

Roller Bearing Fault Type Identification Based on LMD and Logistic Regression

Jingjing Wang, Yonghe Wei, Ruizhi Feng, Chao Wei

School of Mechanical Engineering, Shenyang Ligong University, Shenyang Liaoning
Email: yonghewe@sina.com, 1252808696@qq.com

Received: Jun. 9th, 2016; accepted: Jun. 26th, 2016; published: Jun. 30th, 2016

Copyright © 2016 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Aiming at the nonlinear and non-stationary vibration signal of the rolling bearing, a method based on local mean decomposition (Local Mean Decomposition, LMD) and logistic regression is proposed. This method processed collected vibration signals of rolling bearing inner ring and outer ring by LMD method, then selected the parameter of the model by genetic algorithm (GA) combined with logistic regression, and finally trained and tested the parameter by logistic regression. The result shows that the method can be effectively applied in roller bearing fault type identification.

Keywords

Local Mean Decomposition (LMD), Logistic Regression, Roller Bearing, GA, Fault Type Identification

基于LMD和逻辑回归的滚动轴承故障状态类型识别

王晶晶, 魏永合, 冯睿智, 魏超

沈阳理工大学, 机械工程学院, 辽宁 沈阳
Email: yonghewe@sina.com, 1252808696@qq.com

收稿日期: 2016年6月9日; 录用日期: 2016年6月26日; 发布日期: 2016年6月30日

摘要

针对滚动轴承的故障振动信号的非线性非平稳性,提出了一种基于局部均值分解(Local Mean Decomposition, LMD)方法和逻辑回归(LR)的滚动轴承故障诊断方法。该方法将采集到的滚动轴承内圈、外圈振动信号进行LMD方法处理后,采用遗传算法(GA)和逻辑回归结合进行模型中的参数选择,通过逻辑回归进行训练和测试,结果表明该方法可以有效地对滚动轴承故障类型进行识别。

关键词

局部均值分解(LMD), 逻辑回归, 滚动轴承, 遗传算法(GA), 故障状态识别

1. 引言

随着科学技术的快速发展,机械设备的功能变得越来越强大,越来越高效,变得越来越智能化。人们越来越依赖现代的应用设备,一旦设备发生故障,导致生产效率低,影响产品质量,造成不同程度的经济损失或导致重大事故。滚动轴承是旋转机械的重要组成部分,为监测和故障诊断具有重要意义。据统计,旋转机械中发生故障约有30%是由滚动轴承故障引起的。因此,对于滚动轴承的故障状态类型识别已经越来越重视。逻辑回归在许多科研领域尤其是医学上得到了广泛研究,主要用来评估人体健康状况[1]。机械设备运行过程中振动信号包含丰富的故障信息,然后进行信息的分析和处理,获得机械零件状态信息,进而判断是否有故障判断机械设备或组件。在滚动轴承故障或不正常工作,其振动信号是一种典型的非线性和非平稳信号,因此有必要采取故障诊断过程中适用于处理非线性和非平稳信号的特征提取方法。局部平均分解(LMD) [2]方法是基于信号本身和自适应分解,得到每个组件的PF分量具有一定的物理意义,反映了信号的本质,具有很高的信噪比,非常适用于线性、非平稳信号处理。因此,我们可以采用LMD方法提取故障信息作为特征值,然后采用逻辑回归进行滚动轴承故障状态识别。

随着信号检测技术、数字信号处理技术以及人工智能技术的迅速发展,对于滚动轴承的故障诊断已发展成为集物理、数字信号处理以及人工智能等方面技术于一体的综合性课题。近些年来国内外大量专家对特征选择方法作了深入地研究,出现了各种各样的方法:如,主成分分析法(PCA)、神经网络法、无监督聚类法、粗糙集理论法、遗传算法、基于特征相关性和冗余性分析的特征选择方法、基于动态规划方法的特征选择法等。其中,遗传算法因其简单,鲁棒性强,适用于并行处理,已广泛应用于计算机科学、优化调度、运输问题、组合优化等领域,也被广泛应用于特征选择,并取得了较好的结果。因此,采用遗传算法对逻辑回归模型的参数选择[3]。在国内,北京化工大学的李峰峰采用 Logistic 回归模型的机械状态健康评估研究,取得了较好的效果,湖南大学的史美丽基于 LMD 的滚动轴承故障诊断研究,取得了良好的效果。

