

DOI: 10.13382/j.jemi.B1902663

基于 LS 和 MTS 的两阶段滚动轴承故障特征选择方法

彭宅铭 程龙生 詹君 姚启峰

(南京理工大学 经济管理学院 南京 210094)

摘要:滚动轴承故障预测和健康管理(PHM)方法可以提取大量的故障特征数据,这些数据虽然有很大的潜在价值,但也存在高维、高冗余性的特点,难以直接分析和利用。因此,针对轴承故障特征数据的特点,以去除数据冗余性、筛选敏感特征为目的,提出两阶段特征选择算法。该方法的第1阶段采用拉普拉斯得分(LS)对原始特征按局部保持能力进行排序,利用互信息聚类算法删除特征集中的冗余特征。第2阶段采用多变量模式识别中的马田系统(MTS)方法对剩余特征进行综合评价,挖掘对故障分类更有效的特征。轴承退化仿真试验数据验证结果表明,提出的两阶段特征选择算法可以有效地去除冗余度、提高故障监测准确率,可以有效的运用到滚动轴承的初期故障检测中。

关键词:特征选择;拉普拉斯得分;互信息;马田系统;故障检测

中图分类号: TH133; TN911 **文献标识码:** A **国家标准学科分类代码:** 460

Two-stage feature selection method for rolling bearing diagnosis based on LS-MTS

Peng Zhaiming Cheng Longsheng Zhan Jun Yao Qifeng

(School of Economics and Management, Nanjing University of Science & Technology, Nanjing 210094, China)

Abstract: Rolling bearing prognostic and health management (PHM) method can extract a large number of fault characterization data. Those data are of great potential value, because of their characteristics of high-dimensionality and high-redundancy. However, direct analysis and utilization of them are impossible. Therefore, aiming at reducing the redundancy data and screening sensitive features, a two-stage feature selection algorithm is proposed. In the first stage of the method, the Laplacian score (LS) is used to sort the original features based on their locality preserving power, and the mutual information-based clustering algorithm is utilized to remove the redundant features of the original feature set. In the second stage, the Mahalanobis-Taguchi system (MTS), as a useful multivariate pattern recognition method, is employed to comprehensively evaluate the remaining features, unearthing features which are prone to fault classification. The verification results of the bearing degradation simulation test data show that the proposed two-stage feature selection algorithm can effectively remove redundancy and improve the accuracy of fault monitoring. This method can be effectively applied to the initial fault detection of rolling bearings.

Keywords: feature selection; Laplacian score (LS); mutual information; Mahalanobis-Taguchi system (MTS); fault detection

0 引言

旋转机械被广泛运用到生产和生活中,在航空航天、轨道交通、机械制造等领域发挥着重要的作用。在服役过程中,旋转机械会经历从正常到性能降低再到完全失效的一系列不同的性能退化的状态。滚动轴承作为旋转机械的关键部位,其性能的退化或失效将影响到整个机

械,并有可能导致机械非计划性停机,造成重大的经济损失^[1]。如果能及时检测滚动轴承初期故障,有针对性的对滚动轴承进行维修或更换,将大大降低成本并提高生产效率。因此,滚动轴承初期故障检测对旋转机械运行和维护具有重要意义^[2-3]。

随着物联网和智能传感器的发展,基于振动信号数据的滚动轴承故障检测方法逐渐成为热点。该方法主要是利用统计特征分析、快速傅里叶变换(fast Fourier

transform, FFT)、Hilbert 变换等信号处理方法提取振动信号的时域、频域和时-频域特征,将其作为特征向量进行故障诊断与评估^[4]。王玉静等^[5]提出了一种优化的集合经验模态分解(ensemble empirical mode decomposition, EEMD)特征向量提取方法;赵光权等^[6]利用小波包分解的能量熵特征对滚动轴承进行故障诊断;徐统等^[7]用变分模态分解(variational mode decomposition, VMD)方法来提取故障特征,并运用概率神经网络实现故障模式识别。尽管这些数据从不同的层面反映了轴承的运行状态,但提取的特征数据不可避免的呈现高维、高冗余性等特点,难以直接利用与挖掘。这些冗余特征在数据集成阶段很难得以消除,增加了数据处理的难度。此外,这些数据中存在大量对轴承故障不敏感的特征,如将其保留,不仅对故障检测无太大作用,而且还会引发“维数灾难”,同时还会降低某些对维数敏感的故障检测算法的精度。因此,有必要对提取的故障特征进行筛选,有效降低特征数据的冗余性与无关性,提高故障检测的准确度。

