

基于主成分分析的多源特征融合故障诊断方法

巫茜¹ 蔡海尼² 黄丽丰¹

(重庆理工大学计算机科学与工程学院 重庆 400050)¹ (重庆大学软件学院 重庆 400044)²

摘要 目前故障诊断的实际应用中,因噪声的干扰,基于单传感器的故障诊断稳定性较差,很难达到满意的诊断精确度。提出了一种多传感器多特征数据融合的故障诊断方法。该方法利用多传感器从不同部位获取同一部件的运行状况,并通过构建多源特征融合模型,提高特征信息的抗干扰性,最后通过融合特征信息来完成部件的故障诊断。在将新方法应用于滚动轴承故障诊断的试验中,可以看到新方法能够获得较好的性能,比基于单传感器故障诊断的精确度更高。

关键词 多源信息融合,故障诊断,主成分分析,特征级,分类
中图分类号 TH212, TH213.3 **文献标识码** A

Feature-level Fusion Fault Diagnosis Based on PCA

WU Qian¹ CAI Hai-ni² HUANG Li-feng¹

(College of Computer Science & Engineering, Chongqing University of Technology, Chongqing 400050, China)¹

(College of Software Engineering, Chongqing University, Chongqing 400044, China)²

Abstract In most of current application, fault diagnosis based on single sensor show bad performance and is difficult to achieve satisfactory accuracy owing to noises. In this paper, a multi-sensor multi-feature fusion fault diagnosis method was proposed. The method firstly collects part operating conditions using multiple sensors installed on it. Then a fusion model based on principal component analysis was constructed to fuse all extracted features from the different sensors. Finally, fault diagnosis was carried out according to the fused results. Experiments and results show that new method can achieve perfect performance which is also better than that achieved by single-sensor.

Keywords Feature fusion, Fault diagnosis, PCA, Character level, Classifier

1 引言

机械系统故障诊断是保证生产安全的重要手段之一,一直吸引着相关领域研究者的关注。滚动轴承在机械系统中的使用十分普遍,而且较容易发生故障。这些故障不仅会降低机械系统的运行性能,而且会诱发停机事故。在旋转机械系统运行中,振动是必然产生的现象,振动信号包含了丰富的信息,能够直接反映系统不同的运行性能。因此,振动分析是当前状态监测和故障诊断最为广泛且最为有效的方法之一^[1,2]。随着计算机及传感器等相关技术的发展,实时获取并分析相对丰富的运行状态信息已经成为可能。传统故障诊断研究主要针对来自单个传感器信息源的信号进行分析处理。这些单个传感器可以同时采集多个系统部件的运行状态信息,但也因为多部件运行状态信号的叠加,影响了故障诊断的精确性。为了提高故障诊断的精确度,本文将基于多传感器多源振动信号信息对特征融合故障诊断进行研究。

目前,数据融合方法越来越受到各个领域专家的关注,并成功应用在了目标跟踪、模式识别和图像处理等领域^[3]。数据融合一般包括像素级融合、特征级融合和决策级融合。在

近年的研究报告中,决策级融合的应用研究较为常见,特征级融合的应用则很少受到研究者的关注。而特征分析是状态监测和故障诊断的关键步骤,基于特征的融合能够在消除冗余信息的同时,保留大量不相关的融合信息,这些融合特征综合了不同数据源信息,信息损失量比决策级融合少,更能为状态识别率和故障诊断准确率的提高提供有用信息。

基于以上因素,本文提出了一种基于主成分分析的多数据源多特征融合的故障诊断方法。该方法首先通过多传感器获取诊断部件运行状况信息,然后分别对各传感器获取的信号进行特征提取,再通过主成分分析(Principal Components Analysis, PCA)融合多源特征信息,并同时完成特征量降维处理,最后利用最近中心(Nearest Center, CNN)分类器进行故障类型识别,完成故障诊断。实际试验表明,该特征级融合诊断模型能够获取更有效的故障特征信息,可以大大提高故障诊断的精确度。

2 振动信号特征参数提取

旋转机械系统内的滚动轴承出现异常时,会直接反映到振动信号上,导致信号幅值域上特征参数的改变。所以,为了

到稿日期:2010-02-01 返修日期:2010-05-09 本文受重庆自然科学基金项目(2008BB2065),重庆理工大学青年基金项目(2010ZQ21)资助。
巫茜(1980-),女,硕士,讲师,主要研究方向为信息分析、企业信息化, E-mail: wuqian80@163.com;蔡海尼(1972-),女,博士生,副教授,主要研究方向为软件工程和软件架构。

