

二维非负矩阵分解在齿轮故障诊断中的应用*

李兵¹, 米双山¹, 刘鹏远¹, 刘东升¹, 张培林²

(1. 军械工程学院四系 石家庄, 050003) (2. 军械工程学院一系 石家庄, 050003)

摘要 针对齿轮故障信号时频分布识别问题, 提出采用二维非负矩阵分解技术提取时频分布矩阵特征参数的方法。采用S变换技术将齿轮故障信号变换至时频域, 为克服传统的一维非负矩阵分解对矩阵向量化带来的维数过高和结构信息损失问题, 提出采用二维非负矩阵分解技术直接对信号时频分布矩阵提取特征参数。对齿轮5种状态下信号时频分布矩阵的特征提取和分类结果表明, 二维非负矩阵分解技术无论在计算效率还是分类精度上都明显优于一维非负矩阵分解技术。

关键词 齿轮; 故障诊断; 特征提取; 时频分布; 二维非负矩阵分解

中图分类号 TH165.3

引言

齿轮故障信号是典型的非平稳信号, 时频分析技术可将一维时域信号变换到二维时频域内进行分析, 是分析非平稳信号的有效手段, 已在齿轮故障诊断中获得了广泛的应用^[1-4]。然而, 由于时频分布矩阵维数巨大, 很难直接用于分类, 为节省存储空间, 降低计算复杂度, 必须对信号时频分布矩阵进行进一步的特征提取。

非负矩阵分解(NMF)算法是1999年由Lee等^[5]提出的特征提取方法。该算法实现简便, 分解的结果中不出现负值, 具有可解释性和明确的物理意义, 占用存储空间少, 已广泛用于信号处理、模式识别、计算机视觉和图像工程等研究领域^[6-8]。有学者将非负矩阵分解技术与时频分析相结合应用于发动机故障诊断^[9-10], 取得了较好的效果。

采用原始的非负矩阵分解技术提取二维矩阵特征时, 需要将二维矩阵展开成一维的向量才能进行处理, 这会带来两个问题: 一是二维矩阵转换成一维向量时, 向量维数很大, 带来巨大的计算量, 收敛速度非常慢; 二是将二维矩阵一维化的过程必然会损失原始矩阵中的结构信息, 最后分解所得到的基向量很难精确反映原始矩阵的结构信息^[11]。笔者提出了基于S变换和二维非负矩阵分解技术的齿轮故障诊断方法, 采集了齿轮在5种状态下的振动信号, 利用S变换技术将齿轮故障信号变换至时频域得到时

频分布矩阵, 分别采用一维非负矩阵分解技术和二维非负分解技术对时频分布矩阵进行特征提取, 对比分析两种方法在不同的特征维数下的计算效率及分类性能。

1 S变换

S变换理论是R. G. Stockwell^[12]于1996年提出的时频分析理论, 它融合了短时傅里叶变换和小波变换的优点, 是对连续小波变换和短时傅里叶变换的一种扩展。

设 $h(t)$ 为一维连续信号, 定义 $h(t)$ 的S变换为

$$S(\tau, f) = \int_{-\infty}^{+\infty} h(t) \omega(\tau - t, f) e^{-i2\pi ft} dt = \int_{-\infty}^{\infty} h(t) \frac{|f|}{\sqrt{2\pi}} e^{\left[-\frac{f^2(\tau-t)^2}{2}\right]} e^{-i2\pi ft} dt \quad (1)$$

其中: f 为分析频率; t 为时间; τ 为窗函数 $\omega(\tau - t, f)$ 中心点的位置, 控制高斯窗在时间轴上的位置。

S变换的精髓在于引入了宽度随频率反向变化的高斯窗, 从而在低频段采用较宽的窗以获得较高的频率分辨率, 在高频段采用较窄的窗以获得较高的时间分辨率。

2 二维非负矩阵分解

2.1 非负矩阵分解

非负矩阵分解的主要思想为: 已知非负矩阵 V ,

寻找适当的非负矩阵 \mathbf{W} 和 \mathbf{H} ,使得给定数据向量集合 $\mathbf{V}_{n \times m}$ 为

$$\mathbf{V}_{n \times m} \approx \mathbf{W}_{n \times r} \mathbf{H}_{r \times m} \quad (2)$$