故障特征的筛选问题属于特征选择(或维数约简)问题,该过程可以降低故障检测算法的复杂度,节省时间并提高算法效率。根据维数约简的原理,可将特征选择算法分为线性特征选择和非线性特征选择两种。主成分分析(principal components analysis, PCA)和独立成分分析(independent component analysis, ICA)等是典型的线性维数约简方法,在线性特征数据提取方面效果显著。当面对轴承故障的非线性特征时,这些方法往往很难发挥效用。相比线性特征选择算法,非线性特征选择方法能够保持数据间的局部特性,有利于提高数据理解性、提升故障诊断的准确率。目前,非线性特征选择算法,如等距映射(isometric feature mapping, ISOMAP)、拉普拉斯特征映射(Laplacian eigenmaps, LE)和局部线性嵌入(locally-linear embedding, LLE)等,已经成功地运用到轴承故障特征选择上^[8]。拉普拉斯分值(Laplacian score, LS)是在LE算法基础上提出的非线性特征选择方法,它结合特征的局部信息保持能力和 Variant 分值来进行特征选择,在轴承故障选择方面也得到了广泛的应用^[9-10]。该算法通过独立地评价每一个特征维度,并根据评价指标的得分高低来筛选特征,但也存在如下不足:1)该方法只给出了特征排序结果,无法给出具体筛选的特征数目;2)特征集中存在冗余时,相互冗余的特征会得到相近的得分,使得筛选出的特征都是具有相似局部特性的冗余特征。

马田系统(Mahalanobis-Taguchi system, MTS)是由 Taguchi 在质量工程学基础上发展起来的多变量模式识别技术。与其他的模式识别技术相比,MTS 原理更简单、运算速度更快,不仅可以在不依靠数据分布的情况下处理大样本数据,而且能利用质量工具筛选重要特征,成为评价决策、医疗诊断和质量控制等过程中关键指标筛选

的方法之一。在滚动轴承的健康管理方面,MTS 方法也得到了运用^[11-13]。但是,该方法在轴承故障特征选择方面也会面临挑战:特征维数过高会导致 MTS 无法使用,特征之间出现冗余或者具有较高相关性时 MTS 识别的准确率不高。

因此,针对 LS 和 MTS 方法的不足,结合滚动轴承故障特征数据高维、高冗余的特点,提出基于 LS 和 MTS 的两阶段特征选择算法,对故障特征进行降维处理,使得降维后的特征具有较低的冗余性和较高的故障敏感性,为更准确的故障检测提供数据基础。

1 问题描述

1.1 滚动轴承故障特征分析

目前,对滚动轴承的故障预测和健康管理(PHM)研究大多集中在滚动轴承故障模式识别,而对滚动轴承故障检测的研究较少。滚动轴承是旋转机械的关键部件,故障发生的频率高,维修代价高。因此,有必要对轴承进行故障检测,在故障初期对轴承进行视情维修,提高轴承工作效率。

滚动轴承的振动信号属于非线性、非平稳信号,故障特征提取方法较多,主要集中在时域分析、频域分析和时-频域分析等方法上。时域特征分析主要是利用统计分析方法计算振动信号的特征统计特征,包括均值、均方根、峭度、峰值、波形因子、峰值因子、脉冲因子等,频域特征和时-频域特征分析则是采用 FFT、Hilbert 变换和 Hilbert 包络谱等方法计算得出^[9]。多种不同维度的特征提取方法相结合,可以使不同特征之间相互弥补不足,形成高维特征集。但不可避免的是,在特征提取的过程中,各特征参数之间的关联性使得高维数据中存在无关特征和冗余特征,降低了故障检测方法的效率,不利于实际运用。

因此,本文针对提取的滚动轴承故障特征数据的特点,采用特征选择方法对数据进行特征筛选,去除冗余特征,挖掘对故障敏感的重要特征。

1.2 故障特征选择问题

特征选择的关键是在全部特征上选择其中的理想子集使该子集具有最小冗余性和最大相关性^[14],该过程可以用数学模型进行描述:

假设 $F = \{f_1, f_2, \dots, f_m\}$ 为 m 维特征集合,求解 F 的子集 $S = \{f_1, f_2, \dots, f_k\}$ ($k < m$),使得:

$$\min R(S) \text{ 且 } \max D(S) \quad (1)$$

式中: $R(S)$ 为 S 的冗余性,指的是所有特征间的相关性,包括线性相关性与非线性相关性,而 $D(S)$ 为 S 的相关性,指单个特征对于反映数据样本局部结构作用的大小。

冗余性与相关性是两个不冲突的性质,可以利用两阶段的特征选择方法。第 1 阶段主要消除原始数据集中的非线性冗余特征,对相关特征进行初步筛选,第二阶段对特征进行综合评价,筛选相关特征,既能有效避免冗余性对于相关性筛选的影响,又能避免在第 2 阶段使用复杂的优化算法。因此,本文提出两阶段特征选择方法实现这一思路,得到具有最小冗余性与最大相关性的特征子集 S 。

