基于 EMD 和优化 K-均值聚类算法 诊断滚动轴承故障*

郭艳平,颜文俊

(浙江大学 电气工程学院, 杭州 310027)

摘 要:考虑到滚动轴承振动信号的非平稳特征和实际应用中典型故障样本不易获得等原因,而在实际应用中,故障程度识别和故障类型诊断一样重要,提出一种滚动轴承故障类型及故障程度识别方法。首先对原始振动信号进行 EMD 分解,对含故障特征的 IMF(intrinsic mode function)分量进行信号重构,随后对重构信号进行 Hilbert 包络谱分析,在提取特征量的基础上,应用优化 K-均值聚类算法进行故障类型和故障程度分类。实验结 果表明:基于 EMD 和优化 K-均值聚类的故障类型和故障程度识别算法,可将含不同故障类型的样本集,按故障 类型进行正确分类;也可将含同种故障类型、不同故障程度的样本集,按故障程度进行正确分类。

关键词:滚动轴承;故障诊断;故障程度;EMD;K-均值聚类

中图分类号: TP391 文献标志码: A 文章编号: 1001-3695(2012)07-2555-03 doi:10.3969/j.issn.1001-3695.2012.07.041

Fault diagnosis of bearing based on empirical mode decomposition and K-means clustering

GUO Yan-ping, YAN Wen-jun

(College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China)

Abstract: The vibration signal is nonstationary and the sample with typical fault is difficult to acquire. The severity of fault diagnosis is the same important to style diagnosis. This paper presented a novel fault diagnosis of bearings based on the characteristic fault frequency and K-means clustering. The reconstructed signal could be obtained by some set of IMF components of the vibration signal by EMD. It performed the Hilbert envelope analysis to reconstructed signal. From the power spectrum of Hilbert envelope signal, it could identify the amplitude of the characteristic fault frequency and its integer multiples, which was used to diagnose the style and severity of fault. The result demonstrates that the proposed method based on EMD and K-means clustering can recognize the style and severity of bearing fault.

Key words: rolling bearing; fault diagnosis; severity of fault; EMD(empirical mode decomposition); K-means clustering

0 引言

滚动轴承作为机械设备中一种通用的连接和传递动力的 零部件,几乎在任何大型设备中都具有重要作用。采用先进技 术对轴承进行状态监测和故障诊断,可实现轴承由事后维修、 定期维修到监测维修的转变,减少不必要的人力和物力浪费, 为提高经济效益贡献力量。

现在已经有很多学者在轴承故障诊断和故障程度识别方 法上取得了成绩,小波变换^[1]、Wigner-Ville分布^[2]、Hilbert-Huang变换(Hilbert-Huang transform, HHT)^[3,4]、EMD^[5-7]已经 被用于滚动轴承故障诊断中,并取得了不错的效果。Wigner-Ville分布在对多分量信号进行分析时易产生交叉项,这一缺 陷限制了其广泛应用;固定的小波基和恒定的分辨率使得小波 分析没有摆脱傅里叶分析的局限。EMD 能根据信号的局部特 征,自适应地将一个复杂信号分解为一系列本征模态函数 IMF,每一个 IMF 分量都对应真实的物理意义。与窗口傅里叶 变换、Wigner-Ville、小波变换对比来说, EMD 是一种具有自适

应的多分辨率分析方法,它非常适合对非线性、非平稳信号进 行分析^[8]。已经有许多学者将 EMD 与模式识别方法结合起 来应用于机械系统的故障诊断,但已有的诊断方法大多对训练 样本的数量和质量有一定的要求,且大多诊断方法只对故障类 型进行了分类,而对于在实际应用中至关重要的故障程度并没 有太多涉及。文献[9]将 EMD 与神经网络模式识别方法相结 合对滚动轴承进行故障诊断,此方法需有足够数量的样本来训 练神经网络,且对训练样本的质量要求较高;文献[10]提出了 一种基于经验模态分解和支持向量机的柴油机振动信号分析 方法,该方法也需要一定数量具有典型故障特征的训练样本, 在少量训练样本的情况下,支持向量机的总体分类性能会有所 降低。针对传统信号分析方法对滚动轴承振动信号等非稳定 信号分析不具有自适应性,以及实际典型故障样本不易获得等 原因,本文提出了一种基于经验模态分解和优化 K-均值聚类 的滚动轴承故障诊断方法,该方法可自适应地将样本按故障类 型和故障程度进行分类,该方法具有算法简单、结果准确的 优点。