2. 局部均值分解(LMD)的基本理论

局域均值分解(Local mean decomposition, LMD)是一种自适应分解方法,可以将一个复杂的多分量信号分解为若干个瞬时具有物理意义的乘积函数(Product function, 简称 PF)分量之和,每个 PF 分量均为包络信号和纯调频信号相乘而得到,所有 PF 分量的瞬时幅值和瞬时频率组合,便可以得到完整的原始信号时频分布,然后对其进行分析可以更准确有效的把握原始信号特征信息[4]。当滚动轴承发生故障影响设备运行时,损伤的滚动轴承元件在转动过程中产生脉动时引起的高频冲击振动会激起各个元件的固有频

率振动,同时,各个固有频率振动的幅值和频率还会受到脉动激发力的调制,而每个机械元件又有各自的固有频率,这说明滚动轴承故障振动信号具有多载波多调制的特性,是一种多分量的复杂的调幅-调频信号[5]。LMD方法可以将复杂的多分量调幅-调频信号自适应分解成单分量的调幅-调频信号,分解后得到的PF分量很好的保持了原始数据的幅值和频率变换,因此可以采用LMD方法对滚动轴承故障振动信号进行处理分析来提取主要的故障特征[6]。

如下个图所示:图1为轴承信号LMD分解图。

3. 遗传算法和逻辑回归模型

3.1. 逻辑回归的基本原理

逻辑回归[7](Logistic Regression, LR)是一种多元统计分析方法,它主要适用于描述一组自变量与具有二分性质的响应变量之间的一种最佳映射关系。其中,响应变量是二分类变量(例如发生或者不发生,通常用0或者1表示),而自变量可以是二分类的、连续性的、离散的或者是三者的混合。自变量与响应变量的一类(如取正常值)的概率之间的关系呈S型曲线。对于任意自变量X,在其它自变量不变的情况下,随着它的取值的不断增大,响应变量的概率P也会增大,一开始变化较慢,然后加速,最后又趋于平缓,但始终不会高于1[1]。

逻辑回归模型的数学表达式如公式(1)所示[1]:

$$P = P_i(y_i = 1 | X_1 = x_1, \dots, X_k = x_k) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}} \quad (1)$$

式中, $x = (x_1, x_2, \dots, x_k)^T$ ——自变量, k 维输入值; y_i ——实际观察到的响应变量; $y_i = 1$ ——事件发生; $y_i = 0$ ——事件未发生; P_i ——第 i 个事件发生的概率(取值在 0 和 1 之间); β_0 ——回归截距; β_1, \dots, β_k ——回归系数[8]。