1.3 马氏距离度量与故障检测

距离度量可用来表征两个样本之间的差异度,距离越小,说明两样本之间的差异就越小,反之亦然。马氏距离 (Mahalanobis distance, MD) 是由印度统计学家 Mahalanobis 提出的一种协方差距离,是一种有效的计算样本与样本集“重心”的最近距离或计算两个未知样本集的相似度的方法。其定义如下:

$$MD_{ij}^2 = (x_i - x_j)^T A^{-1} (x_i - x_j) \quad (2)$$

式中: $x_i \in \mathbf{R}^d$ 表示 d 维空间中的向量; A^{-1} 为 d 维空间的协方差矩阵的逆。与欧氏距离相比, MD 不受量纲的影响,同时还能考虑特征之间的联系,排除变量间的相关性的干扰,能够很好的度量观测样本和已知样本之间的距离,很适合用在故障检测上^[15]。

Taguchi 在 MTS 中根据马氏距离定义了如下的标准化马氏距离^[11]:

$$MD_{ij}^2 = \frac{1}{d} (x_i - x_j)^T A^{-1} (x_i - x_j) \quad (3)$$

使得正常样本的 MD 值在 1 附近,而异常样本的 MD 值随着样本与正常类样本的差的增加而变大。不同的 MD 值对应不同的设备运行特征,可通过比较异常样本到正常样本的 MD 值来定性诊断设备的故障。因此,本文将利用 MTS 算法计算的 MD 值构建滚动轴承故障检测模型,以此来监测轴承的运行状态。

2 基于 LS 的第 1 阶段特征选择

2.1 LS 方法

LS 是一种无监督特征选择方法,利用得分表征邻近样本点在某个特征的差异。LS 主要是利用局部保持能力来衡量特征,通过直接对特征集进行学习提取数据内在的信息结构。在初始特征空间中选取分值较小的特征参数,极大程度地保留了振动信号中含有的结构特征,从而有利于滚动轴承故障检测^[9-10]。

令 L_r 为第 r ($r = 1, 2, \dots, l$) 个特征的拉普拉斯分值, LS 特征选择步骤如下。

1) 用 m 个样本点构建近邻图 $G = (V, E)$ 。如果两个样本 x_i 和 x_j 足够近,将两个样本通过边 $(x_i, x_j) \in E$ 相

连。可以取 x_i 的 k 近邻点,建立近邻图。

2) 为了衡量近邻样本点之间的距离,定义 G 的邻接矩阵 $W = (W_{ij})_{m \times m}$:

$$W_{ij} = \begin{cases} e^{-\|x_i - x_j\|^2 / \sigma}, & (x_i, x_j) \in E \\ 0, & (x_i, x_j) \notin E \end{cases} \quad (4)$$

式中: σ 为常数; $\| \cdot \|$ 表示两个样本点之间的欧氏距离。 W_{ij} ($0 \leq W_{ij} \leq 1$) 值越大,说明两个样本点越相近。

3) 定义图 G 的拉普拉斯矩阵 L 为:

$$L = D - W \quad (5)$$

式中: $D = \text{diag}(WI)$ 为对角阵; $I = (1, 1, \dots, 1)^T$ 。

4) 计算第 r 个特征的拉普拉斯得分 L_r 。

$$L_r = \frac{\sum_{ij} (f_{ir} - f_{jr})^2 W_{ij}}{\text{Var}(f_r)} = \frac{\tilde{f}_r^T L \tilde{f}_r}{\tilde{f}_r^T D \tilde{f}_r} \quad (6)$$

$$\tilde{f}_r = f_r - \frac{f_r^T D I}{I^T D I} I$$

式中: \tilde{f}_r 为第 r 个特征的向量值 f_r 去均值化后的结果; $\text{Var}(f_r)$ 为第 r 个特征的方差。

5) 根据每个特征的 L_r , 选取前 t 个 L_r 值较小的特征参数构建评估指标, $t \leq l$ 。事实上,式(4)的分子和分母分别表示特征参数的局部信息保持能力和区分能力,在特征选择时需要选择局部信息保持能力好同时区分能力越大的特征信息,即要求分子越小越好,同时分母越大越好。所以,计算所得的 L_r 与特征重要性成反比,即 L_r 越低,说明该特征对特征选择越重要。

利用 LS 可以将多维特征按照重要性进行排序。

2.2 互信息 (MI) 原理

MI 是信息论中的一种信息度量,它可以看成是某一个随机变量中包含的关于另外一个随机变量的信息量,或者说是某一随机变量由于已知另一随机变量而减少的不确定性。

设两个随机变量 (X, Y) 的联合分布为 $p(x, y)$, 边际分布分别为 $p(x)$ 、 $p(y)$, 则 (X, Y) 的互信息即为 $p(x, y)$ 与 $p(x)p(y)$ 的相对熵,用 $I(X; Y)$ 表示,即:

$$I(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \quad (7)$$

MI 可以有效度量特征集中变量间的冗余性。

2.3 LS-MI 特征选择算法

利用 LS 方法可以准确鉴别无关的和性能较差的特征,但是不能去除冗余特征。为克服这一缺点,选取最优特征子集,本文提出基于 LS 和 MI 的组合式方法 (LS-MI) 筛选特征子集,具体算法描述如下。

