

DOI: 10.19650/j.cnki.cjsi.J2210199

单时频谱非负矩阵编码与解调的特征提取^{*}

张焱, 蔡有鑫, 王平, 韩延, 黄庆卿

(重庆邮电大学工业物联网与网络化控制教育部重点实验室 重庆 400065)

摘要:非负矩阵分解(NMF)作为一种矩阵分解以及非线性维数约简工具,被广泛用于多样本振动时频谱的分解编码以及特征提取,但单样本振动时频谱的NMF编码、尤其NMF分解向量与振动时频谱分量间关联关系尚缺乏探讨。阐述了单时频谱编码与解调的特征提取原理,重点分析了NMF对单时频谱基于部分的特征表示能力、单时频谱NMF基向量的带通滤波幅频特性(BFAC)、以及NMF编码向量与时频谱分量的同步变化特性。提出单时频谱NMF编码与解调的两种特征提取新方法,即基于NMF基向量的滤波解调和NMF编码向量直接解调,定义一种BFAC指数指标和基向量归一化的NMF编码优化迭代规则分别用于NMF低维参数自适应选取和优化求解过程。将所提方法用于仿真信号以及齿轮箱振动信号分析,6.4 k长度数据在给定因子分解秩和NMF最大迭代300次终止条件设定下的特征提取用时约3.5 s,同时实现了对信噪比为-10 dB仿真信号以及多故障齿轮箱振动信号中故障特征的提取。

关键词:特征提取;故障诊断;时频谱;非负矩阵分解;编码;解调

中图分类号: TH165.3 TN911.2 **文献标识码:** A **国家标准学科分类代码:** 510.40

Feature extraction using nonnegative matrix coding and demodulation of single spectrogram

Zhang Yan, Cai Youxin, Wang Ping, Han Yan, Huang Qingqing

(Key Laboratory of Industrial Internet of Things & Networked Control, Ministry of Education, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: Negative matrix factorization (NMF), a tool for matrix factorization and nonlinear dimension reduction, has been widely used for factorization coding and feature extraction of vibration spectrograms from multiple samples. However, the systematic exploration and research on the NMF coding of single vibration spectrogram and the relationships between NMF factorization vector and vibration spectrogram components is still lacking. The basic principles of feature extraction using single spectrogram coding and demodulation are explained. The ability of the NMF to part-based characterize single spectrogram, the bandpass filtering amplitude-frequency characteristics (BFAC) of single spectrogram NMF basis vector, and the synchronous change characteristic between NMF coding vector and spectrogram component are mainly discussed. Two novel feature extraction methods using nonnegative matrix coding and demodulation of single spectrogram are developed, filtering demodulation based on NMF basis vector and demodulation directly performed on NMF coding vector. A BFAC index is defined for adaptive selection of NMF low dimensional parameter, likewise an optimal iteration rule with NMF basis vector normalization for optimization process of NMF coding. The proposed method is applied to analyze the simulated signal and gearbox vibration signal. Under the conditions of given factorization rank and NMF maximum iterations of 300, it takes about 3.5 s to extract the features of data with the length of 6.4 k points. Meanwhile, the fault characteristics of the simulated signal with signal-to-noise ratio of -10 dB and that of the gearbox vibration signal with compound faults are extracted.

Keywords: feature extraction; fault diagnosis; spectrogram; nonnegative matrix factorization; coding; demodulation

收稿日期: 2022-07-26 Received Date: 2022-07-26

^{*} 基金项目: 中国博士后科学基金(2022MD713687)、国家自然科学基金(51705056)、重庆市博士后科学基金(cstc2021jcyj-bshX0094)、重庆市自然科学基金面上项目(cstc2021jcyj-msxmX0556)资助

0 引 言

旋转机械振动信号通常为复杂的多分量调制信号,该类信号具有的非平稳特征和机械系统结构及其运行状态密切相关,包含有丰富的反映如轮齿缺陷、轴承点蚀等局部损伤、以及因制造安装等导致的动静件碰摩、过大齿侧间隙等故障缺陷。基于振动分析的故障特征提取对于判定故障发生与否以及确定故障部位与形式等具有重要意义^[1-4]。

机械故障相关的冲击脉冲往往诱发系统形成一系列固有频率冲击响应,且该冲击响应以及如齿轮啮合等系统高频分量往往也受到冲击脉冲的调制。解调分析是冲击类振动特征提取及故障诊断的常用手段,其关键是对故障相关信号分量的分离和解调,其中滤波频带的选取尤为关键^[5-7]。Osorio 等^[5]提出基于谱峭度选取滤波频带,Berntsen 等^[6]提出基于运行模态分析的解调频带选取用于轴承诊断,张焱等^[7]提出基于同步压缩小波变换和广义解调的无转速计下变工况滚动轴承故障特征提取方法。然而,实测振动信号中故障相关冲击分量通常具有持续时间短、信噪比低、极易淹没于背景噪声中等特点,加大了故障特征提取难度。时频分析联合时域和频域描述,使得聚焦该类含冲击的非平稳信号的局部特性成为可能,短时傅里叶变换、小波分析、希尔伯特黄变换等是常用时频分析方法,Yang 等^[8]对参数化时频分析方法及其工程应用进行了综述,相关研究表明基于振动时频谱的特征提取是进行故障诊断的有效途径。

非负矩阵分解(negative matrix factorization, NMF)作为一种矩阵分解以及非线性维数约简工具^[9],被大量应用于图像、语音、振动等数据的分析处理。Gan 等^[10]对 NMF 的设计原理和特点、各种变体及其应用等进行了综述和分析。在机械故障特征提取与诊断方面,Wang 等^[11]提出一种基于核鲁棒非负矩阵分解的工业过程故障监测方法,Liang 等^[12]提出一种稀疏核非负矩阵分解的特征提取方法并用于轴承故障诊断,Xiu^[13]等提出一种结构化联合稀疏非负矩阵分解方法,并用于多种设备的故障检测,Shen 等^[14]提出一种协同聚类方法并基于 NMF 所得矩阵向量进行多任务齿轮故障诊断,Wang 等^[15]对多个 NMF 层进行堆栈以层次化学习方式提取机械故障特征,Wodecki 等^[16]基于 NMF 和多通道数据融合对振动信号中局部损伤相关分量分离提取。上述研究中,基于 NMF 的故障特征提取与诊断普遍将 NMF 作为维数约简工具,应用于多样本的分解编码及特征提取,形成多阶段级联诊断模式,即“多样本下 NMF 模型训练→基于 NMF 模型的测试样本特征计算→分类器模式识别”。

值得注意的是,单样本振动时频幅值谱满足 NMF 非负矩阵输入要求,更重要的是,从信号组成角度看,时频幅值谱可分解为多个不同时频分量的叠加,这与 NMF 的“纯加性”描述理念完全契合。

