

DOI: 10.13382/j.jemi.B2104385

mRMR 特征筛选和随机森林的故障诊断方法研究*

常梦容 王海瑞 肖 杨

(昆明理工大学信息工程与自动化学院 昆明 650500)

摘要:针对滚动轴承原始振动信号重要特征信息被较强背景噪声淹没以及提取的时域特征冗余度较高、相关性较强的缺点,提出一种基于最大相关-最小冗余(max-relevance and min-redundancy, mRMR)特征筛选和随机森林的滚动轴承故障诊断研究方法。首先将原始信号进行自适应噪声完整集成经验模态分解(CEEMDAN)得到一系列固有模态分量(IMFs),分析IMF并去掉高频噪声和一部分虚假分量,再将信号进行重构并提取其时域特征,通过mRMR去除冗余性和相关性较高的特征向量,使筛选出的特征子集与标签有最大的依赖性,最后将该特征子集输入到随机森林分类器进行分类。实验表明,mRMR具有优良的特征搜索策略,重要特征均靠前得到选取,仅需3个特征便能达到较高的分类准确率,效率高于其余特征选择算法。

关键词:特征筛选;时域特征;mRMR;冗余性;相关性

中图分类号: TH133.33; TN06 **文献标识码:** A **国家标准学科分类代码:** 510.4010

Research on fault diagnosis method based on mRMR feature screening and random forest

Chang Mengrong Wang Hairui Xiao Yang

(College of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China)

Abstract: Aiming at the shortcomings that the important feature information of the original vibration signal of the rolling bearing is submerged by strong background noise, and the extracted time domain features have high redundancy and strong relevance, this paper proposes a new rolling bearing fault diagnosis research method based on maximum relevance-minimum redundancy (mRMR) feature selection and random forest. First, the original signal is subjected to complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) to obtain a series of intrinsic modal functions (IMFs), analyze IMF and remove high frequency noise and part of false component, then reconstruct the signal and extract its time domain characteristics, mRMR is used to remove redundant and highly correlated feature vectors, so that the selected feature subset has the greatest dependence on the label, and finally the feature subset is input to the random forest classifier for classification. Experiments show that mRMR has an excellent feature search strategy, the important features are selected first. Only three features are needed to achieve a higher classification accuracy, and the efficiency is higher than other feature selection algorithms.

Keywords: feature selection; time domain feature; mRMR; redundancy; relevance

0 引言

滚动轴承是旋转机械中不可缺少的一部分,在承受旋转体的重量和工作负荷的同时,能保证旋转体的旋转精度。因此,滚动轴承能够为旋转机械提供可靠、稳定的支撑,其运行状态对旋转机械的性能有很大的影响。如

果轴承失效,可能会造成严重的事故和大量的财产损失。因此,滚动轴承的状态检测和故障诊断具有重要意义。然而噪声、传输路径复杂、信号衰减等因素的影响,振动信号中包含的故障信息通常非常微弱,这使得故障诊断极为困难。

因此,为了更好地分析非线性非平稳信号的特征、解决原始信号中存在的大量噪声致使故障特征提取不易的

问题, 研究人员提出了经验模态分解 (empirical mode decomposition, EMD) 方法。EMD 可以在没有任何先验知识的情况下, 根据输入信号的特征自适应地将信号分解为一些固有模态函数 (IMF) [1]。它被认为是近两百年来基于傅里叶变换 (FT) 的线性和稳态谱分析的突破, 被广泛应用于机器故障诊断、疾病检测、经济分析等诸多方面。然而, 当高频分量在时域存在间歇性时, 该方法的结果将出现“模态混叠”。即一个单一的 IMF 包含频率差异非常大的信号, 或者将同一频率的信号分解成完全不同的 IMF。随之, Zhao 等 [2] 在 EMD 分解的每个阶段加入自适应噪声, 提出 CEEMDAN 方法, 通过计算每个 IMF 的比裕度得到每个 IMF。自适应噪声完整集成经验模态分解 (CEEMDAN) 有效地解决了上述问题, 重构误差几乎为零, 分解完全, 计算代价降低。