输入: 训练数据集 D 。

- 1) 计算 D 中任意两个特征的互信息 I_{ij} ;
- 2) 计算 D 中所有特征的 LS 值 L_r ;

3) 根据 I_{ij} 进行聚类, 得到 k 个特征子集;

4) 根据聚类结果, 在每个特征子集中挑选 L_i 值最大的特征作为该类特征的代表特征, 共 k 个特征构成特征子集 D_k , 即为筛选结果。

输出: 特征子集 D_k 。

经过第1阶段LS-MI算法后, 高相关性特征中LS较高的特征得以保留, 非线性冗余特征被基本删除。在剩余无冗余特征中进行MTS特征评价, 可以很大程度上避免重复选择具有相近相关性的特征。

3 基于MTS的第2阶段特征选择方法

在MTS中, Taguchi综合利用正交表(orthogonal arrays, OA)和信噪比(signal to noise ratio, SNR)来筛选重要特征, 具体步骤如下:

阶段1 基准空间构建

该阶段是MTS的基础, 主要是收集特征数据, 确定正常、异常样本并计算MD值, 具体如下:

1) 收集正常和异常样本数据构建初始特征集。

2) 计算正常样本的MD值。

设 $\mathbf{x}_i (i = 1, 2, \dots, n)$ 为 n 个包含 p 维特征的样品, $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})$, x_{ij} 为该样品的第 j 个变量的观测值, 则该样品的马氏距离平方(马氏距离)为:

$$MD_i = \frac{1}{p} \mathbf{Z}_i^T \mathbf{S}^{-1} \mathbf{Z}_i \quad (8)$$

式中: \mathbf{S} 表示样本的协方差矩阵; $\mathbf{Z}_i = [z_{i1}, z_{i2}, \dots, z_{ip}]^T$ 为样品标准化后的数据, $z_{ij} = (x_{ij} - \bar{x}_j) / s_j$, \bar{x}_j 和 s_j 分别为第 j 个变量的均值和标准差。

阶段2 基准空间有效性验证

在优化之前要对基准空间的有效性进行验证。

3) 利用正常样本特征均值和标准差对异常样本进行标准化处理并计算异常样本的MD值。若MD均明显大于基准空间中正常样本的MD, 则认为基准空间有效, 反之, 则需重新选择特征变量构建基准空间。

阶段3 基准空间优化

Taguchi认为并非所有的特征对分类都有贡献, 可以利用OAs和SNR对其进行筛选, 挑选出其中的重要特征作为分类依据。

4) 根据变量个数设计OAs, 将每个特征安排在正交表的每一列, 用“1”和“0”分别表示该变量是否参与构建基准空间, 根据每一行(一次试验)被选择的特征构建特征空间, 计算异常样品的马氏距离 $MD_k, k = 1, 2, \dots, m$ 。

5) 计算每次试验的SNR:

$$\eta = -10 \log \left(\frac{1}{m} \sum_{k=1}^m \frac{1}{MD_k} \right) \quad (9)$$

6) 计算变量 x_j 的信噪比增量:

$$\Delta(j) = \bar{\eta}^+ - \bar{\eta}^- \quad (10)$$

式中: $\bar{\eta}^+$ 和 $\bar{\eta}^-$ 分别为 x_j 参与试验的SNR均值和未参与试验的SNR均值; $\Delta(j) > 0$ 说明变量可用来反映样本间的差异, 应该保留, 反之, 应该删除该特征。

阶段4 利用优化的系统进行分类或预测

优化后的特征可以作为分类或预测的重要依据, 可利用其计算MD, 进行分类和预测。

7) 利用优化后的基准空间计算正常和异常样本的MD值, 确定分类阈值, 对待测样品进行分类。

在第1阶段特征选择的基础上, 运用MTS能够更好地筛选出与分类有关的关键特征, 同时可以计算基准空间的MD值, 为后续故障检测和退化评估打下基础。

4 滚动轴承初期故障检测模型

4.1 故障检测模型

利用两阶段特征选择算法得到故障敏感特征后, 如何制定判别准则是故障检测的关键。在MTS中, Taguchi根据“经济损失”与“功能波动”之间的函数关系, 提出利用质量损失函数法(quality loss function, QLF)确定分类阈值, 以此作为区分正常样本和异常样本的依据。但是, QLF方法缺乏合理的统计解释, 主要依据领域内知识确定阈值, 具有一定的主观性, 而在实际应用中由难以确定每个样本的相对成本或损失。针对这种情况, SU^[16]等提出的概率阈值方法, 利用切比雪夫定理设计阈值 T :

$$T = u_{MD} + \sqrt{\frac{1}{1 + \lambda - \omega}} s_{MD} \quad (11)$$

式中: u_{MD} 和 s_{MD} 是正常样本的马氏距离均值和标准差; λ 为小参数, 一般取0.05; ω 表示正常样本集中马氏距离小于异常样本马氏距离最小值的比例。

