基于时域特征和改进k-均值聚类算法的风机振动分析*

周云龙[1]， 王锁斌[1]， 赵 鹏[2]
(1.东北电力大学能源与机械工程学院 吉林,132012) (2.华北电力大学能源与动力学院 昌平,102206)

摘要 针对风机振动信号的非平稳和非线性特征,提出了一种基于时域信号分析和改进的k-均值聚类算法的故障识别方法。对离心式风机运行中产生的几种非稳态振动故障信号,提取其时域信号的峰峰值、Hurst 指数和近似熵参数作为特征向量,采用改进的k-均值聚类算法作为故障分类器,设置转子不平衡、联轴器不对中、风机基座松动、转轴径向摩擦和轴承内圈损坏5种故障。对离心式风机试验的结果表明,3种时域特征能较好地反映各故障之间的差异,改进的k-均值聚类算法与原始的k-均值算法相比分类性能更好,稳定性更强,平均识别率达到88.67%。

关键词 故障诊断; 离心式风机; 时域特征; 改进k-均值聚类算法
中图分类号 TP206.3

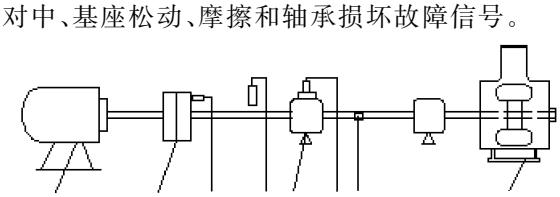
引 言

设备故障诊断的实质是模式识别问题,包括特征提取和故障识别两部分。目前大多数智能诊断方法是以频谱频带能量为特征,以神经网络进行故障识别。这种模式存在以下问题:a.由于傅里叶频谱不能反映任何时域信息,所以只对平稳信号的分析有效,但是实际现场由于电网电压的波动以及设备自身的非线性等因素影响,设备振动信号通常表现出非平稳性[1];b.文献[2-3]利用神经网络对轴承故障进行诊断得到了不错的结果,但是神经网络用于故障识别需要大量的训练样本,而这在现实中往往难以满足,且神经网络的泛化能力不是非常理想。

笔者提出了基于时域特征参数和改进k-均值聚类的智能诊断方法,充分利用信号最直接的时域信息,提取振动信号的峰峰值、Hurst 指数和近似熵等对风机故障敏感度很高的参数来表征非平稳信号的特征。k-均值聚类是一种无监督的学习方法,在不知道样本类别的情况下,根据样本的特征向量来分类样本,在解决小样本问题中表现出独特的优势和良好的应用前景,并具有优良的泛化能力。

1 试验装置及试验方法

试验系统中离心风机的型号为Y5-47315,最大转速为2 900 r/min,风压为803 Pa,风量为1 830 m^3/h,电动机的型号为Y90S-2,功率为1.2 kW,电压为380 V,电流3.4 A,最大转速为2 840 r/min。为便于振动信号的测取,跟实际的风机系统相比,试验装置在风机跟控制电机之间多了两个轴承[4],风机和电机之间由刚性联轴器连接,离心风机轴的垂直和水平方向分别安装非接触式电涡流位移传感器测取径向位移,风机轴承座上平面安装LC0119T 型加速度传感器。风机联轴器的垂直面作为试验测试面,水平安装非接触电涡流位移传感器测取轴向位移,系统测取转子联轴器不对中振动加速度信号。试验过程中,在风机额定转速下,保持离心风机入口调节阀开度,使风机的负荷维持在80%,采样频率为800 Hz,试验系统的整体结构如图1 所示。图2 中的第1 个信号为风机正常运行信号,第2 到第6 依次为不平衡、不对中、基座松动、摩擦和轴承损坏故障信号。