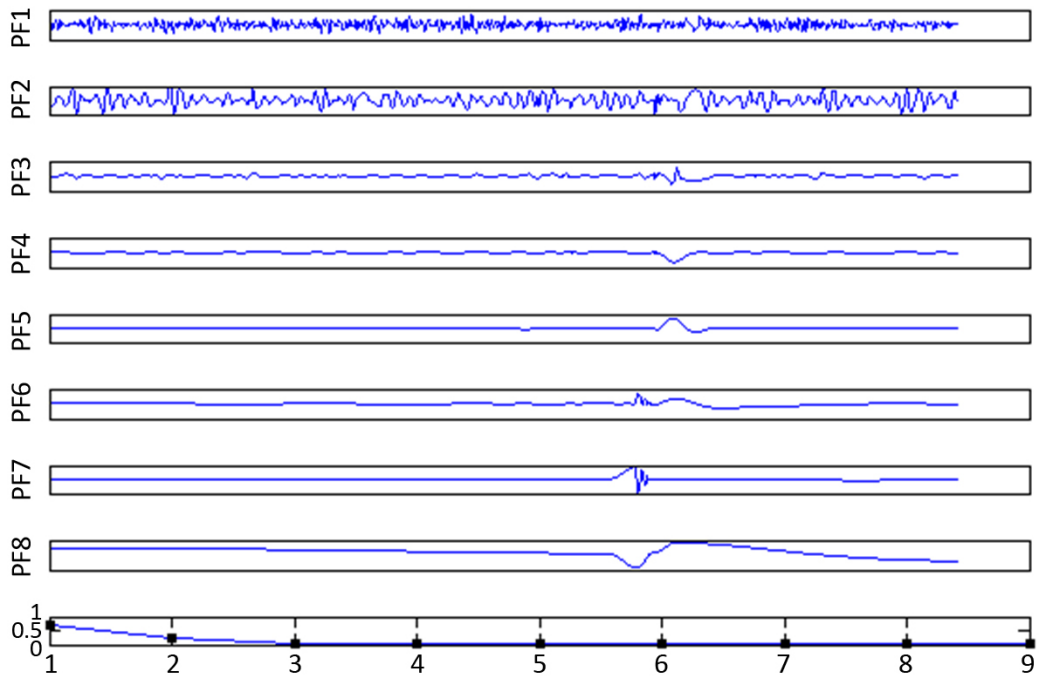


Figure 1. LMD decomposition of bearing signal

图 1. 轴承信号 LMD 分解图

3.2. 遗传算法原理

遗传算法[9]是一种重要的优化算法，它以广泛的适应性越来越多地受到人们的青睐和重视。遗传算法是模拟生物群体的遗传和长期进化过程建立起来的一类搜索和优化算法，它模拟了生物界“生存竞争，优胜劣汰，适者生存”的机制，用逐次迭代法搜索寻优。按照一定的适配值函数及一系列遗传操作对各个体进行筛选，从而使适配值高的个体被保留下来，组成新的群体，新群体中各个体适应度不断提高，直到满足一定的极限条件。此时，群体中适配值最高的个体即为待优化参数的最优解，其优点为从一群点搜索，而不是从一点开始搜索，能快速全局收敛[10]。

遗传算法流程图如图 2 所示。

3.3. 适应度函数的设计

适应度函数的设计主要满足以下条件：

(1) 单值、连续、非负、最大化

适应度函数 $Fit(f(x))$ 应该是实函数，并且单值、连续，但不要求可导，不过， $Fit(f(x))$ 的曲线在重要部分，特别在最优解附近一般不宜太陡也不宜过于平缓[9]。

(2) 合理、一致性

是指适应度函数曲线上，各点的适应度值应与解的优劣成反比例，即

$x_1, x_2, (x_1, x_2) \in [l \min, l \max] \wedge f(x_1) < f(x_2) \rightarrow Fit(f(x_1)) > Fit(f(x_2))$ ，其中 $[l \min, l \max]$ 是函数 $f(x)$ 的定义域[9]。

(3) 计算量小

$Fit(f(x))$ 不应设计的过于繁复，应在上述条件下越简单越好[9]。

(4) 通用性强

一个适应度函数的好坏，还应满足尽可能广泛的通用性，使用户在求解种种问题时，最好无需改变适应度函数中的参数。通用性要求是对适应度函数设计的更高一层的要求。它能使用户在对所求解函数的全局最优解的性质完全“无知”的情况下，由算法在运行过程中自动修正其中的参数值，从而一步一步接近最优解[9]。

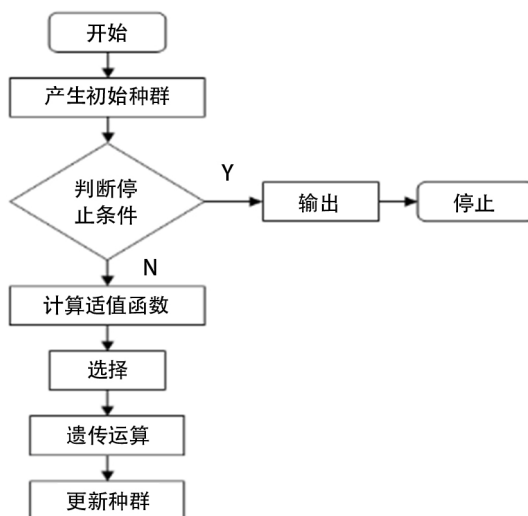


Figure 2. The flow chart of genetic algorithm
图 2. 遗传算法流程图

经过对以上各点的总结设计特征提取的适应度函数表达式如下：

$$Fit(f(x)) = \begin{cases} 1 - 0.5 * \left[\frac{f(x) - b}{a} \right]^\alpha, & |f(x) - b| < a \\ \frac{1}{1 + \left[\frac{f(x) - b}{a} \right]^\beta}, & |f(x) - b| \geq a \end{cases} \quad (2)$$