受此启发,本文考虑对单样本振动时频幅值谱进行 NMF 处理和分析。鉴于 NMF 基向量具备刻画原始非负输入数据局部特征的能力,同时由于 NMF 是以编码向量作为加权因子、通过各基向量的组合以逼近原始输入数据,对于输入为单样本振动时频幅值谱的情况,NMF 的多个基向量所刻画的局部特征实际上分别对应于机械振动中具有实际物理含义的特定频分量,同时编码矩阵中非零元素的位置即指示了这些时频分量在振动时频幅值谱中的位置信息,包括水平方向的时刻信息以及垂直方向的频率信息等。显然,单样本振动时频幅值谱的 NMF 基向量和编码向量与原始机械振动的时频分量间存在密切关联,且 NMF 基向量和编码向量均含有反映机械运行状态的特征信息,这些信息均可作为机械故障判定提供相关依据。然而,现有 NMF 研究中尚缺乏关于单样本振动时频谱的 NMF 编码、以及 NMF 分解向量与振动时频谱分量间关联关系的探讨。

本文分析单样本振动时频幅值谱与其 NMF 分解向量间的关系,提出单时频谱非负矩阵编码与解调的特征提取方法,该方法以单样本时频谱作为载体,无需多样本的输入,也不必借助分类器开展模式识别,创新了 NMF 在机械故障特征提取与诊断方面的应用模式,同时为故障特征提取与诊断提供新思路。主要创新性工作包括:1)论述单时频谱编码与解调的特征提取原理,重点阐述单样本时频谱的 NMF 编码特征表示、单样本时频谱 NMF 基向量的带通滤波幅频特性、以及时频谱分量与 NMF 编码向量的同步变化特性;2)提出单时频谱 NMF 编码与解调特征提取的两种新方法,即基于 NMF 基向量的滤波解调和 NMF 编码向量直接解调,其中 NMF 编码向量直接解调避免了滤波解调中的滤波器构造与信号滤波等操作;3)分析 NMF 参数设置以及求解算法对所提特征提取方法的影响,明确单时频谱 NMF 编码的多目标优化及约束要求,进而提出一种 BFAC 指数指标以及基向量归一化的单时频谱 NMF 编码优化迭代规则分别用于 NMF 低维参数自适应选取和优化求解过程。利用仿真数据和实测齿轮箱振动信号对所提方法进行验证。

1 单时频谱编码与解调的特征提取原理

1.1 单样本时频谱基于部分的特征表示

振动信号时频谱可同时从时域和频域角度刻画信号局部特征,对于以矩阵 $\mathbf{Z} \in \mathbb{R}^{M \times N}$ 表示的时频谱,根据矩阵分解理论,多种算法如 NMF、主成分分析(principal

component analysis, PCA) 和矢量量化(vector quantization, VQ) 均符合式(1)所示的矩阵分解框架。

$$\mathbf{Z} \approx \mathbf{WH}, \mathbf{W} \in \mathbb{R}^{M \times r}, \mathbf{H} \in \mathbb{R}^{r \times N} \quad (1)$$

式中: $\mathbf{W}$ 为 $M \times r$ 维的基矩阵, $\mathbf{H}$ 为 $r \times N$ 维的编码矩阵, 统称 $\mathbf{W}$ 和 $\mathbf{H}$ 中各向量为分解向量, r 为因子分解秩, 一般情况下, r 设置为远小于 N 和 M , 从而 $\mathbf{WH}$ 可看作对矩阵 $\mathbf{V}$ 的压缩表示。

NMF 与 PCA 和 VQ 算法的主要区别在于施加在 $\mathbf{W}$ 和 $\mathbf{H}$ 上的约束不同。NMF 是一种非负约束下的矩阵分解与编码方法, 与 PCA 和 VQ 等算法得到“整体”特征表示不同, NMF 旨在学习基于“部分”的特征表示。

对于满足 NMF 非负输入要求的单样本时频幅值谱 $\mathbf{V} \in \mathbb{R}_+^{M \times N}$, 在矩阵分解中对 $\mathbf{W} \in \mathbb{R}_+^{M \times r}$ 和 $\mathbf{H} \in \mathbb{R}_+^{r \times N}$ 添加非负约束, NMF 表示为^[9]:

$$\mathbf{V} \approx \mathbf{WH} \quad \text{s. t. } v_{ij} \geq 0, w_{ik} \geq 0, h_{kj} \geq 0 \quad (2)$$

$\forall i = 1, \dots, M, \forall k = 1, \dots, r, \forall j = 1, \dots, N$

其中, 称基矩阵 $\mathbf{W} \in \mathbb{R}_+^{M \times r}$ 中各列 $\mathbf{W}_k, \forall k = 1, 2, \dots, r$, r 为基向量, 编码矩阵 $\mathbf{H} \in \mathbb{R}_+^{r \times N}$ 中各行 $\mathbf{H}_k, \forall k = 1, 2, \dots, r$ 为编码向量。

基向量 $\mathbf{W}_k$ 捕捉时频幅值谱 $\mathbf{V}$ 的局部时频特征, 编码矩阵中非零元素的位置即指示了相关局部时频特征在时频幅值谱 $\mathbf{V}$ 中的位置。例如, 时频幅值谱 $\mathbf{V} \in \mathbb{R}_+^{M \times N}$ 中第 j 列 ($j \in \{1, \dots, N\}$) 可由以编码矩阵元素 $[h_{1j}, h_{2j}, \dots, h_{rj}]^T$ 作为加权因子的各基向量 $\mathbf{W}_k, k = 1, 2, \dots, r$ 以线性组合方式逼近, 从而即实现基于部分的单时频谱特征表示。

1.2 单样本时频谱分量与 NMF 分解向量间关系

为便于阐述, 图 1 给出了单时频谱编码与解调的特征提取原理示意, 图 1(a) 为信号幅值谱示意图, 包含位于不同频带内的两个信号分量 C_1 和 C_2 。图 1(b)~(d) 分别对应式(2)中的时频幅值谱 $\mathbf{V}$ 、基矩阵 $\mathbf{W}$ 和编码矩阵 $\mathbf{H}$, 时频幅值谱 $\mathbf{V}$ 的横坐标对应于时间 t , 纵坐标对应于频率 f , 图 1 中以正方形对冲击特征进行示意, 以模拟机械故障振动中常见的周期性冲击, 并以颜色深浅表示幅值大小差异, 以空心圆表示 0 值。