为判断轴承是否出现故障, 利用优化后马氏空间计算待测样本的MD值, 将其与故障阈值 T 进行对比, 若待测样本的马氏距离 $MD_i \leq T$, 则属于正常样本; 反之, 若 $MD_i > T$, 则属于故障样本。

4.2 故障检测流程

根据滚动轴承故障检测框架和本文设计的两阶段特征选择方法, 设计如图1所示故障检测流程。

5 实验及结果分析

5.1 实验设置

本文利用美国辛辛那提大学智能维护系统中心(intelligent maintenance systems, IMS)的轴承退化仿真试验数据来验证所提方法的有效性, 其试验装置如图2所示。试验中共有4个同型号轴承, 轴承1和4固定在机

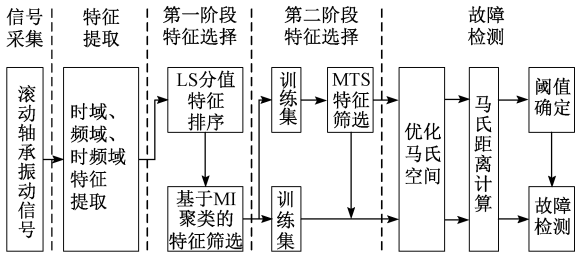


图 1 滚动轴承故障检测流程

Fig. 1 Flowchart of rolling bearing fault detection

体上,轴承 2 和 3 通过杠杆机构向主轴施加恒定径向载荷。试验过程中,电机驱动主轴以 2 000 r/min 的转速持续运行,在轴承端设置的传感器以 20 480 Hz 的采样频率每隔 10 min 采集一次振动信号,并通过磁塞不断收集润滑油中的金属屑用来反映轴承的退化程度,当吸附的金属屑积累到一定水平时,试验结束。

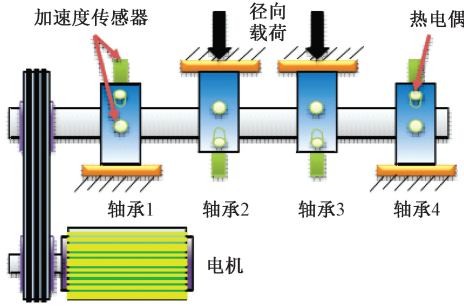


图 2 轴承退化仿真试验平台

Fig. 2 Test platform of bearing degradation simulation

5.2 信号处理与分析

本文选取第 2 个轴承数据集,该过程持续了 164 h,共得到 984 组信号,每组信号包含 20 480 个数据点,解体分析后发现 1 号轴承外圈出现严重剥落。

提取振动信号的 12 个时域特征、13 个频域特征和 Hilbert 包络谱的 13 个频域特征,构建特征集 FS :

$$FS = \begin{bmatrix} FS_1 \\ \vdots \\ FS_k \\ \vdots \\ FS_n \end{bmatrix} = \begin{bmatrix} TF_{1 \times 12} & FF_{1 \times 13} & HF_{1 \times 13} \\ \vdots & \vdots & \vdots \\ TF_{k \times 12} & TF_{k \times 13} & TF_{k \times 13} \\ \vdots & \vdots & \vdots \\ TF_{n \times 12} & TF_{n \times 13} & TF_{n \times 13} \end{bmatrix}_{n \times (12+13+13)}$$

式中: n 为样本总数; FS_k 为第 k 个样本的多域特征集, FS 是 $n \times 38$ 为的特征矩阵; $TF_{k \times 12}$ 、 $FF_{k \times 13}$ 和 $HF_{k \times 13}$ 分别表示第 k 个样本的时域特征集、频域特征集和 Hilbert 包络谱的频域特征。

最终得到 984 个样本的 38 个特征,按样本的先后顺序进行排序,得到如图 3 所示的特征数据。

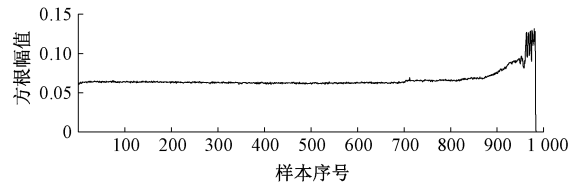
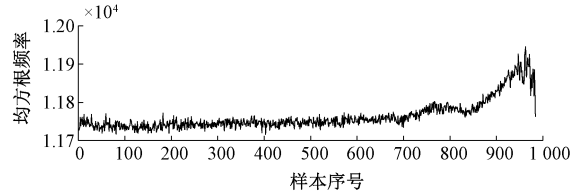


