

旋转机械全息序列相似性匹配故障诊断方法^{*}

吴 薇^{1,2}, 胡静涛¹

(1 中国科学院沈阳自动化研究所工业信息重点实验室 沈阳 110016)

2 中国科学院研究生院 北京 100039)

摘 要: 针对全息诊断分辨率低影响旋转机械故障诊断质量和自动化水平的问题, 将时间序列相似性匹配的基本概念和方法引入故障诊断应用中, 结合全息诊断信息融合分析旋转机械振动全貌的思想, 定义了全息序列及其相似性度量模型, 用类时间轴上的多维序列表征转子系统振动全貌, 进而利用采用近似三角不等式与 B+ 树结合剪枝策略的全息序列相似性匹配算法实现故障诊断。实验结果表明, 该方法能够实现高质量的故障自动分类识别。

关键词: 全息序列; 故障诊断; 多维时间序列; 相似性匹配

中图分类号: TP18 TP206 文献标识码: A 国家标准学科分类代码: 520.2099

Holosphere series similarity matching for rotating machinery fault diagnosis

Wu Wei^{1,2}, Hu Jingtao¹

(1. Key Laboratory of Industrial Informatics, Shenyang Inst. of Automation, Chinese Academy of Sciences, Shenyang 110016, China)

2. Graduate School of the Chinese Academy of Sciences, Beijing 100039, China)

Abstract Low resolution limits the quality and automation level of holosphere technique in machinery diagnosis. To remedy this program, the basic definitions and methods of time series similarity matching are applied to the fault diagnosis and combined with the idea of information fusion analysis for the rotating machinery of holosphere technique. The holosphere series and its similarity measurement model are defined. Pseudo multidimensional time series is used to express the rotation panorama of the rotator system, and the holosphere series similarity matching algorithm based on the searching strategy of the combination of the weaker triangle inequality and the B+ searching tree is used to achieve the fault diagnosis for the rotating machine. Experimental results show that the method proposed above can effectively achieve high quality automatic fault identification and classification.

Key words Holosphere series; fault diagnosis; multidimensional time series; similarity matching

1 引 言

全息诊断是一种综合利用大型旋转机组一个或多个支承载面上两个相互垂直方向上振动信号频率、幅值和相位信息的故障诊断分析方法, 它能够精确直观地揭示转子振动各自由度间的内在联系, 有效克服了传统谱分析方法信息利用孤立的片面性和局限性, 具有很强的故障识别能

力。自 1988 年提出以来, 随着在我国的石化、电力、冶金等行业推广应用, 全息诊断解决了许多生产上的难题, 经受了实践的检验^[1]。然而, 目前的全息诊断仍主要以经验性的人工识别方式实现, 费时费力, 当振动状态复杂, 出现不同频率分量振动椭圆交叠时, 会使全息谱识别分辨率降低, 严重影响了故障诊断质量和自动化水平。

时间序列相似性匹配, 指在时间序列数据库中发现与给定时间序列模式相似的序列, 是一种从时间序列数

据中抽取模式的有效方法,是当前机器学习、统计分析和数据挖掘领域中的重点研究课题,已被成功用于具有相似价格起伏的股票、相似销售量变化的商品发现、笔迹识别以及考古化石物种形状识别分类^[2-4]等。本文将时间序列相似性匹配的概念和方法引入到故障诊断领域中,将时间序列相似性匹配的信息检索和知识自动获取能力与全息诊断信息融合分析旋转机械振动全貌的思想相结合,定义了全息序列及其相似性度量模型,用类多维时间序列表征转子振动状态,利用基于近似三角不等式和 B⁺ 树索引结构的全息序列相似性算法判定故障模式,以提高故障诊断质量和自动化水平。