1-电机; 2-联轴器; 3-轴向位移传感器; 4-垂直位移传感器; 5-轴承; 6-加速度传感器; 7-水平位移传感器; 8-离心风机

图1 离心风机试验装置

2 基于时域分析的振动信号特征提取

2.1 峰峰值

波形峰峰值分析反映了振动信号的局部幅值强

* 吉林省教育厅资助项目(编号:2007047)
收稿日期:2010-06-07;修改稿收到日期:2010-09-28

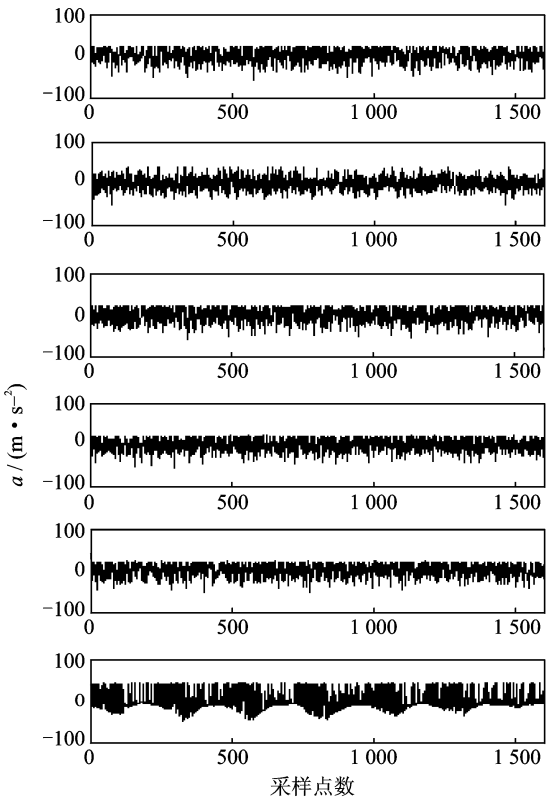


图2 几种典型风机振动信号

度变化，试验中采用加速度信号为分析信号，加速度是单位时间内速度的变化率。风机运行时，在不同故障条件下转子受到的应力不同，其加速度会不同，反映到波形上的峰峰值也会不同；所以，可以选用峰峰值作为故障特征，来反映风机的运行状态。由于实际采样的原始信号没有明确的起始点，不是风机转动工作周期的整数倍，这样会造成信号间的可比性很差，不利于下一步的故障诊断，所以需要对原始信号截取风机转动的 n 个工作周期进行分析，以减小误差对特征提取的影响。定义振动信号的峰峰值为

$$x_{pv} = \max(x_i) - \min(x_i) \tag{1}$$

其中：x_{pv} 为振动信号峰峰值；$\max(x_i)$ 为振动信号波峰值；$\min(x_i)$ 为振动信号波谷值。

n 个周期信号的平均峰峰值为

$$\overline{x}_{pv} = \sum_{i=1}^{n} x_{pv}/n \tag{2}$$

2.2 Hurst 分析计算方法

英国水利学家Hurst 在研究尼罗河水位的涨落问题时发现，大多数自然现象，包括河水水位、温度、降雨、太阳黑子等，不服从布朗运动及高斯分布的特征，而是遵循一种“有偏随机游动”——趋势加噪声。

分形布朗运动是一个能反映广泛的自然物体一些不规则运动性质的分形模型，它的数值变化非常复杂，连续但不可导，是一个非平稳过程，对时间和尺度的变化具有自相似性。

大量试验表明，风机故障振动信号具有非平稳性，因此可以采用分形布朗运动来描述此信号。分形布朗运动(FBM)增量方差为

$$M(B_H(t) - B_H(t_0)) = \sigma_0^2 \, |t - t_0|^{2H} \tag{3}$$

其中：B_H 为分数Brown 函数；t 为时间；σ_0^2 为 t_0 时刻的样本方差；H 为Hurst 指数。

Hurst 指数决定了一个FBM 的不规则程度，并描述了随机过程的长期相关性。笔者采用R/S 分析法计算风机振动信号的Hurst 指数[5]，其计算过程如下：设采集的振动信号时间序列为 $X(t)(t=1,2,\cdots,T)$，令

$$X^*(t) = \sum_{u=1}^{\tau} X(u)$$

则极差 $R(t,\tau)$ 为

$$R(t,\tau) = \max_{0\leqslant u\leqslant \tau} |X^*(t+u) - X^*(t) - \frac{u}{\tau}[X^*(t+u) - X^*(t)]| - \min_{0\leqslant u\leqslant \tau} |X^*(t+u) - X^*(t) - \frac{u}{\tau}[X^*(t+u) - X^*(t)]| \tag{4}$$