图 3 轴承 1 全寿命周期部分特征数据

Fig. 3 Features of bearing 1 during whole lifetime



5.3 特征选择分析应用

利用 LS 得分算法和互信息计算公式分别计算特征的 LS 得分和特征间的互信息 MI,并根据互信息进行特征聚类,其结果如图 4 所示。

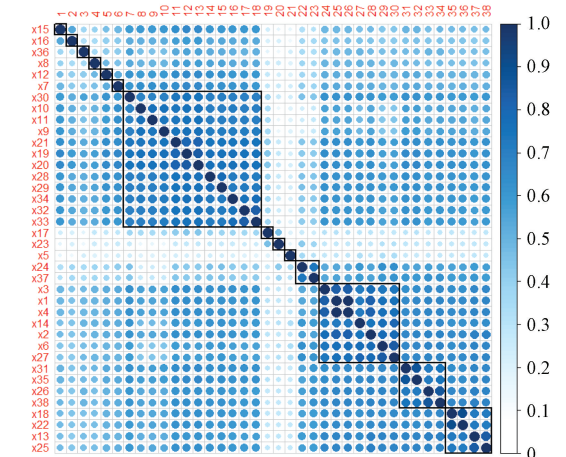


图 4 根据特征互信息聚类结果

Fig. 4 Results of cluster based on Mutual Information

利用第 1 阶段特征选择算法最终筛选出 $F29$ 、 $F37$ 、 $F23$ 、 $F5$ 、 $F16$ 、 $F36$ 、 $F17$ 、 $F35$ 、 $F8$ 、 $F12$ 、 $F1$ 、 $F15$ 、 $F18$ 和 $F7$ 这 14 个特征。

考虑到轴承初期运行不稳定,选取数据集中第 61~260 个样本作为正常样本,第 851~900 个样本作为异常样本共同构成训练集,剩余样本作为测试集。根据 MTS 相关步骤,计算验证基准空间的有效性。通过验证发现,正常样本的马氏距离在 1 附近,而异常样本马氏距离的最小值都远超过 1,说明所建立的基准空间是有效的。

根据 SNR 和 OAs 计算特征的信噪比增量,如图 5 所示。

由图 5 可知,除 $F23$ 、 $F36$ 和 $F18$ 以外,其余特征的信噪比增量明显大于 0,因此,选取剩余 11 个特征作为优化

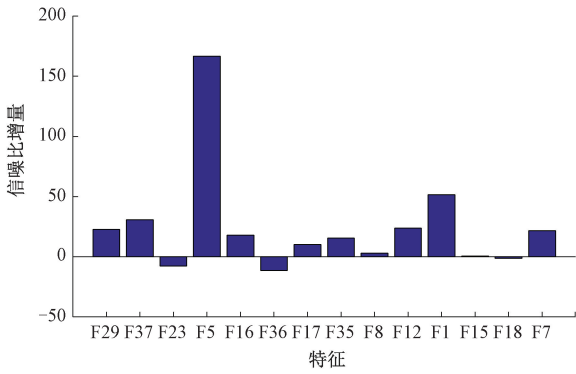


图5 SNR和OAs筛选结果

Fig. 5 Results of features selection based on SNR and OAs

后特征,构建优化基准空间,计算训练集中各样本的马氏距离值MD和阈值T,得到T=3.203。

计算全寿命周期样本的马氏距离,前934个样本的马氏距离和局部马氏距离图像如图6所示(由于从935~984个样本的马氏距离值变化较大,不利于观察变化趋势)。

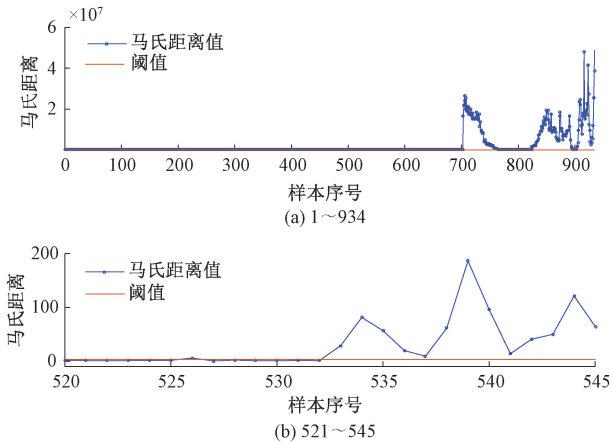


图6 全寿命周期样本的马氏距离值

Fig. 6 MDs of the full life cycle samples

由图6可知,滚动轴承在运行到第533个周期时开始出现初期故障;到第701个周期以后,轴承的工作状态恶化,故障程度加深,直到最终失效。根据检测结果,如果能在发现初期故障阶段(第533~700个周期内)实施维护策略,就可以延长轴承的使用寿命,避免经济损失。

5.4 评价与对比

由于方法的目的是去除数据集冗余性、筛选重要特征并提升故障检测的精度,为此,本文选取剩余冗余度(RED)和检测准确率(ACC)来评价两阶段特征选择方法的有效性。

数据集的剩余冗余度(RED)可根据剩余特征集中特征的平均互信息来确定:

$$RED(S) = \frac{1}{|S|^2} \sum_{f_i, f_j \in S} I(f_i; f_j) \quad (12)$$

式中:S为剩余特征集;I(f_i;f_j)为第i个特征和第j个特征间的互信息。

ACC由剩余特征集的灵敏度(Se)和特异性(Sp)来确定,其公式如下:

$$Se = \frac{TP}{TP + FN}, Sp = \frac{TN}{TN + FP} \quad (13)$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