2 全息序列及其相似性度量模型

全息诊断是一种融合机组振动频域信息的方法,它利用改进的傅里叶变换算法将转子支承截面内相互垂直的两个方向的振动位移信号分解,精确求得幅值、频率和相位信息并集成,用合成的一系列椭圆来描述不同频率分量下转子的振动行为。根据描述振动情况的角度不同,全息诊断主要包括侧重反映转子某个单一支承截面全面振动信息的二维全息谱、反映转子系统不同空间位置支承截面整体振动情况的三维全息谱和反映转子单一截面振动随转速变化情况全息瀑布谱等分析技术。其中二维全息谱为最基本和常用的全息谱,如图 1(a)所示。

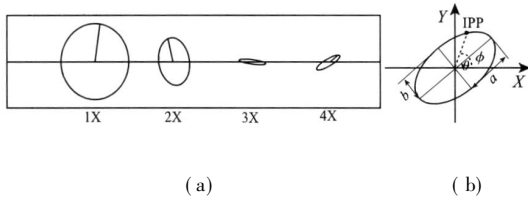


图 1 二维全息谱及阶次椭圆

Fig 1 2D-holospectrum and the frequency phase ellipse

全息谱图的最基本图形特征元素是特定转速下,由同支支承截面相互垂直的 X、Y 两个方向振动位移信号同阶次(频率 ω)简谐波合成的轨迹,轨迹的坐标表示为:

$$\begin{cases} x = A \times \sin(\omega t + \alpha) \\ y = B \times \sin(\omega t + \beta) \end{cases} \quad (1)$$

式中: A 和 B 分别为 X 和 Y 振动位移信号经傅里叶变换分解后 ω 频率正弦分量的幅值, α 和 β 为初始相位。由式(1)可得同频合成振动的轨迹方程:

$$\frac{x^2}{A^2} - \frac{2xy}{AB} \cos(\Delta\varphi) + \frac{y^2}{B^2} - \sin^2(\Delta\varphi) = 0 \quad (2)$$

式中: $\Delta\varphi = \beta - \alpha$ 可见,当 $\cos(\Delta\varphi) \neq 0$, 即 $\Delta\varphi \neq k\pi$ (k 通常取 0 或 1) 时,该方程为椭圆曲线的一般方程形式;当 $\Delta\varphi = k\pi$ 时椭圆退化为直线。

旋转机械的大多数常见故障,都有与之对应明显不

同的全息谱特征^[1]。这些特征主要集中于不同阶次(频率)椭圆(直线和正圆看作椭圆的特殊形式)的大小和旋转角度等形状特征以及初相点方位和进动方向等轨迹运动特征。

目前,全息诊断故障识别还主要依赖于人工经验分析,这样很大程度上制约了诊断的分辨率和质量。本文将全息谱图映射到类时间序列空间,将全息诊断的复杂知识形式转化为便于智能化诊断驱动的信息形式来表征旋转机械的振动全貌。

时间序列是由一组在时间轴上顺序排列的数据点组成,数据点可以是单一数值,也可以是数据向量。这里,将时间轴扩展,给出了类时间序列的概念。

定义 1 一个类时间序列 S 是由一组随某一标识改变的数据点组成,记为 $S = \{s(t) \mid t = 0, 1, 2, \dots, n\}$ 。其中,数据点 $s(t) = (a_1, a_2, \dots, a_m)$ 为 m 元向量; t 为 S 的顺序标识,是具有连贯顺序特性的数据轴上的点; $n = |S|$ 为 S 的长度, m 为 S 的维数。

简单地,可以直接利用轨迹坐标序列将椭圆映射到类时间序列空间,但这样会带来超大规模数据,严重制约识别效率,对于故障诊断是没有意义的。依据椭圆曲线的数学性质以及全息诊断的实践经验,本文定义了全息椭圆特征向量,进而给出了全息序列及其相似性的定义和全息序列相似性的度量模型。

定义 2 全息椭圆特征向量。全息谱中 f 阶次椭圆用于旋转机械故障判定的特征用向量 $v_f = (da, db, \theta, \phi)$ 可以确定。其中,如图 1(b)所示, a 为椭圆的长半轴, b 为短半轴, θ 为长轴倾角, ϕ 为初相点的方位角, d 为符号变量,当椭圆为正进动时为 1, 反之为 -1, 这里称向量 v_f 为全息椭圆特征向量。