1) 时频谱基向量的带通滤波幅频特性

时频幅值谱 $\mathbf{V}$ 中各列实际为不同时刻的瞬时谱向量 $\text{Spec}[f, t[j]], \forall j = 1, 2, \dots, N$, 其中 f 对应原振动信号的全奈奎斯特频率域。根据式(2), NMF 处理可看作对各瞬时谱向量 $\text{Spec}[f, t[j]]$ 的聚类, 基矩阵 $\mathbf{W}$ 中包含 r 个与聚类簇相关的基向量 $\mathbf{W}_k, \forall k = 1, 2, \dots, r$ 。基向量 $\mathbf{W}_k$ 可看作 N 个瞬时谱向量 $\text{Spec}[f, t[j]]$ 在特定频带范围的“均值化”谱向量, 反映了聚类簇所对应的特定谱分量所在频段位置, 即基向量具有类似带通滤波器的幅频特性, 且其通带对应于“均值化”谱向量所在频段。

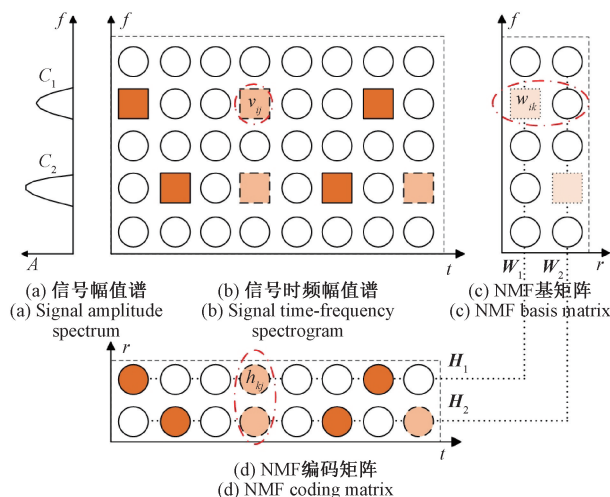


图 1 特征提取原理示意图

Fig. 1 Schematic diagram for feature extraction

因此, 可依据 NMF 基向量 $\mathbf{W}_k, \forall k = 1, 2, \dots, r$ 的“通带”设计滤波器频率响应, 并通过对原振动信号进行滤波处理, 可分离得到特定频带范围的信号分量, 进而通过对滤波信号的解调分析实现特征提取。

为便于阐述, 后文称基向量 $\mathbf{W}_k$ 关于频率 f 的“基向量-频率”曲线为“BFAC 曲线”, 基向量 $\mathbf{W}_k$ 及其对应编码向量 $\mathbf{H}_k$ 统称为关联向量, 记为 $Q_k = \{\mathbf{W}_k, \mathbf{H}_k\}$ 。

2) 时频谱分量与 NMF 编码向量的同步变化特性

设时频幅值谱 $\mathbf{V}$ 中元素 $v_{ij}, i \in \{1, \dots, M\}, j \in \{1, \dots, N\}$ 来自于信号分量 $C_\tau, \tau \in \{1, \dots, \lambda\}, \lambda$ 表示信号分量数目, 根据式(2), v_{ij} 可近似表示为:

$$v_{ij} = \sum_{k=1}^r w_{ik} h_{kj} \quad (3)$$

特别的, 当单信号分量 C_τ 的时频幅值谱 $\mathbf{V}_\tau$ 仅与单一关联向量 Q_k 相关时, 时频幅值谱 $\mathbf{V}_\tau$ 可由且仅由 $\mathbf{W}_k \mathbf{H}_k$ 进行近似重构, 式(3)中元素 v_{ij} 可近似表示为:

$$v_{ij} = w_{ik} h_{kj} \quad (4)$$

在这种情况下, 一方面, 基向量 $\mathbf{W}_k$ 依然具备前述带通滤波幅频特性, 并且将唯一对应信号分量 C_τ 。(例如, 图 1 中信号分量 C_1 和 C_2 分别仅与基向量 $\mathbf{W}_1$ 和 $\mathbf{W}_2$ 相关, 基向量 $\mathbf{W}_1$ 和 $\mathbf{W}_2$ 中非零元素位置恰对应于 C_1 和 C_2 所处频带)。

另一方面, 当单信号分量 C_τ 的时频幅值谱 $\mathbf{V}_\tau$ 仅与单一关联向量 Q_k 相关时, 根据式(4), 单信号分量 C_τ 的幅值与编码向量 $\mathbf{H}_k$ 的幅值间呈比例关系, 即编码向量 $\mathbf{H}_k$ 沿时间轴的幅值变化与信号分量 C_τ 沿时间轴的幅值变化同步(如图 1 中, 信号分量 C_1 中各冲击时刻以及冲击间隔被同步反映在编码向量 $\mathbf{H}_1$ 中, C_2 与 $\mathbf{H}_2$ 间情况类似)。因此, 通过编码向量的直接解调有望提取出故障振

动信号中冲击相关特征。

2 单时频谱编码与解调的特征提取方法

2.1 NMF 参数设置以及求解算法影响分析

NMF 因子分解秩 r 以及求解算法不仅影响式 (2) 中 $\mathbf{WH}$ 对 $\mathbf{V}$ 的近似程度, 同时对单时频谱 NMF 编码与解调特征提取效果有重要影响, 具体说明如下。

根据式 (2) 知, 单样本时频幅值谱 $\mathbf{V} \in \mathbb{R}_+^{M \times N}$ 的 NMF 分解具有非唯一解, 对于信号分量 C_τ 对应的时频幅值谱 $\mathbf{V}_\tau$, 其既可能由单一关联向量 $\mathbf{Q}_k$ 近似重构, 即时频谱分量 $\mathbf{V}_\tau$ 与关联向量 $\mathbf{Q}_k$ 为“一对一”关系, 也可能由基矩阵 $\mathbf{W}$ 中多个基向量以线性组合方式进行近似, 即“多对多”关系。

当呈现为“多对多”关系时, 一方面, 信号分量 C_τ 沿时间轴的幅值变化情况难以被单一编码向量 $\mathbf{H}_k$ 沿时间轴的幅值变化完全反映, 难以通过对编码向量 $\mathbf{H}_k$ 的直接解调分析提取出信号分量 C_τ 中所含故障特征; 另一方面, 单个基向量 $\mathbf{W}_k, k = 1, 2, \dots, r$ 往往也与多个信号分量相关, 依据 $\mathbf{W}_k$ 设计的滤波器难以滤波得到单分量信号, 而是多分量共存于滤波信号中, 这无疑增大了滤波解调故障特征辨识的难度。