其中: n 为数据样本的维数; m 为集合中数据样本的个数。

这个矩阵可以近似分解为矩阵 $\mathbf{W}_{n \times r}$ 和矩阵 $\mathbf{H}_{r \times m}$ 的乘积。一般情况下, r 的选择要满足 $(n+m)r < nm$,从而使 \mathbf{W} 和 \mathbf{H} 的维数小于原始矩阵 \mathbf{V} ,由此可压缩原始数据维数。

笔者以最小化修正的 Kullback-Laebler 散度为目标函数,给定非负矩阵 $\mathbf{V}_{n \times m}$,非负矩阵分解算法可转化为如下的带约束的优化问题^[13]

$$\begin{aligned} \text{minimize } D(\mathbf{V} \parallel \mathbf{WH}) &= \sum_{ij} \left(V_{ij} \log \frac{V_{ij}}{(\mathbf{WH})_{ij}} - \right. \\ &\left. V_{ij} + (\mathbf{WH})_{ij} \right) \text{ subject to } \mathbf{W}, \mathbf{H} \geq 0 \end{aligned} \quad (3)$$

对应于式(3)的乘性迭代规则表示如下,由此可得到基向量矩阵为

$$\mathbf{H}_{au} \leftarrow \mathbf{H}_{au} \frac{\sum_i \mathbf{W}_{ia} \mathbf{V}_{iu} / (\mathbf{WH})_{iu}}{\sum_k \mathbf{W}_{ka}} \quad (4)$$

$$\mathbf{W}_{ia} \leftarrow \mathbf{W}_{ia} \frac{\sum_u \mathbf{H}_{au} \mathbf{V}_{iu} / (\mathbf{WH})_{iu}}{\sum_v \mathbf{H}_{av}} \quad (5)$$

$$\mathbf{W}_{ia} = \mathbf{W}_{ia} / \sum_i \mathbf{W}_{ia} \quad (6)$$

在特征提取时,将原始向量向基向量矩阵投影得到的投影系数即为特征参数。设 \mathbf{X}_i 为样本向量,则表征 \mathbf{X}_i 的系数向量可由式(4)求得

$$\mathbf{H}_i = \mathbf{W}^T \mathbf{X}_i \quad (7)$$

2.2 二维非负矩阵分解方法

针对非负矩阵分解在提取二维矩阵特征参数时所存在的问题,有学者提出了二维非负矩阵分解(2DNMF)的概念^[14-15],与一维非负分解技术(1DNMF)不同,在提取二维矩阵特征时,2DNMF直接对矩阵进行操作,无需进行向量化,不但降低了计算量,而且保留了二维矩阵的结构信息。

二维非负矩阵分解包括矩阵行方向分解和列方向分解两大部分。

设 \mathbf{A} 表示非负矩阵集, \mathbf{A}_k ($k=1,2,\dots,m$) 表示第 k 个时频分布矩阵,其维数为 $p \times q$ 。当对矩阵进行行方向非负矩阵分解时,首先将 m 个矩阵级联为一个新的 $p \times qm$ 矩阵 $\mathbf{U} = [\mathbf{A}_1 \mathbf{A}_2 \dots \mathbf{A}_m]$,将 \mathbf{U} 分解为两个非负矩阵 \mathbf{L} 和 \mathbf{H} 的乘积为

$$\mathbf{U}_{p \times qm} = \mathbf{L}_{p \times d} \mathbf{H}_{d \times qm} \quad (8)$$

其中: \mathbf{L} 为行基矩阵; \mathbf{H} 为系数矩阵; d 为特征维数。

\mathbf{H} 可划分为 m 个 $d \times q$ 的矩阵, $\mathbf{H} = [\mathbf{H}_1 \mathbf{H}_2 \dots \mathbf{H}_m]$,则 \mathbf{H}_k 为矩阵 \mathbf{A}_k 的系数矩阵,即满足

$$\mathbf{A}_k = \mathbf{L} \mathbf{H}_k \quad (k=1,2,\dots,m) \quad (9)$$

同理,当对矩阵集 \mathbf{A} 进行列方向分解时,先将 m 个矩阵级联为一个新的 $q \times pm$ 矩阵 $\mathbf{V} = [\mathbf{A}_1^T \mathbf{A}_2^T \dots \mathbf{A}_m^T]$,将 \mathbf{V} 分解为两个非负矩阵 \mathbf{R} 和 \mathbf{H} 的乘积

$$\mathbf{V}_{q \times pm} = \mathbf{R}_{q \times g} \mathbf{H}_{g \times pm} \quad (10)$$