全息椭圆特征向量的各元素计算方法如下:

1) 长半轴和短半轴。利用椭圆曲线的正交不变量,椭圆的长半轴 a 和短半轴 b 分别通过式(2)的系数计算整理得:

$$\begin{cases} a = \sqrt{\frac{1}{2}(A^2 + B^2 + \sqrt{(A^2 + B^2)^2 - 4A^2B^2 \sin^2(\Delta\varphi)})} \\ b = \sqrt{\frac{1}{2}(A^2 + B^2 - \sqrt{(A^2 + B^2)^2 - 4A^2B^2 \sin^2(\Delta\varphi)})} \end{cases} \quad (3)$$

2) 进动方向。阶次椭圆是中心点为原点的图形。在椭圆轨迹上按采样顺序连续取三点,分别利用坐标求得各点与 x 轴的夹角,若夹角逐渐增大,则为正进动,反之为反进动^[5]。

3) 长轴倾角。将椭圆轨迹进行转轴处理计算

$$\text{得: } \theta = \frac{1}{2} \arctan\left(\frac{2AB \cos(\Delta\varphi)}{A^2 - B^2}\right) \quad (4)$$

当 $A = B$ 时, $\theta = \pm \frac{\pi}{4}$ 。

4)初相点方位角。由(1)得椭圆轨迹的初相点(IPP)的坐标为:

$$\begin{cases} x_0 = A \sin(\alpha) \\ y_0 = B \sin(\beta) \end{cases} \quad (5)$$

初相点方位的变化与转子的重心位置相关。这里的方位是相对的,与具体的坐标值无关,所以选取初相点的方位角 ϕ 来表示:

$$\phi = \arctan\left(\frac{y_0}{x_0}\right) = \arctan\left(\frac{B \times \sin(\beta)}{A \times \sin(\alpha)}\right) \quad (6)$$

根据 x_0 和 y_0 的符号判断 ϕ 所处的象限, $\phi \in [0, 2\pi)$ 。

可见,全息椭圆特征向量可以用 X 、 Y 两方向的振动位移信号傅里叶变换分解后各阶次频率的幅值和相位即可确定。由于泄漏的存在,为了保证精度,需要对傅里叶变换分解直接得到的幅值和相位值进行校正,这里采用基于汉宁窗的比例内插法^[1]。

定义3 全息序列。

依据频率大小将旋转机械转子同一支承截面同一转速下的全息椭圆特征向量在频率轴上顺序排列,抽象得到序列, $V = \{v_{f_1}, v_{f_2}, \dots, v_{f_n}\}$, f_i 为频率标识, k 为观测阶次频率数。按转速或支承截面空间位置,生成序列 $P = \{p(t'_i) \mid i = 0, 1, 2, \dots, n\}$, 其中, t'_i 为转速或空间位置标识,序列元素 $p(t'_i) = V_{t'_i} = \{v_{t'_i, f_1}, v_{t'_i, f_2}, \dots, v_{t'_i, f_n}\}$, n 为序列 P 的长度,称 P 为全息序列。显然,序列 V 可看作长度为 1 的单位全息序列。

定义4 全息序列相似性度量模型。

设 v_{f_1} 和 v_{f_2} 为任意两个全息椭圆特征向量,两向量的相似性定义为:

$$\sigma(v_{f_1}, v_{f_2}) = \begin{cases} 1 & \text{若 } \Delta v_{f_i} \text{ 各元素小于 } v_e \text{ 中对应元素} \\ 0 & \text{其他.} \end{cases} \quad (7)$$

式中: Δv_{f_i} 表示 v_{f_1} 和 v_{f_2} 对应元素差的绝对值组成的向量, $v_e = (a_e, b_e, \theta_e, \phi_e)$ 为特征向量元素误差阈值组成的阈值向量。

设 $p(t'_1)$ 和 $p(t'_2)$ 为任意两个具有相同频率标识的全息序列元素,二者相似性定义为:

$$\text{simE}(p(t'_1), p(t'_2)) = \frac{1}{k} \sum_{i=1}^k \sigma(v_{t'_1, f_i}, v_{t'_2, f_i}) \quad (8)$$