此外, 对于两不同信号分量 C_τ 和 C_ς 及其对应的时频幅值谱 $\mathbf{V}_\tau$ 和 $\mathbf{V}_\varsigma$, 假设 $\mathbf{V}_\tau$ 和 $\mathbf{V}_\varsigma$ 分别仅与关联向量 $\mathbf{Q}_k$ 和 $\mathbf{Q}_\rho$ 呈“一对一”关系, 根据式 (4), 对于特定时频幅值谱元素, 其基向量元素与编码向量元素呈反比例, 从而当两基向量 $\mathbf{W}_k$ 和 $\mathbf{W}_\rho$ 幅值大小关系未知时, 难以根据两编码向量 $\mathbf{H}_k$ 和 $\mathbf{H}_\rho$ 的幅值大小对信号分量 C_τ 和 C_ς 的幅值大小进行判定, 将阻碍基于编码向量直接解调的故障特征定量分析。

综合上述分析, 面向单时频谱编码与解调的特征提取既要求最小化式 (2) 中 $\mathbf{WH}$ 对 $\mathbf{V}$ 的整体近似误差、也要求尽可能满足各时频谱分量与其关联向量呈“一对一”关系、且编码向量能定量反映其对应信号分量的幅值大小, 该多目标优化及约束下 NMF 表示为:

$$\begin{aligned} \mathbf{V} &\approx \mathbf{WH} \\ \mathbf{V}_\tau &\approx \mathbf{W}_k \mathbf{H}_k \\ \mathbf{H}_k &\propto C_\tau \\ \text{s. t. } &v_{ij} \geq 0, w_{ik} \geq 0, h_{kj} \geq 0 \\ &\forall i = 1, \dots, M, \forall k = 1, \dots, r, \\ &\forall j = 1, \dots, N, \forall \tau = 1, \dots, \lambda \end{aligned} \quad (5)$$

然而, 现有关于 NMF 研究中, 参数 r 设置以及 NMF 优化求解算法设计普遍以最小化式 (2) 的近似误差为目标, 难以满足时频谱分量与其关联向量的“一对一”关

系, 同时由于缺乏对基向量的取值约束, 多个基向量间不具有可比性, 无法直接用于单样本时频谱编码与解调的特征提取。

2.2 BFAC 指数与低维参数自适应选取

提出一种 BFAC 指数指标, 以指导 NMF 因子分解秩 r 的选取, 其基本思想是定量评价不同基向量 BFAC 曲线的最大重叠程度。由于时频谱 NMF 基向量的带通滤波幅频特性, 多个 NMF 基向量相当于一组自适应滤波器组。当因子分解秩 r 偏大时, 多个基向量的通带中心频率会相距较近, 导致通带间重叠; 当 r 偏小时, NMF 基向量的滤波通带难以覆盖全奈奎斯特频率域, 导致原信号中一些重要信息无法提取。鉴于多个基向量的主要不同点在于通带的不同, 因此, 通过对各基向量的 BFAC 曲线分布以及重叠程度进行定量评价可指导选取合适的因子分解秩 r_p 。

相关系数可衡量变量间的线性关系, 结合对基向量 BFAC 曲线间皮尔逊相关系数的分析, 定义 BFAC 指数指标为:

$$H_{\text{BFAC}} = \text{MAX}(|\text{corrcoef}(\mathbf{W}) \odot (\mathbf{I}_r - \mathbf{E}_r)|) \quad (6)$$

式中: $\text{corrcoef}(\cdot)$ 为在矩阵不同列向量间计算皮尔逊相关系数, $\odot$ 表示哈达玛积运算, $\mathbf{I}_r$ 表示元素全为 1 的 $r \times r$ 维矩阵, $\mathbf{E}_r$ 表示 r 阶单位矩阵, $\text{MAX}(\cdot)$ 为取矩阵最大元素值操作。 H_{BFAC} 实质表示最相关的两基向量间的相关系数绝对值, 取值范围为 $[0, 1]$ 。

通过计算不同 r 取值下 BFAC 指数指标 H_{BFAC} 可为选择 r_p 提供参考。当 H_{BFAC} 较小时, 表明最相关的两组基向量的 BFAC 曲线间重叠较少, NMF 分解倾向于呈现为“一对一”关系, 随着 H_{BFAC} 增大, 最相关的两组基向量的 BFAC 曲线间重叠逐渐增大, NMF 分解倾向于呈现为“多对多”关系。

需注意的是, 为尽可能保证各时频谱分量可由且仅由单一关联向量进行近似, NMF 因子分解秩 r 应取值较小, 但 r 过小会导致式 (5) 中 $\mathbf{WH}$ 对 $\mathbf{V}$ 近似误差增加。本文综合 BFAC 指数及近似误差, 选取 BFAC 指数小于给定阈值 $H_{\text{Threshold}}$ 的最大 r 值, 选取规则如式 (7)。

$$\begin{aligned} r_p &= \text{MAX}(r) \\ \text{s. t. } &H_{\text{BFAC}} \leq H_{\text{Threshold}} \end{aligned} \quad (7)$$

2.3 基向量归一化的单时频谱 NMF 编码优化

NMF 求解本质是一个最优化问题, 提出一种基向量归一化的单时频谱 NMF 编码优化迭代求解算法。记 NMF 分解中时频幅值谱 $\mathbf{V} \in \mathbb{R}_+^{M \times N}$ 的噪声矩阵为 $\mathbf{E} \in \mathbb{R}^{M \times N}$, 则:

$$\mathbf{E} = \mathbf{V} - \mathbf{WH} \quad (8)$$

当噪声服从不同概率分布时,为最小化 $\|E\|$,通过最大化似然函数会得到不同形式的目标函数(objective function, OF),如基于欧氏距离、KL 散度和泊松误差等^[9]。

考虑到机械振动噪声通常为高斯噪声,在高斯分布噪声假设下的 NMF 求解对数似然函数为:

$$\ln L(\mathbf{W}, \mathbf{H}) = \sum_{i,j} \ln \frac{1}{\sqrt{2\pi} \sigma_{ij}} - \frac{1}{2\sigma_{ij}} \sum_{i,j} [v_{ij} - (\mathbf{WH})_{ij}]^2$$

(9)

为最大化该对数似然函数,假设各数据点噪声方差 σ_{ij} 相同,进而转化为最小化式(10)中基于欧氏距离度量的目标函数。

$$H_j(\mathbf{W}, \mathbf{H}) = \frac{1}{2} \sum_{i,j} [v_{ij} - (\mathbf{WH})_{ij}]^2$$

(10)

在以 $H_j(\mathbf{W}, \mathbf{H})$ 对 w_{ik} 和 h_{kj} 分别求偏导基础上,可采用梯度下降法对 NMF 优化问题进行迭代求解^[9]。