其中: \mathbf{R} 为列基矩阵; \mathbf{H} 为系数矩阵, g 为特征维数。

\mathbf{H} 可划分为 m 个 $g \times p$ 的矩阵, $\mathbf{H} = [\mathbf{H}_1 \mathbf{H}_2 \dots \mathbf{H}_m]$,则 \mathbf{H}_k 为矩阵 \mathbf{A}_k^T 的系数矩阵,即满足

$$\mathbf{A}_k^T = \mathbf{R} \mathbf{H}_k \quad (k=1,2,\dots,m) \quad (11)$$

矩阵 \mathbf{U} 和 \mathbf{V} 分别是根据训练样本的时频矩阵及其转置构成的,因此对它们的分解可以同时进行,分解方法与一维的非负矩阵分解方法一样,由此可分别得到时频矩阵集的行基矩阵 $\mathbf{L}_{p \times d}$ 和列基矩阵 $\mathbf{R}_{q \times g}$ 。

矩阵 \mathbf{A}_i 在 \mathbf{L} 和 \mathbf{R} 上的投影 \mathbf{C}_i 为

$$\mathbf{C}_i = \mathbf{L}^T \mathbf{A}_i \mathbf{R} \quad (12)$$

其中: \mathbf{C}_i 为表征矩阵 \mathbf{A}_i 的系数矩阵,即特征参数矩阵,其维数为 $d \times g$ 。

3 齿轮故障信号分析

3.1 齿轮故障信号描述

本研究的齿轮故障信号来自于一个二级传动齿轮箱,齿轮箱由两对齿轮副组成,齿数分别为 25,50 和 18,91,输入轴转速为 1 491 r/min,采样频率为 5 kHz,采样点数为 4 096。模拟的齿轮局部故障为齿根裂纹和齿面磨损,故障设置在中间轴齿轮和输出轴齿轮上。试验中采集了齿轮在 5 种状态下的振动信号,即正常、中间轴齿轮齿根裂纹、中间轴齿轮齿面磨损、输出轴齿轮齿根裂纹和输出轴齿轮齿面磨损。图 1 为齿轮在 5 种状态下的信号时域波形。