标准偏差 $S^2(t,\tau)$ 为

$$S^2(t,\tau) = \frac{1}{\tau}\sum_{u=t+1}^{1+\tau} X^2(u) - \left[\frac{1}{\tau}\sum_{u=t+1}^{1+\tau} X^2(u)\right]^2 \tag{5}$$

其中：τ 为延迟时间。

Mandelbrot 等[6-7]的研究表明，$R(t,\tau)/S(t,\tau)$ 与延迟时间 τ 之间存在如下关系

$$E(R(t,\tau)/S(t,\tau)) \propto \tau^H$$

$E()$ 表示同一延迟 τ 下对不同初始时刻 t 取平均值，这样可以消除不同初始时刻的端效应对统计计算的影响。通过作 $\log E(R(t,\tau)/S(t,\tau))$-$\log\tau$ 关系图，并回归直线关系的斜率，就可求得Hurst 指数。

2.3 近似熵的定义及其性质

近似熵是用一个非负数来表示某时间序列的复杂性，越复杂的时间序列对应的近似熵越大。

设采集到的原始数据为 $\{u(i), i=0,1,\cdots,n\}$，预先给定模式维数 m 和相似容限 r 的值，则近似熵可以通过以下步骤计算得到：

1) 将序列 $\{u(i)\}$ 按顺序组成 m 维矢量 $\boldsymbol{X}(i)$，即

$$\boldsymbol{X}(i) = [u(i), u(i+1), \cdots, u(i+m-1)]$$
$$(i = 1 \sim n-m+1)$$

2) 对每一个 i 值计算矢量 $\boldsymbol{X}(i)$ 与其余矢量 $\boldsymbol{X}(j)$ 之间的距离

$$d[\boldsymbol{X}(i), \boldsymbol{X}(j)] = \max_{k=0\sim m-1} |u(i+k) - u(j+k)|$$

3）按照给定的阈值$r(r>0)$，对每一个i值统计$d[\boldsymbol{X}(i),\boldsymbol{X}(j)]<r$的数目及此数目与总的矢量个数$n-m+1$的比值，记做$C_i^m(r)$，即

$$C_i^m(r)=\{d[\boldsymbol{X}(i),\boldsymbol{X}(j)]<r\text{的数目}\}/(n-m+1)$$

4）先将$C_i^m(r)$取对数，再求其对所有i的平均值，记做$\Phi^m(r)$，即

$$\Phi^m(r)=\frac{1}{n-m+1}\sum_{i=1}^{n-m+1}\ln C_i^m(r) \tag{6}$$

5）再对$m+1$，重复1～4的过程，得到$\Phi^{m+1}(r)$；

6）理论上此序列的近似熵为

$$\mathrm{ApEn}(m,r)=\lim_{N\to\infty}[\Phi^m(r)-\Phi^{m+1}(r)] \tag{7}$$

一般而言，此极限值以概率1存在。但实际工作中n不可能为∞，当n为有限值时，按上述步骤得出的是序列长度为n时ApEn的估计值，记做

$$\mathrm{ApEn}(m,r,n)=\Phi^m(r)-\Phi^{m+1}(r) \tag{8}$$

ApEn的值显然与m,r,n的取值[8]有关。根据工程上的经验，通常取$m=2$，这样序列在联合概率下进行动态重构时，会含有更多的详细信息。对于参数r和n的选取，为使近似熵具有较为合理、有效的统计特性，并且尽可能地减小伪差，选$r=0.2\mathrm{SD}(u)$（SD表示序列$\{u(i)\}$的标准差），选取n为1 600、时间长度为2 s、风机转动10个周期的数据点。