给定长度分别为 m 和 n , 对应元素频率标识相同的两个全息序列 P 和 Q 的相似性 $\text{Similarity}(P, Q)$ 为:

$$\text{Similarity}(P, Q) = \begin{cases} \text{simE}(P, Q), & \text{若 } m = n = 1 \\ \text{LCSS}_\delta(P, Q) / \min(m, n), & \text{其它.} \end{cases} \quad (9)$$

$$\text{LCSS}_\delta(P, Q) =$$

$$\begin{cases} 0 & \text{若 } P \text{ 或 } Q \text{ 长度为 } 0 \\ 1 + \text{LCSS}_\delta(\text{Head}(P), \text{Head}(Q)), & \\ & \text{若 } \text{simE}(p(t'_m), q(t'_n)) > \delta \\ \max(\text{LCSS}_\delta(\text{Head}(P), Q), \text{LCSS}_\delta(P, \text{Head}(Q))), & \\ & \text{其它.} \end{cases} \quad (10)$$

式中: δ 为全息序列元素相似性阈值, $\text{Head}(P)$ 为全息序列 P 去掉末尾元素后得到的头序列, l 为 P 和 Q 对应元素的向量长度。

全息序列相似性函数可能存在:

$$\text{Similarity}(P, Q) > \text{Similarity}(P, R) + \text{Similarity}(R, Q)$$

所以,该函数不符合距离度量的三角不等式,属于非度量距离(Non-metric Distance)。根据全息序列的特点,可以得到一个近似相似性三角不等式:

性质1 给定全息序列 A, B, C , 则有:

$$\text{Similarity}(A, B) \leq \text{Similarity}(B, C) + (|A| - \text{Similarity}(A, C)) \quad (11)$$

式中: $|A|$ 为序列 A 的长度。

证明: 因为 B 和 C 的最长公共子序列 LCSS_{BC} 包含 A , B 和 C 三者的最长公共子序列 LCSS_{ABC} , 所以 LCSS_{ABC} 减去 A 中既不属于 LCSS_{AB} 也不属于 LCSS_{AC} 的元素数小于等于 LCSS_{BC} 的长度, 即:

$$|A| - (|A| - \text{Similarity}(A, B)) - (|A| - \text{Similarity}(A, C)) \leq \text{Similarity}(B, C) \quad (12)$$

整理得

$$\text{Similarity}(A, B) \leq$$

$$\text{Similarity}(B, C) + (|A| - \text{Similarity}(A, C)) \quad (13)$$

由于工业现场环境复杂,信号测量条件不同,及噪声的干扰,旋转机械同种故障实际测得的全息谱会存在一定程度的不同变形,所以全息序列的相似性度量模型需要对序列变形有一定的鲁棒性。 $\text{Similarity}(A, B)$ 借鉴了字符串识别领域求最大公共子序列的思想,尽量降低异常数据参与相似性计算的几率,利用动态规划方法计算两全息序列一定阈值范围内最长形状相似子序列的长度,表示二者的相似度。该模型支持序列横纵向伸缩和数据点不连续,并且对噪声不敏感。

3 全息序列相似性匹配故障诊断算法

故障诊断实质上是模式分类问题。全息序列相似性匹配故障诊断的基本思想为:将转子历史故障案例的相关振动数据转化为全息序列生成故障全息序列案例库,利用基于全息序列相似性度量模型的 k - m 分类器通过投票机制对旋转机械待测运行状态进行故障类别判定。

全息序列相似性匹配的时间消耗主要集中于全息序列间相似性 $\text{Similarity}(P, Q)$ 的计算代价。随着全息序

列故障案例库的不断扩充, 匹配过程中总的 $Similarity(P, Q)$ 的计算次数决定了匹配的效率。所以, 顺序查找已不实用。为此, 本文设计了全息序列剪枝搜索策略。

1) 剪枝条件 (Pruning Condition)