3.2 齿轮故障信号的 S 变换

图 2 为齿轮在 5 种状态下振动信号的 S 变换结果。

由图 2 可以看出,齿轮正常状态的时频分布主要以齿轮的啮合分量为主,能量主要集中在 1 阶和 2 阶啮合频率处,在时域上比较连续,而且整体能量较

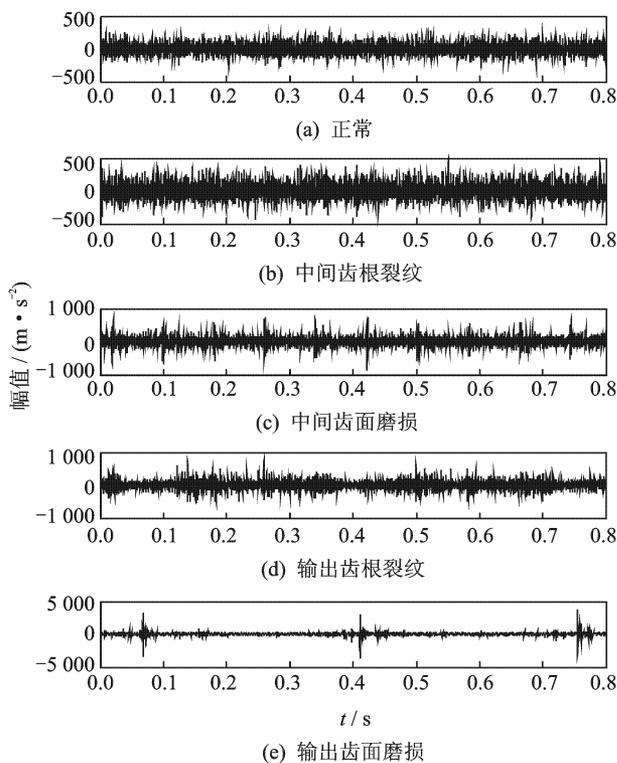


图1 齿轮在5种状态下的振动信号

低;当中间轴齿轮发生齿根裂纹故障时,信号的能量有所增加,主要能量依然集中在1阶和2阶啮合频率处,没有十分明显的冲击特征,但与正常状态相比,高频部分分量有所增加;当中间轴齿轮发生齿面磨损故障时,时频分布中出现了比较明显的周期性冲击成分,其频率为中间轴的转频,约为12 Hz;当输出轴齿轮发生齿根裂纹故障时,时频分布没有明显的冲击,但高频部分的能量有所增加;输出轴齿轮齿面磨损故障信号的冲击特征十分明显,能量几乎全部集中在高频部分。齿轮在5种状态下的S变换时频分布存在各自明显的特点,具有一定的可分性。

3.3 齿轮故障信号时频分布矩阵特征提取

由于时频矩阵的维数十分巨大,本研究中时频分布矩阵维数为 $2\ 048 \times 4\ 096$,所以计算量十分巨大,不利于进行特征参数提取。笔者采用二维小波变换信号的时频矩阵进行3层分解,取第3层的近似系数矩阵进行下一步分析,其维数为 256×512 。这样,在保留信号时频分布矩阵主要分量的前提下大大减小了时频分布矩阵的维数。

从每种齿轮状态下挑选5个时频分布矩阵作为训练样本集 $A = \{A_1, A_2, \dots, A_{25}\}$, $A_k (k=1, 2, \dots, 25)$ 为第 k 个时频分布矩阵。采用一维非负分解技术对

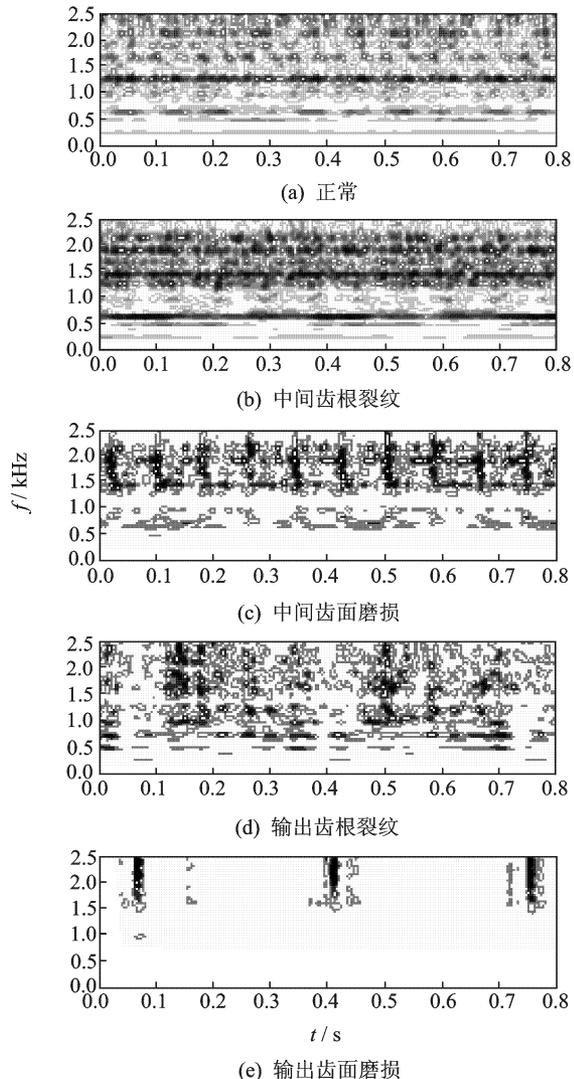


图2 齿轮5种状态下的S变换

时频矩阵进行分析时,要先将时频矩阵进行向量化,其维数为 $131\ 072$,得到训练集为 $V_{131\ 072 \times 25}$ 。

分别采用一维和二维非负矩阵分解技术对齿轮故障信号时频分布矩阵进行特征提取。分解时,迭代次数设置为200,二维非负矩阵分解的行基矩阵秩 d 和列基矩阵秩 g 设置为 $d=g$,取值范围为 $[2, 3, \dots, 9, 10, \dots]$ 。为进行对比分析,相应的一维非负矩阵分解的基向量秩 r 取值为 $[4, 9, 16, \dots, 81, 100]$,这样保证了两种方法提取特征维数的一致性。