收稿日期: 2011-11-29; 修回日期: 2012-01-06 基金项目: 国家"863"计划资助项目(2007AA05Z432926)

作者简介:郭艳平(1983-),女,河北邢台人,博士研究生,主要研究方向为风力发电机在线监测及故障诊断(gouyanping1983@163.com);颜文 俊(1965-),男,浙江温州人,教授,博导,主要研究方向为优化控制、复杂系统建模与分析.

文中所用数据均来自美国 Case Western Reserve University 电气工程实验室的滚动轴承故障模拟实验台^[11],该实验台包 括一个2马力的电动机、一个扭矩传感器和一个功率测试计。 待检测的轴承位于电动机的两端,驱动端轴承型号为 SKF6205,风扇端轴承型号为SKF6203,轴承故障点用电火花加 工而成,损伤点的直径分别为0.1778 mm、0.3556 mm、0.5334 mm。其中,轴承外圈的损伤点分别在时钟:3 点钟、6 点钟、12 点钟三个方向,振动数据由布置在电动机外壳上的振动加速度 传感器采集得到,采样频率为12 kHz,功率和转速通过扭矩传 感器/译码器测得。所采集的数据均存储为*.mat 格式 (MATLAB 文件)。

1 EMD 方法

EMD 被认为是一个筛分的过程,它可以将一个多分量的 复杂信号分解为若干个 IMF,每个本征模态函数须满足如下条 件^[12]:a)极值点的个数和零点数相等或相差为一;b)由局部极 大值点连接形成的上包络线和由局部极小值点形成的下包络 线的平均值为零。

应用 EMD 方法对信号 x(t)进行分解,具体步骤^[12]如下:

a)用三次样条线将所有的局部极大值点连接起来,形成 上包络线;将所有的局部极小值点连接起来形成下包络线,此 时,所有的数据点应该都被上、下包络线包络。

b) 计算上、下包络线的平均值为 m1,得

$$x(t) - m_1 = h_1$$
 (1)

如果 h_1 满足上述两个条件,则 h_1 就是信号 x(t) 的第一个 IMF 分量。

c)如果 h_1 不满足上述两个 IMF 的条件,就把 h_1 作为原始 信号数据,重复进行步骤 a)~c),计算上、下包络线的平均值 m_{11} ;然后判断 $h_{11} = h_1 - m_{11}$ 是否满足 IMF 条件,如果不满足, 则继续循环 k 次,得 $h_{1(k-1)} - m_{1k} = h_{1k}$,直到 h_{1k} 满足 IMF 条件。 令 $c_1 = h_{1k}$,即 c_1 为第一个满足 IMF 条件的分量。

d)计算

$$r_1 = x(t) - c_1$$
(2)

再将 r_1 视为原始信号重复步骤 a) ~ d),得到信号 x(t)的 第二个 IMF 分量 c_2 ,得

$$r_1 - c_1 = r_2 \tag{3}$$

以此类推,直到r_n为一个单调函数为止,循环结束,得

$$x(t) = \sum_{i=1}^{n} c_i + r_n$$
 (4)

其中:r_n为残余函数,代表信号的平均趋势。

2 Hilbert 包络谱

正常轴承的振动信号近似服从正态分布,当轴承出现故障 点时,其振动信号的峭度值明显增大。所以,依次计算 $n \land$ IMF 分量的峭度值, $K = [k_1, k_2, \dots, k_n]$,当其峭度值大于 3 时, 说明这个 IMF 含有较多的故障冲击成分,所以对这些 IMF 进 行重构,得重构信号 c(t),这一步骤可剔除信号中的干扰噪声 成分,提高信噪比。