针对单时频谱 NMF 编码的多目标优化及约束需求,提出在迭代求解中引入基向量归一化处理,最终基向量归一化的时频谱 NMF 编码优化迭代规则为:

$$w_{ik} \leftarrow w_{ik} \frac{(\mathbf{VH}^T)_{ik}}{(\mathbf{WHH}^T)_{ik}}$$

(11)

$$w_{ik} \leftarrow \frac{w_{ik}}{\sum_{l=1}^M w_{lk}}$$

(12)

$$h_{kj} \leftarrow h_{kj} \frac{(\mathbf{W}^T \mathbf{V})_{kj}}{(\mathbf{W}^T \mathbf{WH})_{kj}}$$

(13)

交替迭代更新 $\mathbf{W}$ 和 $\mathbf{H}$ 至收敛,当目标函数小于设定阈值或迭代次数达到设定最大次数时,迭代终止。

2.4 算法流程

所提单时频谱 NMF 编码与解调的特征提取算法的主要步骤如下:

- 1) 计算原始振动信号 $x(t)$ 的时频幅值谱,并作为 NMF 的输入 $\mathbf{V} \in \mathbb{R}_+^{M \times N}$;
- 2) 给定 BFAC 指数阈值 $H_{\text{Threshold}}$,依据 2.2 节低维参数自适应选取方法选取因子分解秩 r_p ;
- 3) 根据 2.3 节基向量归一化的单时频谱 NMF 编码优化方法,计算 NMF 分解向量 $\mathbf{W}_k$ 和 $\mathbf{H}_k, \forall k=1,2,\cdots,r_p$;
- 4) 基于各基向量和编码向量分别开展解调分析,以实现特征提取与故障诊断:

(1) 对于编码向量 $\mathbf{H}_k, \forall k=1,2,\cdots,r_p$,直接对其进行包络解调分析;

- (2) 对于基向量 $\mathbf{W}_k, \forall k=1,2,\cdots,r_p$:
- ① 依据基向量 $\mathbf{W}_k$ 的带通滤波幅频特性构建带通滤波器 F_{wk} ;

- ② 利用 F_{wk} 对 $x(t)$ 滤波得滤波信号 $x_{wk}(t)$;
- ③ 对滤波信号 $x_{wk}(t)$ 进行包络解调分析。

3 试验验证

3.1 试验 1—仿真信号分析

模拟旋转机械局部损伤故障振动信号对所提方法有效性进行验证,故障冲击序列生成模型为^[7]:

$$s(t) = \sum_{m=1}^M A_m e^{-\beta(t-t_m)} \sin(2\pi f_\omega(t-t_m)) \mathcal{E}(t-t_m)$$

(14)

其中, M 为该冲击序列中的冲击个数, A_m 和 t_m 分别为第 m 个冲击的幅值和发生时刻, β 表示结构阻尼, f_ω 表示系统高频响应频率, $\mathcal{E}(t)$ 为单位阶跃函数。综合瞬时转频和冲击间隔,第 m 个冲击发生时刻 t_m 为:

$$t_m \approx t_{m-1} + \frac{1}{FCC \times f_r(t_{m-1})}, m=1,2,\cdots,M$$

(15)

其中, t_{m-1} 为第 $m-1$ 个冲击发生时刻, $f_r(t_{m-1})$ 为 t_{m-1} 时刻瞬时转频, FCC 表示故障特征系数,与实际故障类型相关。

考虑到实测故障振动信号通常包含有背景噪声以及轴转频等干扰,在冲击序列中加入幅值 0.5 频率分别为 50 Hz 及 200 Hz 的两正弦分量、以及高斯白噪声,信噪比为-10 dB。图 2(a)为加干扰后混合信号 $x(t)$ 的时域波形,混合信号相关参数设置如表 1 所示,采样频率 f_s 为 12.8 kHz,取 0.5 s 时长数据进行分析。

表 1 仿真信号参数
Table 1 Parameters of the simulated signal

参数	取值
转频 f_r/Hz	35
故障特征系数 FCC	4
幅值系数 A_m	2
响应频率 f_ω/Hz	2 000
结构阻尼 β	1 000
信噪比 SNR/dB	-10

将所提单时频谱 NMF 编码与解调的特征提取方法用于仿真信号分析。考虑到小波分析能通过伸缩和平移变换对信号进行多尺度细化,基于连续小波变换构建时频谱。图 2(b)所示为仿真信号时频幅值谱(为便于可视化,仅显示前 0.05 s 数据),分析小波采用与机械故障振动信号冲击衰减特征匹配较好的 amor 小波,从图 2(a)和(b)中难以直接辨识冲击间隔、故障频率等特征。

