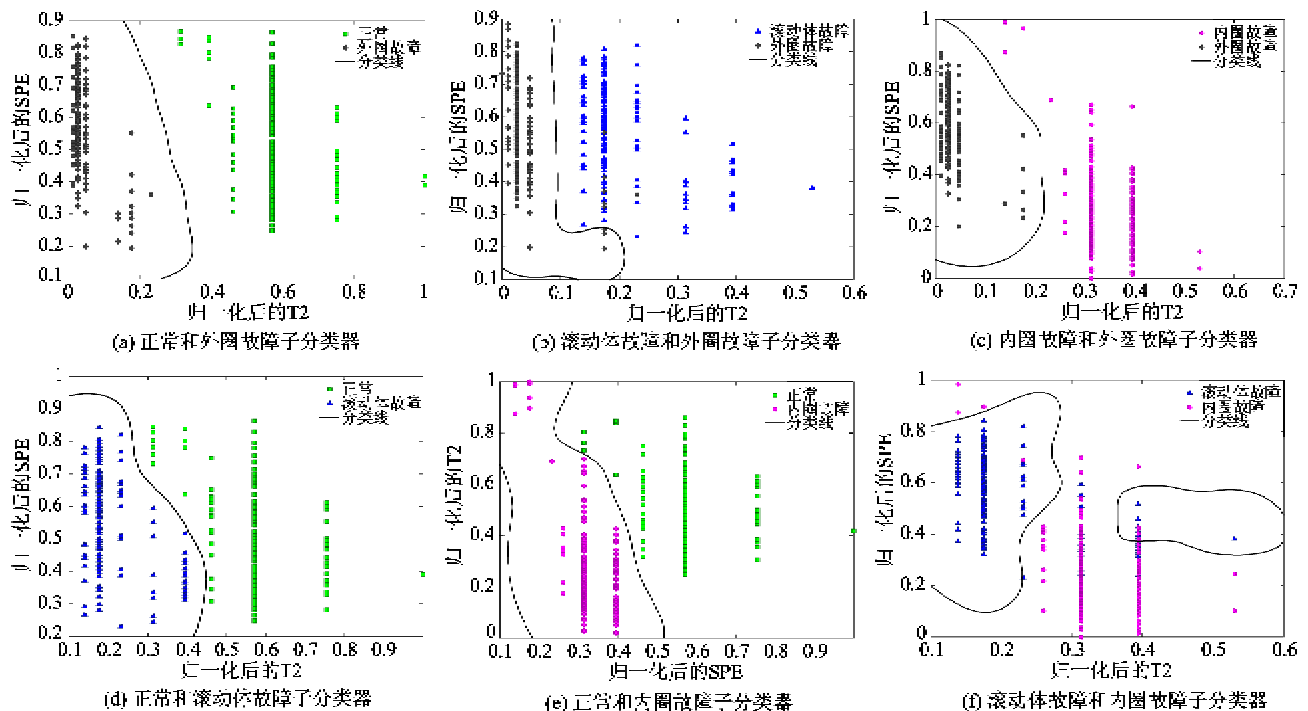


图 6 情况 2 的多分类 LSSVM 中各子分类器的分类效果

表 3 给出了滚动轴承的正常、滚动体故障、内圈故障、外圈故障 4 种状态的辨识结果。可见, 情况 1

和情况 2 的 4 种状态的 DR 均大于 0.9, FAR 值均小于 0.03, 全部样本的整体分类 CR 值分别为 0.9692 和 0.9553, 整体辨识准确率高于 95%, 且两种情况的 FK 值分别为 0.9589 和 0.9404, 十分接近于 1, 整体辨识结果与目标结果的吻合度很高。图 7 直观地表示出 4 种状态的辨识结果, 其中图 7(a) 和图 7(b) 分别为情况 1 和情况 2 的结果。图中横坐标为样本点数, 其中第 1 - 288 个为正常状态样本, 第 289 -