由性质 1 得, 对于任意一个全息序列 R , 故障全息序列案例库中的序列 S_i 以及设备待测状态的全息序列 Q 存在:

$$\begin{aligned} Similarity(S_i, Q) &\leq \\ Similarity(Q, R) + (|S_i| - Similarity(S_i, R)) \end{aligned} \quad (14)$$

选取某一全息序列作为公共比较序列 R_i 在建立故障全息序列案例库的同时, 生成案例库所有全息序列 S_i 与公共比较序列 R 的相似度 $Similarity(S_i, R)$ 的集合, 称为公共比较距离集。这样, 在实现全息序列相似性匹配故障诊断的过程中, 可以利用式 (14) 和预先生成的公共比较距离集, 过滤掉一部分样本序列, 尽可能的减少匹配过程中的相似性度量计算次数。可以看出, 不等式 (14) 右边值越小, 即当 S_i 与 R 越相似, Q 与 R 越不同时, 算法的剪枝能力越强。

2) 索引 (Index)

为了进一步提高搜索效率, 这里采用改进的一维 $B+$ 树对公共比较距离集进行索引。 $B+$ 树可以进行两种查询: 从最小关键字起顺序查找, 另一种是从根结点开始, 进行随机查找。结点包含关键字的含义为 $|S_i| - Similarity(S_i, R)$ 。在叶结点原有结构的基础上, 增加指向案例库中 S_i 位置的指针。可见, 这样通过公共比较距离, 利用简单的一维索引结构就可以对全息序列案例库间接实现了索引。这里, 案例库必须由等长的带有故障标记的样本全息序列构建。该公共比较序列相似度索引树只在有新样本加入案例库时, 才需要更新。

基于全息序列相似性度量模型和上述剪枝搜索策略, 本文提出了全息序列相似性匹配故障诊断算法 HSSM (holospectrum series similarity matching)。算法的形式化描述及相关解释如表 1 所示。

在旋转机械全息诊断实践中, 各阶次椭圆的长短轴长度主要反映转子振动强度大小, 而椭圆间的大小比例关系与故障类型联系地更紧密, 因此, 在使用全息序列前, 需要对全息椭圆特征向量中的长短半轴进行规格化, 使其映射到同一比例观测范围内。以消除振动强度对故障分类公正性的影响, 规格化采用最大最小规格化方法:

$$\begin{cases} a' = \frac{a - \min a}{\max a - \min a} \times (1 - 0) + 0 \\ b' = \frac{b - \min b}{\max a - \min b} \times (1 - 0) + 0 \end{cases} \quad (15)$$

表 1 HSSM 算法
Table 1 HSSM algorithm

算法: $FaultClass = HSSM(Tree, HS_Database, Q, R, k)$
输入: $Tree: (S_length - Similarity(S_i, R))$ 构成的 B^+ 树; $HS_Database$: 故障全息序列案例库, 数据记录为带有类别标记的 S_i ; Q : 待测全息序列; R : 公共比较序列; k : 近邻数
输出: $FaultClass$ 故障类别
1 $knn_array = Zeros(k); // k$ 近邻结果集 $// k$ 个类别标记为 NULL 的 0 对象组成的数组;
2 $L_{QR} = Similarity(Q, R);$
3 $p_leaf = Tree.sqt // Tree$ 中最小关键字指针;
4 $while(p_leaf) \{$
5 $\quad min_BestSoFar = knn_array[1];$
6 $\quad threshold = min_BestSoFar - L_{QR}; //$ 剪枝条件
7 $\quad p_leaf = SearchBplusTree(Tree, threshold);$
8 $\quad V_array = Tree.Larger(threshold, ASC);$ $//$ 将 $Tree$ 中所有大于 $threshold$ 的叶节点
v_1, v_2, \dots, v_h 升序排列存入 V_array ;
9 $\quad for\ each\ v_i\ in\ V_array \{$
10 $\quad\quad L_{vQ} = Similarity(v_i, Q);$
11 $\quad\quad if\ (L_{vQ} > min_BestSoFar) \{$
12 $\quad\quad\quad InseUpdate(knn_array, L_{vQ});$ $//$ 将 L_{vQ} 插入 knn_array 中, 且保留最大 k 个值升序存于 knn_array ;
13 $\quad\quad\quad break$
14 $\quad\quad \}$
15 $\quad \}$
16 $\} FaultClass = knnVote(knn_array); //$ 投票分类