式中:TP表示预测正确的正常样本个数;TN表示预测正确的异常样本个数;FP表示预测错误的正常样本个数;FN表示预测错误的异常样本个数。

分别计算全部特征集F、第1阶段特征选择后特征集S₁和文献[9]的LS方法选择同样数目特征集S₂,计算剩余冗余度,结果如表1所示。

表1 特征选择的剩余冗余度结果

Table 1 RED of feature selection

方法	特征个数	RED
全部特征	38	0.555 6
LS-MI	14	0.437 9
LS ^[9]	14	0.621 6

结果显示,第1阶段特征选择后特征集的剩余冗余度明显低于全部特征集的剩余冗余度,说明该方法可以有效的降低特征冗余。通过与文献中方法对比发现本文方法在去除特征冗余方面要优于单独使用LS方法,而且效果明显。

分别利用第1阶段特征选择算法(LS-MI)和两阶段特征选择算法(LS-MI-MTS)计算各样本的马氏距离,确定阈值,结果如图7所示。

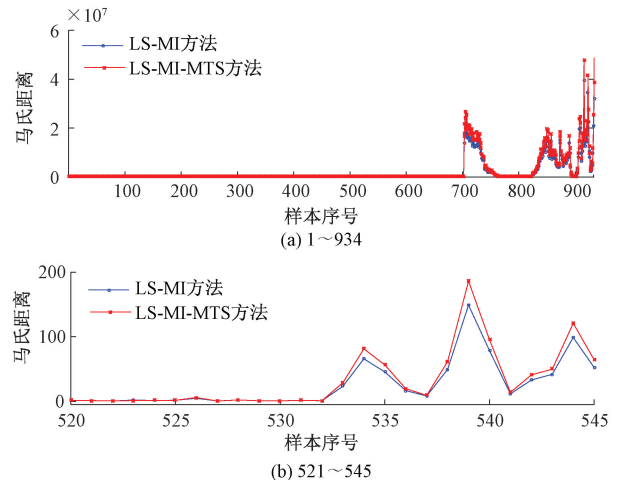


图7 不同特征选择方法结果对比

Fig. 7 Comparison of the feature selection methods

由图 7 可知,两种方法计算的马氏距离变化趋势相同,都是从第 533 个样本点开始 MD 值超过阈值,但是利用两阶段特征选择算法计算的 MD 值变化趋势更明显,更有利于故障检测。

根据图 7 的结果,在全寿命周期中,前 532 样本属于正常状态,后 452 个样本属于故障状态。根据阈值和 MD 值,对测试集样本进行检测,计算不同特征选择算法的故障检测的准确度,结果如表 2 所示。

表 2 特征选择的检测准确率结果

Table 2 ACC of feature selection

方法	特征个数	阈值	ACC
LS-MI	14	2.879	0.975 5
LS-MI-MTS	11	3.203	0.976 8

由表 2 结果可知,两阶段特征选择方法不仅具有更高的特征选择效率,同时还能提高故障检测的准确率。所以,两阶段特征选择效果要优于单独使用第一阶段特征选择方法。

为说明本文方法在滚动轴承早期故障检测的优势,将本文方法和文献[13]使用的 MTS 方法对比发现,文献[13]方法检测到滚动轴承出现故障的时间为第 670 个周期,比本文方法检测到故障的时间晚。因此,本文方法可以更早的识别轴承的故障,更有利于滚动轴承的视情维修。

6 结 论

针对 PHM 中提取的故障特征数据高维、高冗余性的特点,以去除数据冗余性、筛选敏感特征、改善故障检测效果为目的,提出了两阶段特征选择算法。基于 LS 算法可以对特征按重要的进行排序的特性,提出利用互信息聚类将每个特征子集中 LS 最高的特征筛选出来作为代表,得到了低冗余性的特征子集。结合 MTS 算法特征筛选的优势,在剩余子集中筛选出其中的重要特征,从而达到改善故障检测效果的目的。通过实验平台数据上验证,并与 LS 和 MTS 算法对比,验证了本文方法去除冗余性的能力更强,检测效果更优,对初期故障检测更有效。

参考文献

[1] 阙子俊,金晓航,孙毅. 基于 UKF 的轴承剩余寿命预测方法研究[J]. 仪器仪表学报, 2016, 37(9): 2036-2043.
 QUE Z J, JIN X H, SUN Y. Remaining useful life prediction for bearings with the unscented Kalman filter based approach [J]. Chinese Journal of Scientific Instrument, 2016, 37(9): 2036-2043.

[2] 黄良沛,吴超威,王靖. 小波包分析和 BP 神经网络在滚动轴承故障模式识别中的应用[J]. 电子测量技术, 2016, 39(4): 164-168.
 HUANG L P, WU CH W, WANG J. Fault pattern recognition of rolling bearing using wavelet package analysis and BP neural network [J]. Electronic Measurement Technology, 2016, 39(4): 164-168.