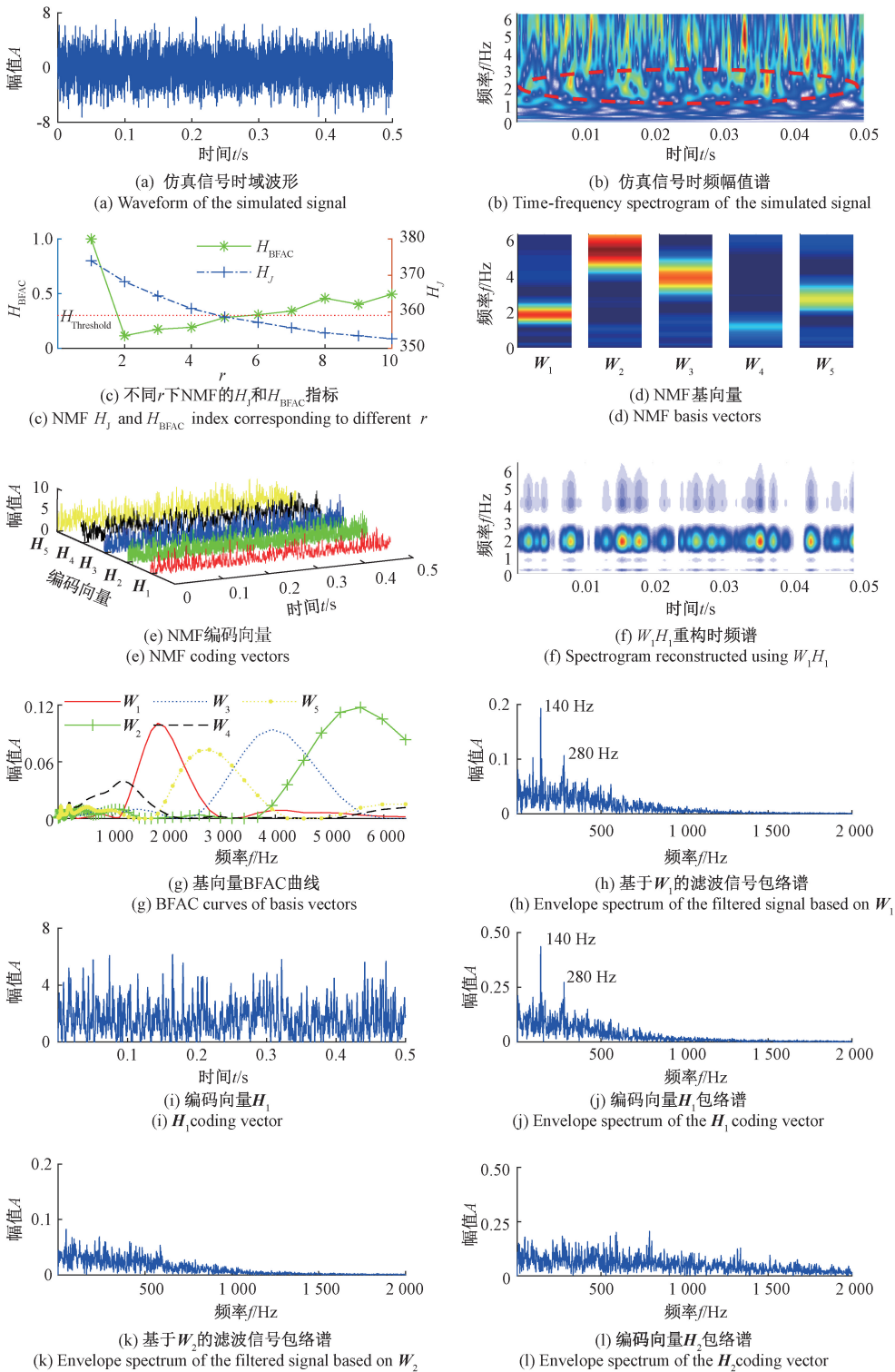


图 2 仿真信号及其分析结果

Fig. 2 Simulation signal and its analysis results

计算不同因子分解秩 r 下的 BFAC 指数以及式 (10) 的目标函数, 如图 2(c) 所示, 在同一 r 值下进行 10 次独立实验并取平均值以抑制随机初始化的影响, 且 NMF 迭

代终止条件设定为最大 300 次。综合考虑 BFAC 指数及重构误差, 给定 BFAC 指数阈值 $H_{Threshold}$ 为 0.3, 并确定因子分解秩 r_p 为 5。图 2(d) 和 (e) 分别为时频幅值谱 NMF

基向量和编码向量,图 2(e)中各编码向量沿时间轴呈较明显的冲击,对比图 2(d)和(b)可知,基向量较好的指示了时频谱中主要时频分量所处频带位置,如基向量 W_1 对应于原信号 2 000 Hz 处高频响应。

图 2(f)为由分解向量 W_1 和 H_1 相乘所得重构时频谱,对比图 2(b)和(f)知,重构时频谱分量较好的匹配了原信号中 2 000 Hz 处高频响应,极大的抑制了其余频段时频谱分量和噪声的干扰,其周期性冲击特征更加明显。图 2(g)为基向量 BFAC 曲线,显然 BFAC 曲线具有类似带滤波器的幅频特性,可依据基向量的“通带”构建带通滤波器,并实现对原信号中特定信号分量的滤波分离。图 2(h)给出了基于基向量 W_1 所得滤波信号的包络幅值谱。图 2(i)和(j)分别为编码向量 H_1 、以及对其直接包络解调所得幅值谱。观察图 2(h)和(j)知,在故障特征频率($f_r \times FCC = 140$ Hz)及其倍频 280 Hz 位置存在明显谱线,表明所提方法有效提取出了故障相关特征。

基于基向量 $W_2, \dots, W_5$ 和编码向量 $H_2, \dots, H_5$ 进行解调分析。图 2(k)和(l)中给出了基于 W_2 的滤波信号包络幅值谱,以及编码向量 H_2 的包络幅值谱,从图 2(k)和(l)知,其中均无故障相关特征,且二者包络谱具有较低的幅值,这与仿真信号实际组成情况相符。与基于 W_2 和 H_2 所得结果相似,基于分解向量 W_3, W_4, W_5 和编码向量 H_3, H_4, H_5 所得各包络谱中均无故障相关特征。限于篇幅,不再赘述各包络谱情况。

为评估所提方法计算效率,基于 MATLAB 实现的算法,在配置为 11 代 Intel Core i7 2.80 GHz 处理器与 16 GB 内存硬件上多次测试,6.4 k 长度数据在给定因子分解秩以及 NMF 最大迭代 300 次终止条件设定下的单时频谱编码与解调特征提取平均耗时约 3.5 s,其中基向量归一化的单时频谱 NMF 编码优化过程耗时约 3 s。

3.2 试验 2—齿轮箱多故障振动信号分析

齿轮箱广泛应用于旋转机械系统中,将本文方法应用于齿轮箱局部损伤故障特征提取。振动信号采集于 DDS 动力传动故障诊断实验台,如图 3 所示,其包含 1 级行星齿轮箱和 2 级平行轴齿轮箱,电机与行星齿轮箱中太阳轮相连。表 2 列出了齿轮箱主要结构参数,行星齿轮箱传动比 4.57,平行轴齿轮箱传动比 0.725。平行轴齿轮箱中间轴上的滚动轴承存在内圈故障,同时在该轴输出端齿轮上存在断齿故障,并通过增加齿间间隙以增加噪声和旋转窜动干扰。多种故障以及干扰相互耦合,加大了故障特征提取难度。

在电机转速 1 176 rpm 下进行振动信号采集,表 3 为根据转速和齿轮箱结构计算所得相关特征频率。图 4 为所测振动信号时频幅值谱,信号由径向安装在平行轴齿轮箱中间轴输出端轴承座上的传感器测取,采样频率 25.6 kHz,取 2 s 时长数据进行分析(图 4 仅展示前 0.1 s

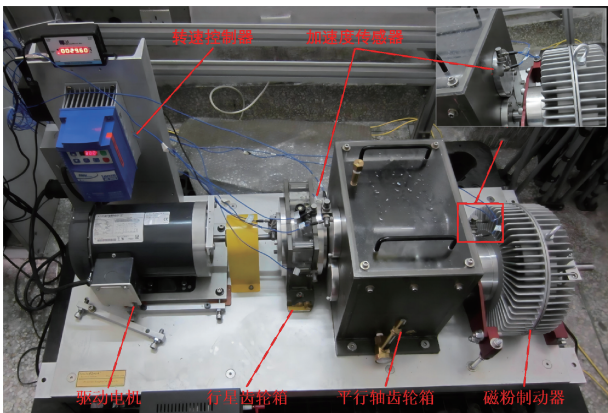


图 3 DDS 实验台
Fig. 3 DDS experimental platform

表 2 齿轮箱结构参数
Table 2 Structure parameters of the gearbox

齿轮类型		齿数
行星齿轮箱	内齿圈	100
	行星轮	36
	太阳轮	28
第 1 级平行轴	主动轮	100
	从动轮	29
第 2 级平行轴	主动轮	36
	从动轮	90