4 旋转机械故障诊断实例分析

实验采用 ZHS-2 型双跨多功能转子实验台对旋转机械的运行状态进行仿真, 基本结构及配置如图 2。该实验台采用额定电流 2 A、输出功率 300 W 的永磁式直流伺服电动机经联轴器直接驱动转子。通过调速器可以手动实现 0~10 000 r/min 范围的无级调速。实验中, 模拟旋转机械七种运行状态的升速过程: 无故障、转子不平衡、转子不对中、动静摩擦、转子横向裂纹、轴承座松动、转子弯曲。转子转速由 300 r/min 上升到 2400 r/min 每增加一定速度, 水平 X 和垂直 Y 方向的电涡流传感器进行一次采样, 每种运行状态每个特定转速采集 50 组振动位移数据, 每组包含 X 和 Y 两个方向各 1 024 个采样点, 采样频率为 1 024 Hz。选取的特征频率点包括: 1x 工频; 0.2x, 0.4x, 0.6x, 0.8x 四个分倍频; 2x~8x 七个高倍频。采用 Bootstrapping 法对不同阈值设置下的全息序列相似性匹配故障诊断方法进行评估和验证。

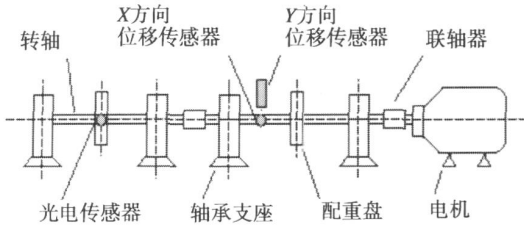


图 2 转子试验台基本结构和配置

Fig. 2 Configuration of the rotor test rig

1) 单位长度全息序列 HSSM 算法分析

分别从每类旋转机械运行状态原始振动位移数据中随机选取 150 组, 共 1 050 组数据构成样本集。每组数据经过除噪等初步预处理后, 转化为一个单位全息序列。对于特定全息椭圆特征向量阈值, 循环 300 次随机从样本集抽取 1 个全息序列作为测试样本 Q_i , 其他数据作为训练样本, 利用 HSSM 法对全息序列 Q_i 进行相似性匹配 3- nn 分类。图 3 显示了不同特征向量阈值设置下 HSSM 法的平均正确率。