可以看出，近似熵的计算实际上是在确定一个时间序列在模式上的自相似程度有多大，即在衡量维数变化时该时间序列中产生新模式的概率的大小。产生新模式的概率越大，序列就越复杂；因此从理论上讲，近似熵能够表征信号的不规则性（复杂性），振动情况越复杂的信号近似熵越大。近似熵只是希望从统计的角度来区别时间过程的复杂性，因此只用较短的数据就可以估计出合理的近似熵。文献[9]指出，近似熵大致相当于维数变化时新模式出现的对数条件概率的均值，在衡量时间序列的复杂性方面具有一般意义，而不仅仅是一个非线性动力学参数，因此近似熵的估计对随机过程和确定性过程都适用。同时，当噪声的幅度低于相似容限时，该噪声将被抑制，若时间序列中存在较大的瞬态干扰时，干扰产生的数据（即野点）与相邻数据组成的矢量与$\boldsymbol{X}(i)$的距离必定很大，因而在阈值检测中将被去除；因此，近似熵具有很好的抗噪、抗野点能力。

3　改进的k-均值聚类算法

假设有n个未知标号的样本$(x_1,x_2,\cdots,x_n)$，如何根据样本的特征向量，将样本分为k类：$\alpha_1,\alpha_2,\cdots,\alpha_k$。假设第$k$类的样本数目为$n_k$，则$n=\sum_{i=1}^{k}n_i$，每类$\alpha_k$的均值为$m_1,m_2,\cdots,m_k$，则$m_k=\frac{1}{n_k}\sum_{i=1}^{n_k}x_i$，$K=1,2,\cdots,k$。

k-均值聚类是基于误差平方和准则，即k-均值聚类最小化的目标函数为

$$J=\sum_{k=1}^{K}\sum_{i=1}^{n_k}\|x_i-m_k\|^2 \tag{9}$$

3.1　k-均值算法思想

1）任意选取样本中的k个对象为初始聚类中心；

2）对于其他对象，根据其与选定的k个聚类中心的距离（相似度），把它们归类到最相似的聚类中，并且重新计算所获聚类的中心；

3）如果聚类最小化的目标函数达到精度要求，则聚类中心不移动，算法终止，否则转到第2步。

3.2　k-均值算法的改进

由于k-均值算法对于初始聚类中心的选取是随机的，很容易陷入局部最优值，导致分类误差，所以需要把局部聚类中心移动到更有利于分类的位置[10]。这里定义变形误差公式为

$$I=S-N[d(w,x_0)]^2 \tag{10}$$

其中：S为某一聚类里所有对象与欧式空间中心的距离平方和，N为属于这一聚类的对象个数，$d(w,x_0)$为这一聚类中心到欧式空间中心x_0的距离。

定义$\Delta M=\Delta I-\Delta D$为聚类中心移动准则，其中：$\Delta I$为移出聚类中心引起的整体变形误差增大；$\Delta D$为插入新的聚类中心引起的变形误差下降，当$\Delta M<0$时，聚类中心的移动可以是整体的变形误差减小。

1）任意选取样本中的k个对象为初始聚类中心；

2）把训练样本中每一个对象归于距离其最近的聚类中，并重新计算聚类中心；

3）如果聚类最小化的目标函数达到精度要求，则聚类中心不移动，转到第4步；

4）根据聚类中心移动准则，若有一个聚类中心可以移到更好的位置来减小整体的变形误差和，则把它转到更好的位置，然后转到步骤2，否则停止。

4　试验结果与分析

对试验中采集到的风机在正常、联轴器不对中、转子不平衡、基座松动、摩擦和轴承内圈损坏工况下的振动信号，分别提取其时域信号的峰峰值、混沌特

性的Hurst 指数以及近似熵数据如表1 所示。

试验中,对风机运行中出现的6 种工况,提取300 个样本,每种工况为50 个样本,在每种振动信号中选取30 组,共180 组作为学习样本,剩余120 组作为测试样本,使用改进的k-均值聚类算法进行分类,当前后两次迭代的整体变形误差小于ε时,算法终止。这里取k 为6,ε为10^{-4},改进前、后的k-均值取聚类算法的一些数据对比见表2。

从试验结果可以看出,由于原始的k-均值聚类算法采用随机选取聚类中心,很容易陷入局部最小值,所以其平均识别率不高;而改进的k-均值聚类算法由于采用了移动局部最优聚类中心的步骤,使其分类性能大大提高,稳定性加强,但是由于其算法的复杂度较初始算法高,所以识别时间较改进前要长一些。