信号包络谱可反映周期性的冲击及其剧烈程度,对上述重构信号 c(t)进行 Hilbert 变换后,得到

$$\hat{c}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{c(t')}{t-t'} \,\mathrm{d}t' \tag{5}$$

可得解析信号如下

$$z(t) = c(t) + \hat{c}(t) = a(t) e^{j\varphi(t)}$$
(6)

Hilbert 包络信号 a(t) 如下

$$a(t) = \sqrt{c^{2}(t) + \hat{c}^{2}(t)}$$
(7)

当轴承的圆周上发生一处点蚀时,滚动体依次滚过工作表 面缺陷时,受到反复冲击而产生低频脉冲,称为轴承的通过频 率或故障特征频率,其发生周期可从转速和轴承的尺寸求得, 故障特征频率因点蚀的位置不同而不同,内环故障特征频率, 滚动体故障特征频率,外环故障特征频率依次表示为*f_i、f_r、 f_a*,其计算式分别为

$$f_i = \frac{1}{2} \left(1 + \frac{d}{D_m} \cos \alpha \right) f_n Z \tag{8}$$

$$f_r = \frac{1}{2} (1 - \frac{d^2}{D_m^2} \cos^2 \alpha) f_n \frac{D_m}{d}$$
(9)

$$f_o = \frac{1}{2} \left(1 - \frac{d}{D_m} \cos \alpha\right) f_n Z \tag{10}$$

其中:d为滚动体直径; D_m 为滚动轴滚道节径; α 为滚动轴承的 压力角; f_n 为轴的旋转频率;Z为滚动体数量;

3 特征量提取



图1 具有内环点蚀故障的滚动轴承振动信号Hilbert包络谱图

当滚动轴承发生故障时,在频谱图上会出现轴的旋转频率 及其倍频、轴承的故障特征频率及其倍频。图1是在以12 kHz 为采样频率,故障点位于驱动端内环、故障点直径为0.177 8 mm、电动机转速为1730 rpm的情况下采集到的振动信号故障 样本经 EMD 和 Hilbert 变换得到的包络谱图。图中数据点 (30.03,2466)和(60.06,713.5)的横坐标分别为轴的旋转频 率 f_n 及其二倍频 $2f_n$ 、(161.9,2587),(323.7,1310)和(485.6, 1019)的横坐标分别为内环故障特征频率 f_i 及其二倍频 $2f_i$ 和 三倍频 $3f_i$ 。所以选取能表征故障类型的旋转频率及其二倍 频、故障特征频率及其二倍频和三倍频(依次表示为: f_n 、 $2f_n$ 、 f_i 、 $2f_i$ 、 $3f_i$ 、 f_o 、 $2f_o$ 、 $3f_o$ 、 f_r 、 $2f_r$ 、 $3f_r$)处的幅值为特征量,依次表示 为 $C = [A_{f_n}, A_{2f_i}, A_{2f_i}, A_{f_0}, A_{2f_0}, A_{3f_0}, A_{2f_0}, A_{3f_0}, A_{2f_0}, A_{3f_r}]。$

4 优化 K-均值聚类

4.1 原始 K-均值聚类算法概述

K-均值聚类又称为快速聚类算法,由于它计算简单、可动态聚类、自适应强等特点而被广泛应用^[13,14],其基本步骤为

a)选 k 个样本作为初始聚类中心;

b)对除聚类中心之外的所有样品逐个归类,每个样本被 归入离它最近的聚类中心点所在的类,该类的中心点更新为这 一类目前的均值,以此类推,直至所有样本都归类;

c)重复上述步骤 b),直至所有样本不能再归类为止。

4.2 K-均值聚类算法的优化

原始 K-均值聚类算法的最终聚类结果在一定程度上依赖 于初始聚类中心的选择,且易受孤立点的影响,聚类个数也无 法选择。针对以上问题,对原始算法作了如下优化:

a)基于数据密度和散度指标选择初始聚类中心

将样本集中数据密度最大的点作为第一个聚类中心 z₁,将 与 z₁ 散度值最大的高密度点作为第二个聚类中心,依次类推, 直到初始聚类中心可反映数据发布特征为止。

b)在迭代过程中排除噪声点的影响

以数据密度和散度指标来寻找远离数据集中心的极端数 据点,在每次迭代计算新的聚类中心时,也将噪声点排除在外, 这样可使聚类中心更能反映正常数据的特征。

5 实验结果及分析

由于篇幅原因,表1只列出了部分故障样本的特征量,其 中,以n开头的样本为正常样本;i开头的样本为内环有点蚀故 障的样本;以o开头的样本为外环有点蚀故障的样本;以r开 头的样本为滚动体有点蚀故障的样本。用优化 K-均值聚类算 法,把样本分成了四类,结果显示,优化 K-均值聚类算法将所 有样本按故障类型分成了四类(正常、内环故障、外环故障、滚 动体故障),即第一类:n-097;n-098;n-099;n-100;第二类i-105, *i*-106, *i*-107, *i*-108, *i*-169, *i*-170, *i*-171, *i*-172, *i*-209, *i*-210, *i*-211, *i*-212;第三类:*o*-130;第四类:*r*-3006。对表1中具有内环故障 的 12 个样本(i-105,i-106,i-107,i-108,i-169,i-170,i-171,i-172,*i*-209,*i*-210,*i*-211,*i*-212)进行分类,其中前四个样本(*i*-105,*i*-106,*i*-107,*i*-108)的故障点直径为0.1778 mm,中间四个 样本(i-169,i-170,i-171,i-172)的故障点直径为0.355 6 mm, 后四个样本(i-209,i-210,i-211,i-212)的故障点直径 0.533 4 mm。结果显示,K-均值聚类算法将这12个样本按故障程度分 成了三类,分类结果与实际故障情况完全吻合。

表1 不同故障类型样本特征量

样本	特征量										
类型	A_{fn}	A_{2fn}	A_{fi}	$A_{2\!/\!i}$	$A_{_{3fi}}$	A_{fo}	$A_{_{2fo}}$	$A_{_{3fo}}$	A_{fr}	A_{2fr}	$A_{_{3fr}}$
<i>n-</i> 097	487.4	202. 8	14.20	3.800	14.90	7.800	9.700	11.60	3.600	7.600	3.000
n-098	202.9	153.5	4.300	2.500	3.600	16.80	15.80	5.900	6.300	7.100	3.000
n-099	193. 3	92.20	1.100	10.00	13.30	11.70	4.500	5.900	9.400	14.40	3.600
<i>n</i> -100	151.8	53.70	7.300	9.200	5.800	13.10	0.700	10.70	14.80	4.500	11.10
i-105	250.3	924.1	2111	917.6	464.6	34.30	27.90	41.20	56.10	186.6	6.300
i-106	316.0	1138	2678	1095	464.8	54.50	20.70	69.00	2.900	37.80	21.60
i-107	416.5	1419	3042	788.1	441.1	54.50	23.90	18.40	24.90	9.700	15.40
i-108	653.9	1043	3493	1539	450.8	20.80	5.400	15.90	24.80	65.70	29.80
i-169	1273	602.6	1010	640. 2	160.3	59.20	20.60	24.90	16.80	51.90	5.900
i-170	829.2	460.1	1312	531.2	128.5	33.40	30. 20	8.100	31.30	97.20	29.80
i-171	848.3	529.6	880. 9	501.2	87.40	8.900	36.50	3.400	26.70	61.30	7.700
i-172	891.2	728.4	1008	438.4	152.4	82.50	71.70	36.80	16.70	37.90	7.300
i-209	4496	1030	5109	2528	1861	119.0	78.10	98.20	26.40	152.3	8.300
i-210	2476	727.6	3784	2087	1676	55.80	15.90	72.00	67.90	114.7	54.60
i-211	3101	998.8	5277	2793	1792	11.90	30.70	65.70	22.80	70.60	14.90
i-212	2694	949.6	4355	2821	1536	92.60	62.40	44.20	65.50	303.4	13.20
o-130	1800	859.3	34.30	105.4	9.700	10720	6029	2842	69.70	32.50	66.30
r-3006	6084	8109	386. 8	594. 7	417.2	1025	312.7	387.2	5879	2318	1375