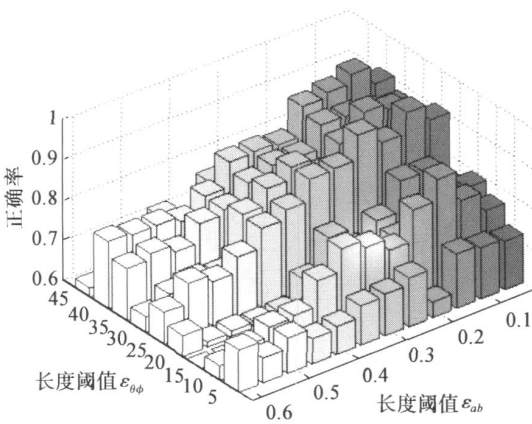


图 3 单位全息序列 HSSM 故障诊断正确率

Fig. 3 Correct rate of the HSSM fault diagnosis based on unit hologram series

图 3 只显示实验结果中, 全息椭圆特征向量阈值向量 v_e 的长度和角度元素分别选取等值的部分, 即 $\varepsilon_{ab} = \alpha_e = \beta_e$, $\varepsilon_{\phi} = \theta_e = \phi_e$ 。其中, 阈值的取值范围为: $\varepsilon_{ab} \in [0.01, 0.6]$, $\varepsilon_{\phi} \in [5, 45]$, 算法中阈值调节值为 $\Delta\varepsilon_{ab} = 0.01$, $\Delta\varepsilon_{\phi} = 5$, 图中每个数据柱为 ε_{ab} 轴上每 5 点的平均值, 即图中 $\Delta\varepsilon_{ab} = 0.05$ 。从图 3 可以看出, 不同 v_e 对 HSSM 算法的正确性影响不同, 随着长度阈值的减小和角度阈值增大, 诊断正确率明显提高, 但角度阈值增大到一定值后, 又出现降低情况, 当阈值范围为 $\varepsilon_{ab} \in [0.01, 0.2]$, $\varepsilon_{\phi} \in [25, 40]$ 时, 诊断正确率集中在 90% 以上, 最高为 93.47%。

2) 长度大于 1 的全息序列 HSSM 算法分析

每类旋转机械运行状态选取 60 个连续转速情况下各 50 组数据, 每组数据利用本文所述方法转化为长度等于 60 的全息序列。依次从每类全息序列集中随机抽取一个全息序列作为公共比较序列 R , 同样每次以 Bootstrapping 的方式从全息序列集中抽取 50 个序列作为测试序列。表 2 显示了阈值向量 $v_e = [0.1, 0.1, 40, 40]$, 采用不同 δ 设置时, HSSM 故障诊断的部分识别结果。

表 2 HSSM 方法不同 δ 的识别结果Table 2 Test result of the HSSM method with different δ

δ	0	2/12	4/12	6/12	8/12	10/12
正确率 (%)	26.86	39.14	78.86	97.71	98.78	89.67
识别时间 (s)	1.218	1.203	1.202	1.193	1.189	1.184

实验得 δ 在 $[5/12, 8/12]$ 的范围内, 能够达到 97% 以上的正确率, 明显高于单位长度全息序列的情况, 最高可达 98.78%。单次识别时间接近 1 s。正确率提高主要由于加入了旋转机械振动随转速变化的特征, 使得转子横向裂纹和弯曲、动静摩擦等故障的全息序列特征更明显。如, 转子弯曲和不平衡故障的单位长度全息序列很难区分, 但前者不同转速的初相角不同, 使得二者长度大于 1 的全息序列区别很大; 横向裂纹故障在临界转速附近状态变化剧烈, 同样体现在全息序列特征的明显不同。

全息序列定义中的阈值 v_e 和 δ 限定了相似全息序列间的变形尺度, 即同类状态间的差异程度。选取的 v_e 元素过大或 δ 过小会增大相似性匹配的漏报率, 反之则会增大错报率, 而且设备的类型和工况等, 也会对全息序列的形态产生影响。所以, 具体阈值的选取需要根据工业现场的实际需求来确定, 可以采用实验 1) 和 2) 的实验搜索法, 选取最高故障诊断正确率所对应的 v_e 和 δ ; 也可以采用案例统计法, 利用案例库中同类全息序列全息椭圆特征向量元素间对应的标准差范围来确定 v_e 元素的范围, 进而利用各类别中同类全息序列元素间相似度的平均值与标准差差值的最小值来确定 δ 。其中, 实验搜索法为首选。

3) HSSM 算法剪枝能力分析

剪枝率 (Pruning Rate) 可以反应算法的剪枝能力, 其中

$$\text{PruningRate} = 1 - \frac{\text{Similarity}(P, Q) \text{ 计算次数}}{\text{样本集大小}} \quad (16)$$

显然, 不采用任何剪枝策略的顺序查找 k - nn 分类器的剪枝率为 0。分别选取四类故障典型数据作为公共比较序列 R , 比较 HSSM 算法的剪枝率, 如表 3 所示。

表 3 HSSM 方法不同公共比较序列的剪枝率

Table 3 Pruning rate of the HSSM method with different public comparative holospectrum series

公共比较序列 R	正常	不平衡	横向裂纹	动静摩擦
Pruning rate(%)	32.34	32.92	17.09	16.12

如表 3 所示, 采用正常和不平衡两类公共比较序列的 HSSM 算法剪枝率较高。分析原因, 主要是由于相对于横向裂纹和动静摩擦, 这两类状态与其他各类状态的全息序列间相似成分较多且较均衡。