表 3 齿轮箱特征频率
Table 3 Characteristic frequencies of the gearbox

特征频率		频率/Hz
行星齿轮箱	太阳轮转频 f_r	19.60
	啮合频率 f_m	428.89
第 1 级平行轴	啮合频率 f_{m1}	428.89
	输入轴转频 f_{r1}	4.29
第 2 级平行轴	啮合频率 f_{m2}	532.38
	输入轴转频 f_{r2}	14.79
输出端轴承	内圈故障频率 f_{BPF1}	80.27
	外圈故障频率 f_{BPFO}	52.80
	滚动体故障频率 f_{BSF}	34.32

数据),图 4 中周期性冲击难以辨识。此外,也对原振动信号整体进行了包络谱分析,所得频谱分量主要对应于平行轴齿轮箱中间轴的 2 倍转频,即 $2f_{r2}$,而故障相关特征频率难以辨识。

利用所提方法分析该多故障耦合振动信号,计算不同因子分解秩 r 下的 NMF 目标函数和 BFAC 指数,如

图 5 所示,综合考虑 BFAC 指数及重构误差后选择因子分解秩 r_p 为 4。图 6 和 7 分别为齿轮箱信号时频谱 NMF 基向量及其 BFAC 曲线,图 8 为时频谱 NMF 编码向量,图 8 中各 NMF 编码向量的幅值变化沿时间轴呈较为明显的周期冲击。

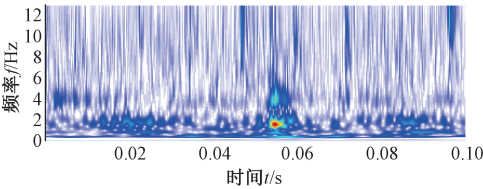


图 4 齿轮箱振动信号时频幅值谱

Fig. 4 Time-frequency spectrogram of gearbox vibration signal

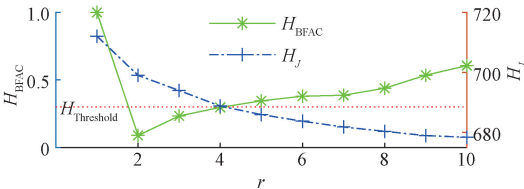


图 5 不同 r 下 NMF 的 H_j 和 H_{BFAC} 指标

Fig. 5 NMF's H_j and H_{BFAC} indices corresponding to different r

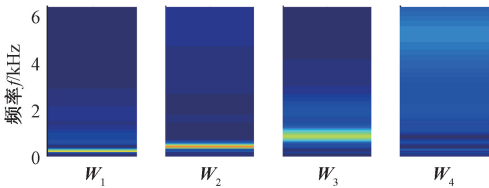


图 6 齿轮箱时频谱 NMF 基向量

Fig. 6 NMF basis vectors for gearbox time-frequency spectrogram

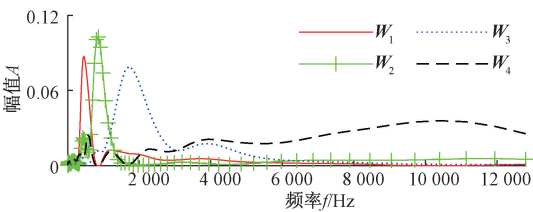


图 7 齿轮箱时频谱基向量 BFAC 曲线

Fig. 7 BFAC curves of basis vectors for gearbox time-frequency spectrogram

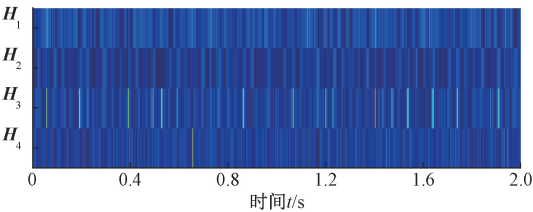
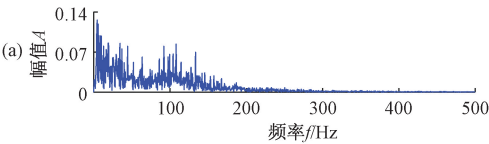


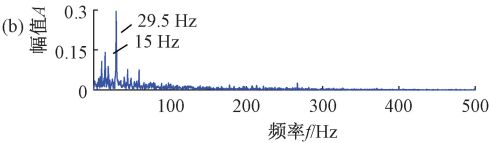
图 8 齿轮箱时频谱 NMF 编码向量

Fig. 8 NMF coding vectors for gearbox time-frequency spectrogram

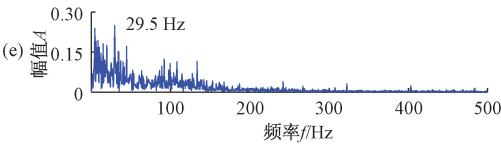
图 9(a)~(c)分别为基于基向量 $W_1 \sim W_4$ 所得滤波解调包络谱,图 9(e)~(h)分别为编码向量 $H_1 \sim H_4$ 的直接包络解调结果。从图 9 可知,图 9(d)和(h)中 80.5 Hz 及其倍频位置(如 161.5、242、322.5 Hz 等)谱线幅值突出,这与第 2 级平行轴位置的轴承内圈故障特征匹配,且这些谱线周围普遍存在较明显的 15 Hz 边频带;此外,图 9(b)和(f)中 15 Hz 及其二倍频位置谱线幅值突出,同时图 9(c)和(g)中 29.5 Hz 及其二倍频位置谱线也存在幅值突出现象,这与第 2 级平行轴转频及其输出端齿轮断齿故障特征频率相关。上述所提特征可为齿轮箱中齿轮故障和轴承故障的判定提供支撑。



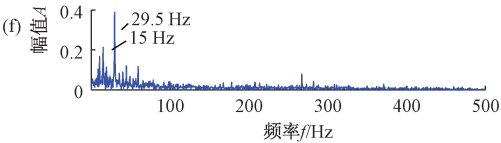
(a) 基于 W_1 的滤波信号包络谱
(a) Envelope spectrum of the filtered signal based on W_1