6 结束语

介绍了基于 EMD 和优化 K-均值聚类算法的滚动轴承故

障诊断和故障程度识别新方法。通过对滚动轴承正常、内环故 障、外环故障和滚动体故障实验信号的分析结果表明:经验模 态分解可去除信号噪声,提高信噪比,从而凸显故障特征;提取 的特征量可准确表征故障类型和故障程度;优化 K-均值聚类 算法可正确地将含不同故障类型的样本集进行正确分类;在对 含同一类型故障、不同故障程度的样本集分类中,优化 K-均值 聚类算法正确地完成了不同故障程度的分类。因此,基于 EMD 和优化 K-均值聚类的滚动轴承故障诊断方法,可在准确 提取故障特征的基础上,有效识别滚动轴承的故障类型,也可 正确识别同种故障类型、不同故障程度的样本。文中所提出的 方法是以故障特征频率处的幅值为特征量,在以后的研究工作 中可在特征量的选取方面作进一步的有益探索。

参考文献:

- [1] EREN L, DEVANEY M J. Bearing damage detection via wavelet packet decomposition of the stator current [J]. IEEE Instrumentation and Measurement, 2004, 53(2):431-436.
- [2] DEVANEY M, EREN L. Detecting motor bearing faults [J]. IEEE Instrumentation and Measurement, 2004, 7(4): 30-50.
- YU De-jie, CHENG Jun-sheng, YANG Yu. Application of EMD method and Hilbert spectrum to the fault diagnosis of rolling bearings
 [J]. Mechanical Systems and Signal Processing, 2005, 19(2): 259-270.
- [4] CHENG Jun-sheng, YU De-jie, YANG Yu. A fault diagnosis approach for rolling bearings based on EMD method and AR model [J]. Mechanical Systems and Signal Processing,2006,20(2):350-362.
- [5] 蔡艳平,李艾华,石林锁,等.基于 EMD 与谱峭度的滚动轴承故 障检测改进包络谱分析[J].振动与冲击,2011,30(2):167-172.
- [6] 窦东阳,赵英凯. 基于 EMD 和 Lempel-Ziv 指标的滚动轴承损伤 程度识别研究[J]. 振动与冲击,2011,29(2):5-8.
- [7] 李辉,郑海起,杨绍普.基于 EMD 和 Teager 能量算子的轴承故障 诊断研究[J]. 振动与冲击,2008,27(10):15-18.
- [8] YAN Ru-qiang, GAO R X. A tour of the Hilbert-Huang transform: an empirical tool for signal analysis[J]. IEEE Instrumentation and Measurement Magazine,2007,10(5):40-45.
- [9] 杨宇,于德介,程军圣.基于 EMD 与神经网络的滚动轴承故障 诊断方法[J]. 振动与冲击,2005,24(1):85-88.
- [10] 沈志熙, 黄席樾, 马笑潇. 基于 EMD 和支持向量机的柴油机故 障诊断[J]. 振动、测试与诊断, 2010, 30(1): 19-22.
- [11] The Case Western Reserve University. Bearing data center seeded fault test data [EB/OL]. http://casegroups.case.edu/bearing datacenter/home/.
- HUANG N E, SHEN Zheng, LONG S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis [J]. Proceeding of Royal Society A, Mathematical physical and Engineering Sciences, 1998, 454 (1971): 903-995.
- [13] KANG S, RYU J, LEE J, et al. Analysis of spacetimennd adaptive processing performance using K-means clustering algorithm for normalisation method in non-homogeneity detector process [J]. IET Signal Processing, 2011,5(2): 113-120.
- [14] LEE J W, PARK R H, CHANG S. Local tone mapping using the Kmeans algorithm and automatic gamma setting [J]. IEEE Transactions on Consumer Electronics, 2011, 57(1):209-217.