5 结 论

本文利用全息序列表示旋转机械的整体振动状态特征, 将经傅里叶变换分解且通过比例内插校正后的旋转机械各支承载面不同方向的振动信号集成映射为类时间轴上的多维序列, 结合故障领域应用的特点, 定义了全息序列的相似性度量模型, 并提出适用于全息序列相似性匹配故障分类的 HSSM 算法。实验结果表明, HSSM 故障诊断过程中不需要借助专家经验和人工识别确认, 在一定阈值范围内能够实现高质量的旋转机械故障诊断, 对提高旋转机械故障识别的准确率及自动化水平是有效的。

参考文献

- [1] 屈梁生. 机械故障的全息诊断原理 [M]. 北京: 科学出版社, 2007
- QU L S. Holspectrum and hobbalancing technique in machinery diagnosis [M]. Beijing: Science Press, 2007.
- [2] GAVRILOV M, ANGUELOV D, et al. Mining the stock market Which measure is best [C]. Proc. of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Boston, Massachusetts, United States, 2000: 487-496.
- [3] M ICHAIL V, MARIDS H. Indexing multidimensional time series with support for multiple distance measures [C]. Proc. of the 9th ACM SIGKDD International Conference on Knowledge Discovery and data mining, 2003: 216-255.
- [4] EAMONN K, LIW. LBK eough supports exact indexing of shapes under rotation invariance with arbitrary representations and distance measures [C]. Proc. of the 32nd International Conf on Very Large Data bases Table of Contents, Korea, 2006: 882-893.

- [5] 万书亭, 李和明. 汽轮发电机组轴心轨迹进动方向自动识别新方法 [J]. 中国电机工程学报, 2003 (23): 146-149.
- WAN SH T, LIH M. A new method for automatically identifying the axis traces moving direction of turbine generator unit [J]. Proceedings of the CSEE, 2003 (23): 146-149.
- [6] 杨苹, 吴捷. 火电厂锅炉常见故障的数据挖掘诊断方法 [J]. 仪器仪表学报, 2005, 26 (7): 696-701.
- YANG P, WU J. Fault diagnosis for boilers in thermal power plant by data mining [J]. Chinese Journal of Scientific Instrument, 2005, 26 (7): 696-701.
- [7] CHEN L, RAYMOND N G. On the marriage of Lp-norms and edit distance [C]. Proc. of the 30th VLDB Conference, Toronto, Canada, 2004: 792-803.
- [8] 黄河, 史忠植. 基于形状特征 k-d 树的多维时间序列相似性搜索 [J]. 软件学报, 2006, 17 (10): 2048-2056.
- HUANG H, SHI ZH ZH. Similarity search based on shape k-d tree for multidimensional time sequences [J]. Journal of Software, 2006, 17 (10): 2048-2056.

作者简介



吴薇, 2003 年于辽宁大学获学士学位, 现为中国科学院沈阳自动化研究所博士研究生, 主要研究方向为旋转机械故障诊断与预测, 时间序列挖掘等。

E-mail: wuwe@sia.cn

Wu Wei received BSc from Liaoning University in 2003. She is now a PhD candidate in Shenyang Institute of Automation, Chinese Academy of Science. Her research interests are rotating machinery fault diagnosis and prognosis and time series data mining.

E-mail: wuwe@sia.cn



胡静涛, 分别于 1985 年和 1988 年于大连理工大学获得学士学位和硕士学位, 现为中国科学院沈阳自动化研究所研究员, 中国科学院研究生院教授, 博士生导师, 主要研究方向为远程设备监测与故障诊断。

E-mail: hu.jingtao@sia.cn

Hu Jingtao received BSc from Dalian Institute of Technology in 1985 and MSc from Dalian University Science and Technology in 1988. He is now a professor both in Shenyang Institute of Automation and Graduate School of Chinese Academy of Science. His main research interest covers the areas of remote monitoring and diagnostics of equipment.

E-mail: hu.jingtao@sia.cn