理想情况下： b 的值是 $\min f(x) = y^*$ ，当适应度值为 0.5 时， α 是 $f(x)$ 到 $\min f(x)$ 的距离，本文将 β 值取为 2，将 α 值分别取 0.5。这里的群体大小，即群体中所含个体的数量 N 为 20，遗传运算的终止进化代数 T 为 100，交叉概率 P_c 为 0.9，变异概率 P_m 为 0.005。

3.4. 基于 GA 和逻辑回归模型的特征选择和参数优化

基于 GA 和逻辑回归的滚动轴承状态健康评估的具体过程如图 3 所示。

- (1) 分别在正常状态、外圈故障状态和内圈故障状态下采集原始振动信号数据，经过降噪、滤波、LMD 处理后，根据遗传算法和逻辑回归结合选取特征值作为特征向量，建立逻辑回归模型的训练样本；
- (2) 利用正常状态和每个故障状态对模型进行训练，得到所需要的逻辑回归模型参数；
- (3) 在测试状态下采集数据，采用(1)中相同的方法得到测试样本；
- (4) 将测试样本代入逻辑回归模型中，得到性能评估指数——CV (Confidential Value)，它在数值上与公式(1)有关，并以此为根据进行故障状态识别；
- (5) 将利用逻辑回归模型计算的 CV 值评估结果与实际状态进行对比，从而验证模型的有效性。

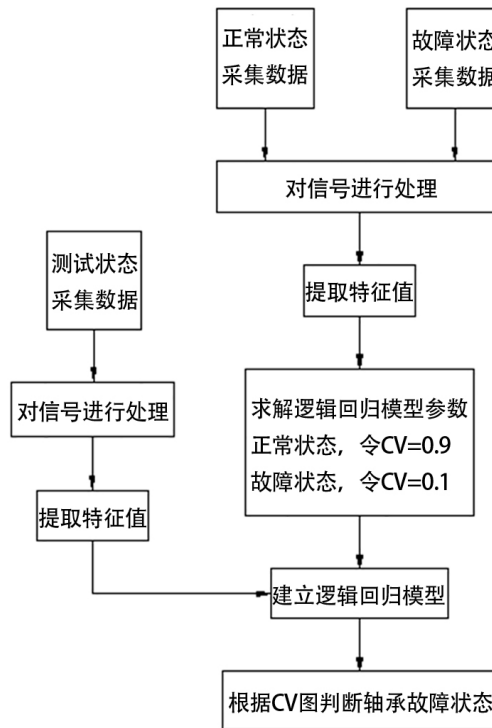


Figure 3. Equipment state recognition based on logistic regression model

图 3. 基于逻辑回归模型的设备状态识别

4. 应用实例分析

实验所用轴承破坏性数据来自于美国凯斯西储大学(Case Western Reserve University, CWRU)轴承实验室标准数据库[1]。轴承实验模拟装置如图4所示,轴承实验模拟装置由功率为1.5 kW的3相感应电机、扭矩传感器/译码器、电器控制装置和测力计组成,电动机带动输入轴,输入轴转速1750 r/min,由输出轴带动负载,本实验所用的轴承型号为SKF6205-2RS,基本几何参数如下表1。试验时分别在电动机端轴承的内圈、外圈和滚动体上使用电火花方法人为的加工了直径为0.021英寸的故障。将有故障的轴承重新装到电机上进行测试和记录振动数据,采样频率为12,000 HZ。根据轴承的参数以及故障频率公式求得滚动轴承理论的内圈故障频率为157.94 HZ,外圈故障频率为104.57 HZ。

将经过LMD方法处理后的滚动轴承故障振动信号分别求出7个有量纲时域参数,5个无量纲参数,还有9个PF能量值[11],统计特征分别为平均值、方差值、方根幅值、均方根值、峰值、偏度值、峭度值、波形指标、脉冲指标、裕度指标、偏斜度指标、峭度指标、PF分量的能量值。