监测滚动轴承的运行状态并及时有效地对其进行故障诊断,振动信号幅值域参数和无量纲指标,包括偏度系数、峭度系数、峰值因数、裕度因数、波形因数和脉冲因数,被提取出来进行分析。设振动采样信号 $x(t)$ 的长度为 n ,则在机械系统监测和诊断中常用的时域特征参数如下:

$$\text{平均幅值 } p_1 = \frac{1}{n} \sum_{i=1}^n x(i) \quad (1)$$

$$\text{标准差值 } p_2 = \left(\frac{\sum_{i=1}^n (x(i) - p_1)^2}{n-1} \right)^{\frac{1}{2}} \quad (2)$$

$$\text{均方根值 } p_3 = \left(\frac{1}{n} \sum_{i=1}^n x(i)^2 \right)^{\frac{1}{2}} \quad (3)$$

$$\text{方根幅值 } p_4 = \left(\frac{1}{n} \sum_{i=1}^n \sqrt{|x(i)|} \right)^2 \quad (4)$$

$$\text{峰值 } p_5 = \max |x(i)| \quad (5)$$

$$\text{偏度系数 } p_6 = \frac{\sum_{i=1}^n (x(i) - p_1)^3}{(n-1)p_2^3} \quad (6)$$

$$\text{峭度系数 } p_7 = \frac{\sum_{i=1}^n (x(i) - p_1)^4}{(n-1)p_2^4} \quad (7)$$

$$\text{峰值因数 } p_8 = \frac{p_5}{p_3} \quad (8)$$

$$\text{裕度因数 } p_9 = \frac{p_5}{p_4} \quad (9)$$

$$\text{波形因数 } p_{10} = \frac{p_3}{\frac{1}{n} \sum_{i=1}^n |x(i)|} \quad (10)$$

$$\text{脉冲因数 } p_{11} = \frac{p_5}{\frac{1}{n} \sum_{i=1}^n |x(i)|} \quad (11)$$

这 11 个特征参数中, p_1 和 $p_3 - p_5$ 反映振动信号在时域上的振幅和能量,而 p_2 和 $p_6 - p_{11}$ 则反映振动信号在时域上的时间序列分布情况。当滚动轴承发生故障时,旋转机械系统振动信号的时域信息将发生变化,无论是幅度信息还是分布情况都可能与正常运行状态下的情况不同。

3 特征融合故障诊断方法

3.1 主成分分析

主成分分析^[4]的主要目的是对包含大量相关信息的数据集进行降维处理,尽可能用最少量的信息进行表示,并保证信息损失较小。这些少量的用来重新组建新数据集的信息被称为主成分,它们之间是不相关的,且由大到小排列。假定 $X = (x_1, x_2, \dots, x_n)$ 是输入空间的数据集,其中, $x_i, i = 1, 2, \dots, n$ 是 m 维向量, n 是数据集中的总样本数。若存在一个映射函数 ϕ 将输入空间数据映射到一个高维空间,则可以得到高维特征空间的数据集 $\Phi(X) = (\phi(x_1), \phi(x_2), \dots, \phi(x_n))$,将该新数据集的协方差定义为:

$$C = \frac{1}{n} \sum_{i=1}^n \phi(x_i)^T \phi(x_i) \quad (12)$$

其符合特征方程:

$$Cv = \lambda v \quad (13)$$

式中, λ 和 v 分别对应协方差矩阵的特征值和特征向量。且由定义^[5]知,特征向量 v 可以用 $\Phi(X) = (\phi(x_1), \phi(x_2), \dots, \phi(x_n))$ 张程表示为:

$$v = \sum_{i=1}^n \alpha_i \phi(x_i) \quad (14)$$

式中, α_i 是张程系数。将式(12)和式(13)代入式(14),可以得

到:

$$K\alpha = m\lambda\alpha \quad (15)$$

式中, $K = \Phi(x)^T \Phi(x)$ 。求得张程系数 α 后,通过式(14)可以计算核主成分 v 。

上面的推导假定所有数据都具有零均值,如果不是,可以通过 $\tilde{K} = K - I_n K - K I_n + I_n K I_n$ 得到^[6],其中, $I_n = \left[\frac{1}{n} \right]_{n \times n}$ 。