576 个为滚动体故障样本, 第 577 - 864 个为内圈故障样本, 第 864 - 1152 个为外圈故障样本, 如图中用虚线划分的 4 个区域, 纵坐标由下向上分别表示 4 种轴承的状态。由图可见, 4 种状态的样本数据绝大部分能够被正确分类到相应类别下, 误判的样本个数较少。此结果表明本文提出的 EMD-PCA-LSSVM 的方法用于滚动轴承多状态辨识是有效的和可行的。

表 3 正常、滚动体故障、内圈故障、外圈故障四种状态的分类结果

	情况 1				情况 2			
	正常	滚动体故障	内圈故障	外圈故障	正常	滚动体故障	内圈故障	外圈故障
DR	1.0000	0.9241	0.9626	0.9898	0.9966	0.9007	0.9517	0.9725
FAR	0.0023	0.0137	0.0252	0.0000	0.0034	0.0230	0.0298	0.0034
CR	0.9692				0.9553			
FK 值	0.9589				0.9404			

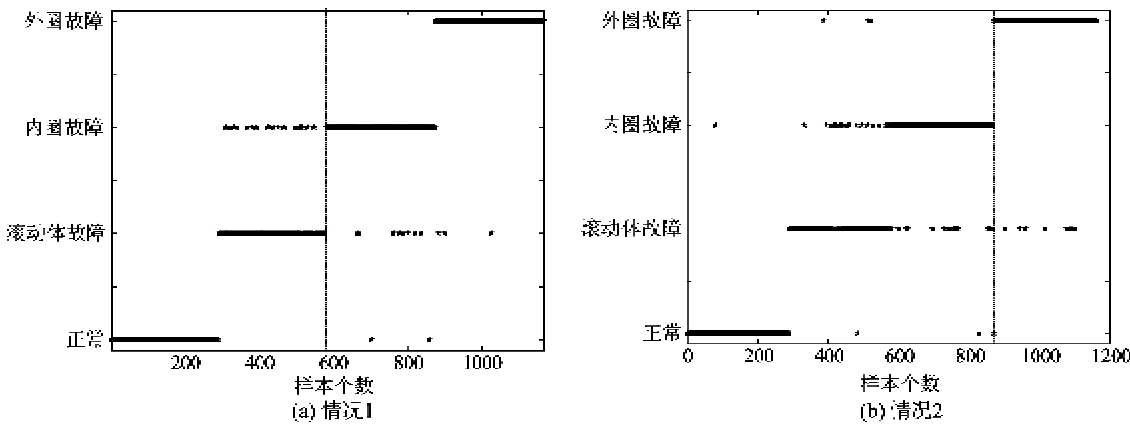


图 7 滚动轴承四种状态的辨识结果

此外,对于情况 1,由图 7(a)可见,滚动体故障的样本被辨识到内圈故障类别下的样本个数最多,导致表 3 中情况 1 的滚动体故障 DR 最低(为 0.9241)而内圈故障 FAR 最高(为 0.0252)。表 2 中情况 1 的滚动体故障与内圈故障子分类器 DR 在 6 个子分类器中最低(为 0.9485),以及图 5(f)的子分类器效果相对较差,可见这是由于用于滚动体故障和内圈故障二分类的子分类器性能相对较低造成的。而对于情况 2,由图 7(b)可见,滚动体故障的样本被辨识到其他三个类别中的样本个数最多,因此表 3 中情况 2 的滚动体故障 DR 最低(为 0.9007),同时其他类别的样本被误判为滚动体故障和内圈故障的样本个数较多,因此滚动体故障和内圈故障的 FAR 相对其他两类较大(分别为 0.0230 和 0.0298)。由表 2 中情况 2 的 6 个子分类器的分

类精度和图 6 中各子分类器的分类效果可见,这是由于滚动体故障与内圈故障的子分类器的 DR 最低(为 0.9267),且分类效果亦不十分理想,即两种情况下的滚动体故障和内圈故障的辨识精度有待提高。

综上所述,基于 EMD-PCA-LSSVM 的滚动轴承运行状态安全域估计以及滚动轴承正常、滚动体故障、内圈故障、外圈故障 4 种状态辨识的试验结果均验证了 EMD-PCA-LSSVM 方法的有效性和可行性。