分别对正常、外圈故障、内圈故障和滚动体故障进行分组,每种状态分为100组,每组包含1024个数据,用每组状态的前50组训练逻辑回归模型,后50组作为测试数据。通过遗传算法和逻辑回归结合选出4个特征值,分别为频带上3,5,9的频带上的能量和脉冲指标。令正常的CV值为0.9,令故障的CV值为0.1,建立逻辑回归模型,训练出的模型参数如表2所示。

前50组数据为正常数据,后50组数据分别为轴承外圈、轴承内圈的故障数据。代入逻辑回归模型,得到CV值如图5、图6所示。

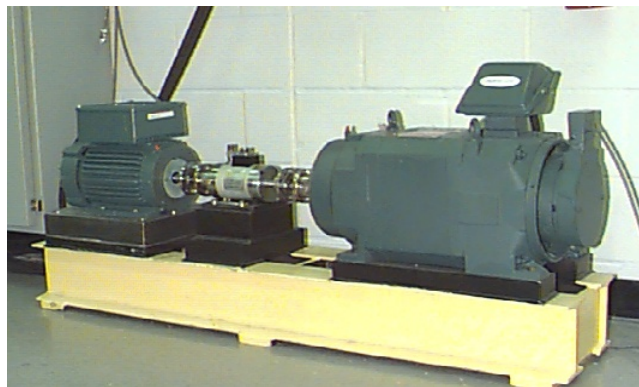


Figure 4. Rolling ball bearing experiment device
图4. 滚动球轴承实验装置

Table 1. Geometric parameter table of 6205-2RS deep groove ball bearing
表1. 6205-2RS 深沟球轴承几何参数表

轴承型号	外径(mm)	内径(mm)	节径(mm)	接触角	滚珠个数	滚动体直径(mm)
6205-2RS	52	25	39	0	9	7.938

Table 2. Logistic regression parameter
表2. 逻辑回归参数

系数	α	β_1	β_2	β_3	β_4
轴承内圈故障	1.1363	0.0088	0.0025	-0.0001	0.0004
轴承外圈故障	2.7765	0.0001	0.0002	-0.0003	-0.2119