[3] 陈运胜. 发电机传动轴承的异常振动谱特征提取算法[J]. 国外电子测量技术, 2016, 35(5): 20-38.
 CHEN Y SH. Abnormal vibration spectrum feature extraction algorithm for generator drive bearing [J]. Foreign Electronic Measurement Technology, 2016, 35(5): 20-38.

[4] YAN X A, JIA M P. A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing [J]. Neurocomputing, 2018, 313: 47-64.

[5] 王玉静,姜义成,康守强,等. 基于优化集合 EMD 的滚动轴承故障位置及性能退化程度诊断方法[J]. 仪器仪表学报, 2013, 34(8): 1834-1840.
 WANG Y J, JIANG Y CH, KANG SH Q, et al. Diagnosis method of fault location and performance degradation degree of rolling bearing based on optimal ensemble EMD [J]. Chinese Journal of Scientific Instrument, 2013, 34(8): 1834-1840.

[6] 赵光权,姜泽东,胡聪,等. 基于小波包能量熵和 DBN 的轴承故障诊断[J]. 电子测量与仪器学报, 2019, 33(2): 32-38.
 ZHAO G Q, JIANG Z D, HU C, et al. Bearing fault diagnosis based on wavelet packet energy entropy and DBN [J]. Journal of Electronic Measurement and Instrumentation, 2019, 33(2): 32-38.

[7] 徐统,王红军,宋智勇,等. 基于 K-L 散度得 VMD 瞬时能量与 PNN 得滚动轴承故障诊断[J]. 电子测量与仪器学报, 2019, 33(8): 117-123.
 XU T, WANG J H, SONG ZH Y, et al. Rolling bearing fault diagnosis using VMD energy feature and PNN based on Kullback-Leibler divergence [J]. Journal of Electronic Measurement and Instrumentation, 2019, 33(8): 117-123.

[8] LU L, YAN J H, DE SILVA C W. Dominant feature selection for the fault diagnosis of rotary machines using modified genetic algorithm and empirical mode decomposition [J]. Journal of Sound and Vibration, 2015, 344: 464-483.

[9] 欧璐,于德介. 基于监督拉普拉斯分值和模糊 C 均值聚类的滚动轴承故障诊断[J]. 中国机械工程, 2014, 25(10): 1352-1357.

- OU L, YU D J. Rolling bearing fault diagnosis based on Laplaian score and fuzzy C-means clustering [J]. China Mechanical Engineering, 2014, 25(10):1352-1357.
- [10] ZHANG J D, PAN H Y, YANG S B, et al. Generalized composite multis-scale permutation entropy and Laplacian score based rolling bearing fault diagnosis [J]. Mechanical Systems and Signal Processing, 2018, 99: 229-243.
- [11] WANG Z P, CHEN L, WANG Z L, et al. Fault diagnosis and health assessment for bearings using the Mahalanobis-Taguchi system based on EMD-SVD [J]. Transactions of the Institute of Measurement and Control, 2013, 35(6):78-807.
- [12] SHAKYA P, KULKARNI M S, DARPE A K. Bearing diagnosis based on Mahalanobis-Taguchi-Gram-Schmidt method [J]. Journal of Sound and Vibration, 2015, 337: 342-362.
- [13] CHEN J X, CHENG L S, YU H, et al. Rolling bearing fault diagnosis and health assessment using EEMD and the adjustment Mahalanobis-Taguchi system [J]. International Journal of Systems Science, 2018, 49(1): 147-159.
- [14] ZHAO Z, WANG L, LIU H. Efficient spectral feature selection with minimum redundancy [C]. Twenty-Fourth AAAI Conference on Artificial Intelligence, 2010.
- [15] 陈海永,徐森,刘坤,等. 基于 Gabor 小波和加权马氏距离的带钢表面缺陷检测 [J]. 电子测量与仪器学报, 2016, 30(5):786-793.
- CHEN H Y, XU S, LIU K, et al. Strip defect detection based on Gabor wavelet and weighted Mahalanobis distance [J]. Journal of Electronic Measurement and Instrumentation, 2016, 30(5): 786-793.
- [16] SU C T, HSIAO Y H. Muti-class MTS for simultaneous feature selection and classification [J]. Transactions on Knowledge and Data Engineering, 2009, 21(2): 192-205.

作者简介



彭宅铭, 2016 年于安徽工业大学获得硕士学位, 现为南京理工大学经济管理学院博士研究生, 主要研究方向为质量管理和质量工程、应用统计建模。

E-mail: pz_ming@njust.edu.cn

Peng Zhaiming received his M. Sc. degree from Anhui University of Technology in 2016. Now he is a Ph. D. candidate at Nanjing University of Science and Technology. His main research interests include quality management and quality engineering, applied statistical modeling.