(b) 基于 W_2 的滤波信号包络谱
(b) Envelope spectrum of the filtered signal based on W_2



(e) 编码向量 H_1 包络谱
(e) Envelope spectrum of the H_1 coding vector



(f) 编码向量 H_2 包络谱
(f) Envelope spectrum of the H_2 coding vector

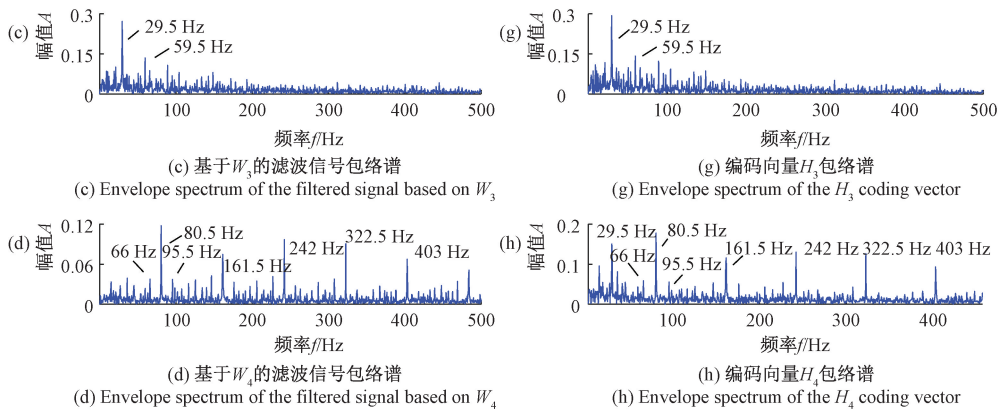


图 9 齿轮箱振动信号解调分析结果

Fig. 9 Demodulation results of gearbox vibration signal

4 结 论

结合机械振动时频幅值谱特点,阐述了单时频谱 NMF 编码与解调的特征提取基本原理以及两种新方法:基于 NMF 基向量的滤波解调和 NMF 编码向量直接解调,创新了 NMF 在机械故障特征提取与诊断方面的应用模式,同时为机械故障特征提取与诊断提供新思路,得到以下结论:

- 1) NMF 可实现单样本时频幅值谱基于部分的特征表示,NMF 基向量捕捉时频幅值谱局部特征,为时频谱分量的选择性近似重构和分离提供基础;
- 2) 单样本时频谱的 NMF 基向量具备带通滤波幅频特性,BFAC 曲线的通带位置直接指示了时频谱分量所在频段,为基于 NMF 基向量的滤波解调特征提取提供理论基础,同时可为滤波频带选取以及滤波器构建提供依据;
- 3) 当时频谱分量与其关联向量呈“一对一”关系,时频谱分量和编码向量二者沿时间轴的幅值变化同步,为 NMF 编码向量直接解调的特征提取奠定理论基础;此外,NMF 编码向量直接解调无需滤波器构造以及信号滤波等操作,算法实现更简单;
- 4) 基于 BFAC 指数的低维参数自适应选取以及基向量归一化的单时频谱 NMF 编码优化方法的联合运用促进了单时频谱 NMF 编码优化及约束的多目标有效达成,进而保证后续解调分析所提故障特征的可辨识度以及定量表征能力。

参考文献

[1] WEI Y, LI Y, XU M, et al. A review of early fault diagnosis approaches and their applications in rotating

machinery[J]. Entropy, 2019, 21(4): 409.

[2] CHAVAN N P, KALE M S, DESHMUKH S P, et al. A review on development and trend of condition monitoring and fault diagnosis [J]. ECS Transactions, 2022, 107(1): 17863.

[3] LAVAL X, MAILHES C, MARTIN N, et al. Amplitude and phase interaction in Hilbert demodulation of vibration signals: Natural gear wear modeling and time tracking for condition monitoring[J]. Mechanical Systems and Signal Processing, 2021, 150: 107321.

[4] KIM Y, HA J M, NA K, et al. Cepstrum-assisted empirical wavelet transform (CEWT)-based improved demodulation analysis for fault diagnostics of planetary gearboxes[J]. Measurement, 2021, 183: 109796.

[5] OSORIO SANTANDER E J, SILVA NETO S F, VAZ L A, et al. Using spectral kurtosis for selection of the frequency bandwidth containing the fault signature in rolling bearings [J]. Marine Systems and Ocean Technology, 2020, 15(4): 243-252.

[6] BERNTSEN J, BRANDT A, GRYLLIAS K. Enhanced demodulation band selection based on Operational Modal Analysis (OMA) for bearing diagnostics[J]. Mechanical Systems and Signal Processing, 2022, 181: 109300.

[7] 张焱,何姝钊,王平,等. 无转速计下变工况滚动轴承故障特征量化表征提取[J]. 仪器仪表学报, 2021, 41(8): 104-114.

ZHANG Y, HE SH B, WANG P, et al. Tacholeless quantitative characterization of rolling bearing fault feature under varying conditions [J]. Chinese Journal of

- Scientific Instrument, 2021, 41(8):104-114.
- [8] YANG Y, PENG Z, ZHANG W, et al. Parameterised time-frequency analysis methods and their engineering applications: A review of recent advances [J]. Mechanical Systems and Signal Processing, 2019, 119: 182-221.
- [9] LEE D D, SEUNG H S. Learning the parts of objects by non-negative matrix factorization [J]. Nature, 1999, 401(6755):788-791.
- [10] GAN J, LIU T, LI L, et al. Non-negative matrix factorization: A survey [J]. The Computer Journal, 2021, 64(7): 1080-1092.
- [11] WANG Y, SUN T, DING M, et al. A novel industrial process fault monitoring method based on kernel robust non-negative matrix factorization [J]. Measurement Science and Technology, 2021, 32(10): 105025.
- [12] LIANG L, DING X, LIU F, et al. Feature extraction using sparse kernel non-negative matrix factorization for rolling element bearing diagnosis [J]. Sensors, 2021, 21(11): 3680.
- [13] XIU X, FAN J, YANG Y, et al. Fault detection using structured joint sparse nonnegative matrix factorization [J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 3513011.
- [14] SHEN F, CHEN C, XU J, et al. A fast multi-tasking solution: NMF-theoretic co-clustering for gear fault diagnosis under variable working conditions [J]. Chinese Journal of Mechanical Engineering, 2020, 33(1): 1-15.
- [15] WANG X, WANG T, MING A, et al. Semi-supervised hierarchical attribute representation learning via multi-layer matrix factorization for machinery fault diagnosis [J]. Mechanism and Machine Theory, 2022, 167: 104445.
- [16] WODECKI J, MICHALAK A, ZIMROZ R, et al. Separation of multiple local-damage-related components from vibration data using Nonnegative Matrix Factorization and multichannel data fusion [J]. Mechanical Systems and Signal Processing, 2020, 145: 106954.

作者简介



张焱(通信作者), 2011 年和 2016 年分别于重庆大学获得学士学位和博士学位, 现为重庆邮电大学副教授, 主要研究方向为机电装备智能诊断与安全服役、工业物联网。

E-mail: yanzhang@cqupt.edu.cn

Zhang Yan (Corresponding author) received his B.Sc. and Ph.D. degrees both from Chongqing University in 2011 and 2016, respectively. He is currently an associate professor at Chongqing University of Posts and Telecommunications. His main research interests include equipment intelligent diagnosis and safety management, and industrial internet of things.