3.2 最近中心分类法

最近中心分类器(Nearest Center Classifier, NC)^[7]按以下思想对测试数据进行分类:

设 $i(i = 1, 2, \dots, m)$ 表示训练数据样本中的一类样本, n 维特征向量 $X_j^i = (X_{1j}^i, X_{2j}^i, X_{3j}^i, \dots, X_{nj}^i)$ 表示类 i 中的第 j ($j = 1, 2, \dots, p$) 个训练样本,其中分量 X_{ij}^i 是对类 i 中第 j 个样本的属性 A_t ($t = 1, 2, \dots, n$) 的度量。

定义类 i 的中心 C^i 为该类样本的平均,即:

$$C^i = \sum_{j=1}^p X_j^i / p = \left(\sum_{j=1}^p X_{1j}^i / p, \sum_{j=1}^p X_{2j}^i / p, \dots, \sum_{j=1}^p X_{nj}^i / p \right) \quad (16)$$

式中, p 为类 i 中的样本个数。

又设 $X = (X_1, X_2, \dots, X_n)$ 为测试样本空间的一个样本,如果把样本集里的样本看成欧氏空间里的点,则可定义样本 X 到类 i 的中心 C^i 的距离为:

$$d^i = \| X - C^i \| \quad (17)$$

根据最近中心法思想,NC 分类器判决规则可定义为:如果对于训练样本空间的某个类 c ,存在以下等式成立: $d^c = \| X - C^c \|$,则判决 X 为来自于类 c 的样本。

3.3 特征融合故障诊断方法

故障诊断可以被视为模式识别过程,即对监控信息进行识别,并最终完成分类决策。目前常用的故障诊断方法主要基于单数据源信息,诊断性能存在一定的限制,容易受到干扰数据信息的影响。为了更好地提高分类性能,即提高故障诊断精确度,本文将基于多数据源信息进行分析,并提出一种新的基于特征级数据融合故障诊断方法。

多源特征故障诊断方法就是将多传感器源信号数据中提取的特征量进行融合,减少数据冗余,减少环境噪声的影响,提高特征量的故障诊断有效性。假定存在 S 个传感器在故障诊断系统中,那么第 s 个($s \in \{1, \dots, S\}$)传感器获取信号所提取的特征量为 P_s ,多源特征融合就是通过前文所述的主成分分析方法完成融合特征 P 求取: $P = \varphi(P_1, P_2, \dots, P_S)$ 。融合特征被求取之后,将被直接用来进行故障诊断。多源特征融合故障诊断方法的算法流程如下所示:

Step 1 通过多传感器从不同部位获取系统部件运行状况信号信息;

Step 2 分别对各个传感器获取的信号提取其相关特征信息;

Step 3 通过主成分分析构造融合函数,融合多数据源特征信息,获取融合特征量;

Step 4 利用 NC 分类器对融合特征量进行训练学习,并利用完成学习任务的 NC 分类器来进行实时故障诊断。

4 试验及其结果分析

本节将通过滚动轴承故障信号数据集来验证基于特征级

数据融合的故障诊断方法。这些故障信号数据从 Case Western Reserve University^[9] 获取。其实验平台为一个三相感应电动机连着一个发电机,滚动轴承是其中重要的部件。

采集滚动轴承运行状态振动信号的加速度传感器分别被安装在感应电动机的驱动端和风扇端。加速度传感器可以同时采集滚动轴承内圈、外圈和滚轴的运行状态信息。在本文实验中,考虑了 4 种不同负载运行状况(0,1,2 和 3 马力)下各故障类型 3 种不同故障程度(7,14,21 密耳)的故障诊断问题。首先构建了两个数据集进行试验分析,如表 1 所列。数据集 A 是简单的故障类型识别诊断问题,数据集 B 是一个进一步加大难度的故障诊断任务,该任务不仅要完成故障诊断,还要识别当前故障类型的故障程度。

表 1 数据集描述

数据集	训练样本数	测试样本数	运行状况	故障程度	类别
A	80	80	正常状态	0	1
	80	80	外圈故障	7 密耳	2
	80	80	内圈故障	7 密耳	3
	80	80	滚动体故障	7 密耳	4
B	80	80	正常状态	0	1
	80	80	外圈故障	7 密耳	2
	80	80	内圈故障	7 密耳	3
	80	80	滚动体故障	7 密耳	4
	80	80	外圈故障	14 密耳	5
	80	80	内圈故障	14 密耳	6
	80	80	滚动体故障	14 密耳	7
	80	80	外圈故障	21 密耳	8
80	80	内圈故障	21 密耳	9	
80	80	滚动体故障	21 密耳	10	