4 结 论

本研究提出的 EMD-PCA-LSSVM 方法利用 EMD 和 PCA 相结合来取状态特征,将振动信号分段并进行 EMD 处理获得各个 IMF 分量组成的 IMF

矩阵,利用 PCA 计算 IMF 矩阵的 T^2 和 SPE 统计量及其控制限,将此控制限作为滚动轴承状态特征,然后利用 LSSVM 求取安全域边界,完成滚动轴承正常和故障两种运行状态的识别,进而利用多分类 LSSVM 进行滚动轴承的正常及滚动体故障、内圈故障、外圈故障 4 种运行状态的辨识。本研究的创新之处在于:(1)将安全域的概念和估计方法引入滚动轴承的状态监测研究中;(2)将 EMD、PCA 和 LSSVM 三种方法相结合,将 EMD 分解出的各 IMF 分量进行 PCA 处理,提出采用 T^2 和 SPE 统计量控制限作为滚动轴承运行状态的特征量,并基于所提取的特征量,利用二分类 LSSVM 估计滚动轴承运行状态的安全域边界,并利用多分类 LSSVM 进行正常及多种故障状态的辨识。试验结果验证了本文提出方法的有效性,该方法为滚动轴承的状态监测提供了可行途径。

后继研究工作主要集中于滚动体故障和内圈故障的辨识精度的提高,以及基于本文提出方法的滚动轴承监测系统的初步实现。

参考文献

- [1] 秦太龙,杨勇,程珩等. 基于 IMF 能量矩和神经网络的轴承故障诊断. 振动、测试与诊断, 2008,28(3): 229-232
- [2] Rai V K, Mohanty A R. Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert-Huang transform. *Mechanical Systems and Signal Processing*, 2007, 21(6): 2607-2615
- [3] Yang Y, Yu D J, Cheng J S. A rolling fault diagnosis method based on EMD energy entropy and ANN. *Journal of Sound and Vibration*, 2006, 294:269-277
- [4] 王太勇,何慧龙,王国锋等. 基于经验模式分解和最小二乘支持矢量机的滚动轴承故障诊断. 机械工程学报, 2007, 43(1): 88-92
- [5] 李兵,张培林,任国全等. 运用 EMD 和 GA-SVM 的齿轮故障特征提取与选择. 振动、测试与诊断, 2009,29(4): 445-448
- [6] Jackson J E. A User's Guide to Principal Components. New York: Wiley, 1991. 12-20
- [7] Sharmin R, Shah S L, Sundararaj U. A PCA based fault detection scheme for an industrial high pressure polyethylene reactor. *Macromolecular Reaction Engineering*, 2008, 2(1): 12-30
- [8] Moura E P, Souto C R, Silva A A. Evaluation of principal component analysis and neural network performance for bearing fault diagnosis from vibration signal processed by RS and DF analyses. *Mechanical Systems and Signal Processing*, 2011, 25: 1765-1772
- [9] Gopalakrishnan K, Kim S. Support vector machines approach to HMA stiffness prediction. *Journal of Engineering Mechanics*, 2011, 137(2): 138-146
- [10] Gencoglu M T, Uyar M. Prediction of flashover voltage of insulators using least squares support vector machines. *Expert Systems with Applications*, 2009, 36: 10789-10798
- [11] Liu L M, Wang A N, Sha M, et al. Multi-class classification methods of cost-conscious LS-SVM for fault diagnosis of blast furnace. *Journal of Iron and Steel Research, International*, 2011, 18(10): 17-23, 33
- [12] Li X, Cao G Y, Zhu X J. Modeling and control of PEM-FC based on least squares support vector machines. *Energy Conversion and Management*, 2006, 47(7-8): 1032-1050
- [13] Wu F, Kumagai S. Steady-state security regions of power systems. *IEEE Transactions on Circuits and Systems*, 1982, 29(11): 703-711
- [14] Zhang Y, Qin Y, Jia L M. Research on methodology of security region estimation of railway system operation safety assessment. In: Proceedings of World Congress on Engineering and Technology, Shanghai, China 2011. 803-807
- [15] Huang N E, Shen Z, Long S R. The Empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 1998, 454(1971): 903-995
- [16] Xu Y L, Chen J. Characterizing nonstationary wind speed using empirical mode decomposition. *Journal of Structural Engineering*, 2004, 130(6): 912-920
- [17] Tse P W, Peng Y H, Yam R. Wavelet analysis and its envelope detection for rolling element bearing fault diagnosis their affectivities and flexibilities. *Journal of Vibration and Acoustic*, 2001, 123: 303-310
- [18] 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
- [19] Qin S J. Statistical process monitoring: basics and beyond. *Journal of Chemometrics*, 2003, 17(8-9): 480-502
- [20] Richard A J, Dean W W. Applied Multivariate Statistical Analysis (Sixth Edition). New Jersey, USA: Prentice Hall, 2007. 256-378
- [21] Vapnik V N. The Nature of Statistical Learning Theory. New York: Springer-Verlag, 1999. 121-243
- [22] Suykens J A K. Nonlinear modeling and support vector machines. In: Proceedings of the IEEE Instrumentation