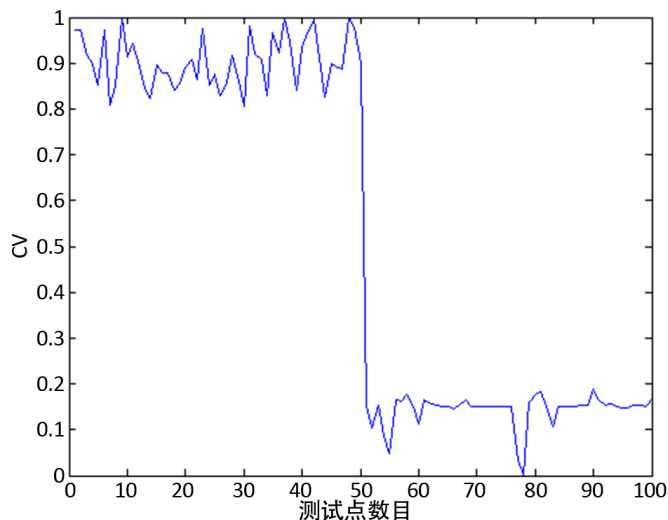


Figure 5. CV graph of normal and outer ring fault
图 5. 正常与外圈故障 CV 图

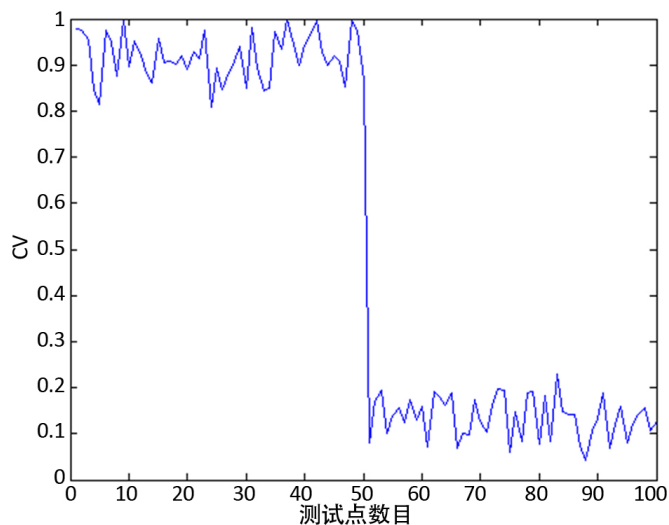


Figure 6. CV graph of normal and inner ring fault
图 6. 正常与内圈故障 CV 图

根据图 5、图 6 所示，正常轴承经评估后得到的 CV 值在 0.9 附近，而有故障的轴承的 CV 值在 0.1 附近，参考以前应用实例可以得到基于逻辑回归进行滚动轴承故障状态识别比较准确。

5. 结论

LMD 是一种自适应的分析方法，可以处理非线性、非平稳信号，分析所得的 PF 分量可以反映信号的局部特征，将分解后的 PF 分量进行能量提取，可以更好的反映轴承特征。逻辑回归模型算法分析设备运行状态与历史数据分布之间的关系，用设备当前数据与设备历史状态数据之间的差异相似性来评估设备状态的健康程度，提供了一种简单实用的设备状态健康评估方法。本文将 LMD 方法和逻辑回归方法相结合，应用到滚动轴承故障状态识别中。通过 LMD 方法分解得到 PF 分量，对 PF 分量进行能量分析，通过遗传算法和逻辑回归结合选取能量值和脉冲指标作为特征向量，通过逻辑回归训练、测试。结果表明，基于 LMD 和逻辑回归的方法能有效的应用于滚动轴承的故障状态识别中。

参考文献 (References)

- [1] 钟鑫, 刘文彬, 杨剑锋. 基于逻辑回归的滚动轴承性能退化评估[J]. 科技信息, 2010(16): 504-505.
- [2] 钟先友. 旋转机械故障诊断的时频分析方法及其应用研究[D]: [博士学位论文]. 武汉: 武汉科技大学, 2014.
- [3] 陈果, 邓堰. 遗传算法特征选取中的几种适应度函数构造新方法及其应用[J]. 机械科学与技术, 2011, 30(1): 124-132.
- [4] 史美丽. 基于 LMD 的滚动轴承故障诊断研究[D]: [硕士学位论文]. 长沙: 湖南大学, 2011.
- [5] Ma, J., Wu, J.D., Fan, Y.G., Wang, X.D. and Yan, X.G. (2015) The Rolling Bearing Fault Feature Extraction Based on the LMD and Envelope Demodulation. *Mathematical Problems in Engineering*, **2015**, Article ID: 429185. <http://dx.doi.org/10.1155/2015/429185>
- [6] 张亢. 局部均值分解方法及其在旋转机械故障诊断中的应用研究[D]: [博士学位论文]. 长沙: 湖南大学, 2012.
- [7] Pandya, D.H., Upadhyay, S.H. and Harsha, S.P. (2014) Fault Diagnosis of Rolling Element Bearing by Using Multi-nomial Logistic Regression and Wavelet Packet Transform. *Soft Computing*, **18**, 255-265. <http://dx.doi.org/10.1007/s00500-013-1055-1>
- [8] 李峰峰. 基于 Logistic 回归模型的旋转机械状态健康评估研究[J]. 维修与管理, 2009(5): 20-23.
- [9] 刘英. 遗传算法中适应度函数的研究[J]. 兰州工业高等专科学校学报, 2006, 13(3): 1-4.
- [10] 周培毅, 张新燕, 张华中. 基于遗传算法与 BP 神经网络的风力发电机齿轮箱故障诊断研究[J]. 华北电力技术, 2010(7): 6-11.
- [11] 崔春英, 段礼祥, 张来斌. 基于 LMD 和 FCM 的滚动轴承故障诊断方法[J]. 科学技术与工程, 2013, 13(7): 1764-1767.

再次投稿您将享受以下服务:

1. 投稿前咨询服务 (QQ、微信、邮箱皆可)
2. 为您匹配最合适的期刊
3. 24 小时以内解答您的所有疑问
4. 友好的在线投稿界面
5. 专业的同行评审
6. 知网检索
7. 全网络覆盖式推广您的研究

投稿请点击: <http://www.hanspub.org/Submission.aspx>