表 2 给出了试验的性能结果。从表 2 可以看出,本文提出的基于特征融合故障诊断模型可以获得满意的性能,即使是在多故障类型多故障程度的故障诊断任务中。

表 2 多源特征融合方法故障诊断性能

诊断方法	数据集 A	数据集 B
本文方法	100%	100%

为了更好地分析基于特征融合故障诊断对性能所带来的影响,本文分别对各传感器的信号进行特征提取,然后利用 NC 分类器进行模式识别。同时,也将全部传感器特征信息进行简单联合(求平均)后直接利用 NC 分类器进行故障诊断,省去了 PCA 进行融合函数构造的过程。在这个试验中,我们获得了如表 3 所列的试验结果。从表 3 可以看出,通过多传感器特征数据融合,能够获取更多有效的决策信息,提高了决策精确度,比基于单个传感器数据源的故障诊断准确度更高。同时,通过表 3 和表 2 的性能比较,也可以发现,融合函数的构建以及构建的好坏直接影响到融合特征量对故障诊断带来的影响。表 3 中对多源特征量进行简单联合所获得的性能远远低于表 2 中通过主成分分析构造融合函数所获得的性能,特别是在数据集 B 中。

表 3 故障诊断性能

诊断方法	数据集 A	数据集 B
驱动端传感器+NC 分类器	76.88%	64.75%
风扇端传感器+NC 分类器	65.63%	38.5%
多源特征融合+NC 分类器	80%	69%

为了进一步分析基于多源特征融合的性能,将本文方法和决策级融合中最大值方法、平均值方法进行了比较,比较结果如图 1 所示。从图 1 可以看出,基于多源特征值的融合,能够更有效地排除干扰,提高精确度。

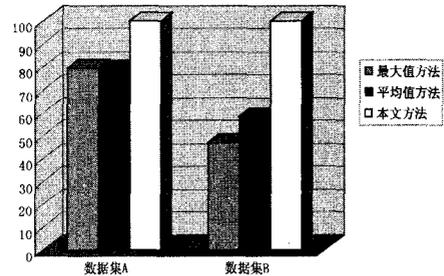


图 1 本文方法与决策级融合方法的性能比较

结束语 本文讨论了一种基于主成分分析的多源特征融合故障诊断模型。该模型通过多源信息的获取及通过构造融合函数完成多源特征信息的融合来进行故障识别与诊断。通过实际滚动轴承故障数据试验的结果分析表明,本文方法能够大大提高故障诊断的精确度。多传感器采集信号的融合研究能够减少噪声对决策结果的影响,同时,也能够互相补充信息,避免单传感器性能低造成的影响,能更好地获取故障特征信息,进一步提高决策的精确度。

参考文献

- [1] Widodo A, Yang B S, Han T. Combination of independent component analysis and support vector machines for intelligent faults diagnosis of induction motors[J]. Expert Systems with Applications, 2007, 32(2): 239-312
- [2] Lei Ya-guo, He Zheng-jia, Zi Yan-Yang. A new approach to intelligent fault diagnosis of rotating machinery[J]. Expert Systems with Applications, 2008, 35(4): 1593-1600
- [3] Niu Gang, Yang Bo-suk. Intelligent condition monitoring and prognostics system based on data-fusion strategy[J]. Expert Systems with Application, 2010, 37(12): 8831-8840
- [4] Shlens J. A Tutorial on Principal Component Analysis[R]. CA 92037. Systems Neurobiology Laboratory, Salk Institute for Biological Studies, La Jolla, April 2009
- [5] Scholkopf B, Smola A. Learning with Kernel: Support Vector Machines, Regularization, Optimization and Beyond[M]. London: MIT Press, 2002
- [6] Scholkopf B, Smola A, Muller K R. Nonlinear component analysis as a kernel eigenvalue problem[J]. Neural Computation, 1998, 10(5): 1299-1319
- [7] Gao Qing-bin, Wang Zheng-zhi. Center-based nearest neighbor classifier[J]. Pattern Recognition, 2007, 40(1): 346-349
- [8] <http://www.eecs.case.edu/laboratory/bearing/download.htm>, 2010