- and Measurement Technology Conference, Budapest, Hungary, 2001. 287-294
- [23] He S T, Song T X, Rong C, et al. Online weighted LS-SVM for hysteretic structural system identification. *Engineering Structures*, 2006, 28(12): 1728-1735
- [24] Chih W H, Chih J L. A comparison of methods for multi-class support vector machines. *IEEE Transactions on Neural Networks*, 2002, 13(2): 415-425
- [25] Platt J C, Cristianini N, Shawe-Taylor J. Large margin DAG's for multiclass classification. *Advances in Neural Information Processing Systems*, 2000, 12: 547-553
- [26] The Case Western Reserve University Bearing Data Center Website. Bearing data center seeded fault test data. <http://csegroups.case.edu/bearingdatacenter/pages/download-data-file>, 2012
- [27] 陶新民, 徐晶, 杨立标等. 基于 GARCH 模型 MSVM 的轴承故障诊断方法. *振动与冲击*, 2010, 29(5): 11-15
- [28] Boumahdi M, Dron J P, Rechak S, et al. On the extraction of rules in the identification of bearing defects in rotating machinery using decision tree. *Expert Systems with Applications*, 2010, 37: 5887-5894
- [29] Kankar P K, Sharma S C, Harsha S P. Rolling element bearing fault diagnosis using autocorrelation and continuous wavelet transform. *Journal of Vibration and Control*, 2011, 17(14): 2081-2094
- [30] Sim J, Wright C C. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical Therapy*. 2005, 95(5): 257-268

Safety region estimation and state identification of rolling bearings based on LMD-PCA-LSSVM method

Zhang Yuan*, Qin Yong*, Xing Zongyi**, Jia Limin*, Liao Guiling*

(* School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044)

(** Department of Automation, Nanjing University of Science and Technology, Nanjing 210094)

Abstract

The idea of safety region was introduced into the condition monitoring of rolling bearings, and the research on estimation of a rolling bearing's safe operating region and identification of a rolling bearing's operating state (normal or at fault) was performed by combinative use of empirical mode decomposition (EMD), principal component analysis (PCA) and least square support vector machine (LSSVM). Firstly, the vibration data of a rolling bearing was collected and it was segmented at regular intervals, and intrinsic mode functions (IMFs) of each segment's data were obtained by using EMD. Then two statistical variables' control limits as the state characteristics of the rolling bearing were calculated based on PCA. At last, the boundary of the safety region was estimated by a two-classification LSSVM, and the normal condition and three fault conditions were identified by a multi-classification LSSVM. The experimental results indicated that the accuracy of safety region estimation and that of state identification were both more than 95%, so the effectiveness of the above method was verified.

Key words: rolling bearings, condition monitoring, safety region, empirical mode decomposition (EMD), principal component analysis (PCA), least square support vector machine (LSSVM)