表1 部分风机振动信号的特征数据

序号	振动类型	峰峰值	Hurst 指数	近似熵
1	正常	13.438	0.169 2	0.558 9
1	正常	15.653	0.169 8	0.564 3
2	不平衡	35.034	0.177 3	1.534 2
2	不平衡	34.159	0.176 2	1.541 8
3	不对中	89.159	0.153 4	4.652 1
3	不对中	90.243	0.152 1	4.765 1
4	基座松动	61.340	0.087 5	2.865 0
4	基座松动	59.280	0.083 7	2.854 9
5	摩擦	68.534	0.065 8	3.352 4
5	摩擦	68.033	0.066 6	3.348 7
6	轴承内圈磨损	37.862	0.113 7	2.259 4
6	轴承内圈磨损	36.643	0.112 5	2.223 8

表2 改进的k-均值聚类算法与原始算法的识别性能对比

分类器	分类耗时/s	平均识别准确率/%
k-均值聚类算法	1.2	76.22
改进的k-均值聚类算法	2.3	88.67

5 结 论

1) 风机振动信号的峰峰值、Hurst 指数和近似熵很好地反应了风机振动信号的非平稳性、复杂性,是有效的时域信号识别度量参数。

2) 改进的k-均值聚类算法用于模式识别的实现步骤比较简单,不需要长时间的训练过程,克服了由于随机选取初始聚类中心导致的局部最小值问题,但是由于其算法复杂度高一些,所以分类时间会长一点。

3) 试验证明,基于时域混合特征与改进的k-均值聚类算法相结合的风机故障诊断方法是可靠的。必须指出的是,上述试验是在小样本情况下得到的,如何提高其在大样本情况下的分类稳定性和正确率是今后研究的关键。

参 考 文 献

[1] 周云龙,柳长昕,赵鹏. 基于自回归-连续隐马尔科夫模型的离心水泵故障诊断[J]. 中国电机工程学报,2008,28(20):88-93.

[2] 李萌,陆爽,马文星. 滚动轴承故障诊断的分形特征研究[J]. 农业机械学报,2005, 36(12):162-164.

[3] 刘良顺,魏立东,宋希庚,等. 基于RBF 神经网络的滚动轴承故障诊断方法[J]. 农业机械学报,2006, 37(3): 163-165

[4] 程军圣, 于德介, 杨宇. EM D 方法在转子局部碰摩故障诊断的应用[J]. 振动、测试与诊断, 2006, 26(1): 24-25.
Cheng Junsheng, Yu Dejie, Yang Yu. Application of EMD to local rub-impact fault diagnosis in rotor systems[J]. Journal of Vibration, Measurement & Diagnosis, 2006, 26(1): 24-27. (in Chinese)

[5] 杨洁明, 田英. 基于EMD 和球结构SVM 的滚动轴承故障诊断[J]. 振动、测试与诊断, 2009, 29(2): 155-158.
Yang Jieming, Tian Ying. Roller bearing fault diagnosis by using empirical mode decomposition and sphere-structured support vecator machine[J]. Journal of Vibration, Measurement & Diagnosis, 2009, 29(2): 155-158. (in Chinese)

[6] Mandelbrot B B, Wallis J R. Some long-run properties of geophysical records [J]. Water Resources Res, 1969, 5(2): 321-340.

[7] Mandelbrot B B, Wallis J R. Robustness of the re-sealed range R/S in the measurement of noncyclical long run statistical dependence[J]. Water Resources Res, 1969, 5(5): 967-988.

[8] 齐子元,徐章遂,卢志才. 近似熵在发动机故障诊断中的应用研究[J]. 军械工程学院学报,2008,20(2):39-42.

[9] 胥永刚,李凌均,何正嘉. 近似熵及其在机械设备故障诊断中的应用[J]. 信息与控制,2002,31(6):547-551.

[10] Fritzke, B. The LBG-U method for vector quantization-an important over LBG inspired from neural networks[J]. Neural Processing Letter 5, 1997, (1): 35-45.

第一作者简介:周云龙 男,1960 年生,教授。主要研究方向为振动测试技术。曾发表《基于自回归-连续隐马尔可夫模型的离心水泵故障诊断》(《中国电机工程学报》2008 年第28 卷第20 期)等论文。
E-mail:ylzhou@mail.nedu.edu.cn