表1给出了采用一维和二维非负矩阵分解技术对齿轮故障信号时频分布矩阵进行非负矩阵分解的计算时间。由表1可以发现,二维非负矩阵分解的计算效率要明显高于一维非负矩阵分解,而且随着特征维数的增加,这种优势也更加明显。

采用一维非负矩阵分解时,训练结束可得到基向量矩阵 $W_{131\ 072 \times r}$,将原始时频分布矩阵向量化后

向基向量矩阵投影(见式(4)),即得到描述该样本的特征参数。

表1 1DNMF和2DNMF的计算时间

特征维数	4	9	16	25	36
1DNMF	0.862 3	1.996 6	2.412 5	4.525 2	5.723 6
2DNMF	0.609 5	1.083 9	2.244 3	3.329 1	2.710 7
特征维数	49	64	81	100	
1DNMF	9.592 5	11.928 7	17.512 1	21.275 6	
2DNMF	3.935 3	4.946 3	5.229 8	6.430 9	

采用二维非负矩阵分解时,训练结束后可得到行基矩阵 $L_{256 \times d}$ 和列基矩阵 $R_{512 \times g}$,将原始样本矩阵向行基矩阵和列基矩阵投影(见式(9)),即可得到描述该样本的特征参数。

图3给出了时频分布矩阵训练集对应的编码矩阵,其维数为 10×10 。图中每一行代表一种齿轮的故障状态,每种状态下5个样本。

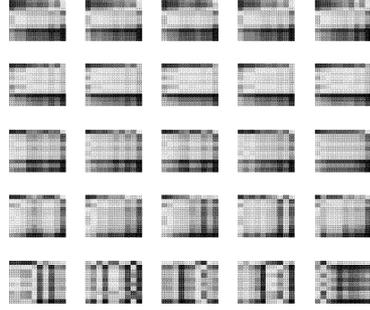


图3 二维非负矩阵分解提取出的齿轮时频矩阵特征参数

为测试两种方法提取的特征子集的故障诊断能力,采用朴素贝叶斯分类器(NBC)对齿轮5种状态进行分类。每种状态采集了20个样本,随机选择其中10个样本作为训练样本,其余10个样本作为测试样本。为保证结果的有效性,重复此过程20次,并取其平均值作为最终的结果。图4给出在不同的特征维数下,两种方法所提取的特征子集的分类精度。

由图4可以看出,在不同的特征维数下,二维非负矩阵分解技术所提取的特征子集分类精度始终优于一维非负矩阵分解技术提取的特征子集。同时也可以发现,并不是基向量秩的维数越高,分类性能越好。

由图4可以看出,当基向量秩为16时,即二维非负矩阵分解基矩阵秩和列基矩阵秩为4时,所提取的特征参数集取得了最优的分类精度。继续增加特征维数时,分类精度反而出现了下降的趋势。因此,

在采用非负矩阵分解提取特征时,特征维数的选择也是一个非常关键的问题。

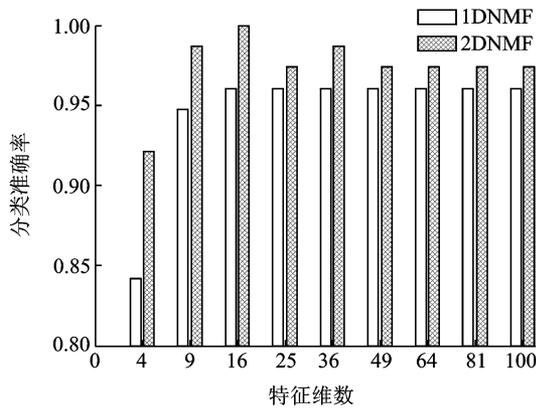


图4 两种特征提取方法的分类精度对比

4 结 论

1) S变换技术能够有效地反映齿轮在5种状态下信号的时频分布特点,可用于对齿轮故障状态的判别。

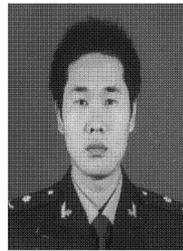
2) 与传统的一维非负矩阵分解技术相比,二维非负矩阵分解在提取时频分布矩阵特征时,无需对矩阵向量化,避免了分析维数过大带来的巨大计算量,因此具有更高的计算效率。

3) 二维非负矩阵分解避免了矩阵向量化带来的结构信息损失,对齿轮5种状态下的时频分布矩阵分类结果表明,二维非负矩阵分解提取的特征参数具有非常理想的故障诊断精度。

参 考 文 献

- [1] 姜鸣,陈进,汪慰军.几种Cohen类时频分布的比较及应用[J].机械工程学报,2003,39(8):129-134.
Jiang Ming, Chen Jin, Wang Weijun. Comparison and application of some time-frequency distributions belonging to Cohen class[J]. Chinese Journal of Mechanical Engineering, 2003, 39(08): 129-134. (in Chinese)
- [2] 林京,屈梁生.基于连续小波变换的信号检测技术与故障诊断[J].机械工程学报,2000(12):95-100.
Lin Jing, Qu Liangsheng. Feature detection and fault diagnosis based on continuous wavelet transform[J]. Chinese Journal of Mechanical Engineering, 2000 (12):95-100. (in Chinese)
- [3] Meltzer G, Ivanov Y Y. Fault detection in gear drives with non-stationary rotational speed—part I: the

- time-frequency approach[J]. *Mechanical Systems and Signal Processing*, 2003,17(5):1033-1047.
- [4] Oehlmann H, Brie D, Tomczak M, et al. A method for analysing gearbox faults using time-frequency representations[J]. *Mechanical Systems and Signal Processing*, 1997,11(4):529-545.
- [5] Lee D D, Seung H S. Learning the parts of objects by non-negative matrix factorization[J]. *Nature*, 1999, 401(6755):788-791.
- [6] Liu W, Zheng N. Non-negative matrix factorization based methods for object recognition [J]. *Pattern Recognition Letters*, 2004,25(8):893-897.
- [7] Zhang T, Fang B, Tang Y Y, et al. Topology preserving non-negative matrix factorization for face recognition[J]. *IEEE Transactions on Image Processing*, 2008,17(4):574-584.
- [8] Bucak S S, Günsel B. Incremental subspace learning via non-negative matrix factorization [J]. *Pattern Recognition*, 2009,42(5):788-797.
- [9] Qing H W, Yun Z Y, Lei C, et al. Fault diagnosis for diesel valve trains based on non-negative matrix factorization and neural network ensemble[J]. *Mechanical Systems and Signal Processing*, 2009,23(5):1683-1695.
- [10] 蔡蕾,朱永生. 基于稀疏性非负矩阵分解和支持向量机的时频图像识别[J]. *自动化学报*, 2009,35(10):1272-1277.
Cai Lei, Zhu Yongsheng. Time-frequency spectra recognition based on sparse non-negative matrix factorization and support vector machine[J]. *Acta Automatica Sinica*, 2009,35(10):1272-1277. (in Chinese)
- [11] 高宏娟,潘晨. 基于(2d)-2 Nmf 及其改进算法的人脸识别[J]. *计算机应用*, 2007,27(7):1660-1663.
Gao Hongjuan, Pan Chen. Face recognition based on (2D)-2NMF and its improvement [J]. *Journal of Computer Applications*, 2007,27(7):1660-1663. (in Chinese)
- [12] Stockwell R G, Mansinha L, Lowe R P. Localization of the complex spectrum: the S transform[J]. *IEEE Transactions on Signal Processing*, 1996,44(4):998-1001.
- [13] Li S Z, Hou X W, Zhang H J, et al. Learning spatially localized, parts-based representation[C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2001.
- [14] Zhang D, Chen S, Zhou Z H. Two-dimensional non-negative matrix factorization for face representation and recognition[J]. *Lecture Notes in Computer Science*, 2005,3723:350-363.
- [15] 高宏娟,潘晨. 基于非负矩阵分解的人脸识别算法的改进[J]. *计算机技术与发展*, 2007,17(11):63-67.
Gao Hongjuan, Pan Chen. Improved face recognition algorithm based on non-negative matrix factorization [J]. *Computer Technology and Development*, 2007, 17(11):63-67. (in Chinese)



第一作者简介:李兵,男,1982年9月生,博士研究生。主要研究方向为信号处理、模式识别、状态监测和故障诊断。曾发表《基于形态梯度解调算子的齿轮故障特征提取》(《振动、测试与诊断》2010年第30卷第1期)等论文。

E-mail:rommandy@163.com