同时, 由于时域、频域以及时频域特征中的统计特性通常描述了整个设备关键部件的运行状态, 能够查看全部信号在每个频带中整个信号的分布和规律, 因此被进行广泛研究 [3]。但随着特征数量的增加, 容易产生维数灾难等问题, 并且原始特征集中包含了较多的无效特征, 相关性和冗余性问题较为突出。因此, 特征筛选 (feature selection) [4-5]、特征提取 (feature extraction) [6] 和特征降维 (feature dimension reduction) [7-8] 在故障诊断领域中显得尤其重要。特征提取后进行特征筛选, 一方面可以得到原始特征空间的低维表示, 以减少维数和信息冗余。另一方面从候选特征中选出“优秀”的特征, 可以提升模型性能以及提高轴承故障诊断的准确性和效率。近年来, 许多学者对这一领域进行了探索, 并提出了许多方法。Maleki 等 [9] 一种基于遗传算法的 k-最近邻技术应用于疾病分期诊断中, 通过有效的特征选择降低数据集的维数, 提高分类器的速度。颜弋凡等 [10] 提出基于最大互信息系数 (MIC) 属性选择的机械性能预测建模方法。该方法首先利用 MIC 算法计算各性能指标和工艺参数之间的相关性度量, 然后根据各相关度量选择形成工艺参数属性子集用于性能预测模型建模及预测。张俐等 [11] 通过最大相关系数和对称不确定性度量准则, 计算出每个特征与标签之间的相关度量值, 并按照数值大小顺序进行排序; 其次, 通过最大相关系数和近似马尔可夫毯式原理进行无关特征和冗余特征的筛选, 最终选择出最优特征子集。孙曙光等 [12] 针对机械振动环境下交流接触器电寿命预测中的失效特征量进行研究, 提出了一种基于优化小波去噪与核主成分分析-皮尔逊相关系数法相结合的方法以获取有效的失效特征量。Chen 等 [13] 提出一种基于 Pearson 相关系数的特征选择算法, 对多个特征进行特征规范化处理, 大大减少了需要处理的安全数据量, 有效降低了数据的维数, 增加了入侵性检测效率。

但上述方法在进行特征选择的同时, 并未考虑到实

际工况的复杂性, 使得提取特征向量并不能完全反应原始信号的实际规律、冲击信号被淹没, 特征提取存在较多误差, 导致选择效果被压缩, 并且提取出的时域、频域以及及时频特征中会产生许多无效特征向量对故障诊断效率造成较大影响。针对上述问题, 本文提出一种基于最大相关-最小冗余 (mRMR) 特征筛选和随机森林的故障诊断方法, 首先对原始信号进行 CEEMDAN 分解与重构, 剔除无效信息, 然后通过冗余性和相关性计算来验证 mRMR 特征提取的高效性和收敛性以及分类结果的影响。

1 CEEMDAN

CEEMDAN [14] 方法是基于 EMD 的。通过自适应加入白噪声, 克服了 EMD 方法的模态混叠问题, 获得了较好的模态分离谱, 同时提高了运算效率。

设 $E_k(\cdot)$ 为 EMD 得到的第 k 个 IMF 的产生算子, 设 $\omega(i)$ 为零平均值单位方差白噪声的实现。对于给定的信号 D , CEEMDAN 的主要步骤如下。

1) 对于信号 $D^i = D + \beta_0 \omega^{(i)}$, $i = 1, 2, \dots, N$, 通过 EMD 分解得到第 1 个 IMF $d_1^{(i)}$, 将 CEEMDAN 的第 1 个分量 $\overline{d_1}$ 定义为:

$$\overline{d_1} = \sum_{i=1}^N d_1^{(i)} \quad (1)$$

2) 计算第 1 个残差为:

$$r_1 = D - \overline{d_1} \quad (2)$$

3) 对于信号 $r_i + \beta_i E_i(\omega^{(i)})$, $i = 1, 2, \dots, N$, 通过 EMD 获得第 1 个 IMF。将 CEEMDAN 的第 2 个分量 $\overline{d_2}$ 定义为:

$$\overline{d_2} = \frac{1}{N} \sum_{i=1}^N E_1(r_1 + \beta_1 E_1(\omega^{(i)})) \quad (3)$$

4) 计算 KTH 残差为:

$$r_k = r_{(k-1)} - \overline{d_k} \quad (4)$$

其中 $k = 1, 2, \dots, K$ 。

5) 信号 $r_k + \beta_k E_k(\omega^{(i)})$, $i = 1, 2, \dots, N$, 通过 EMD 获得第 1 个分量, 将 CEEMDAN ($k+1$) 次 $\overline{d_{k+1}}$ 为:

$$\overline{d_{k+1}} = \frac{1}{N} \sum_{i=1}^N E_1(r_k + \beta_k E_1(\omega^{(i)})) \quad (5)$$

6) 下一个 k 到步骤 4)

重复步骤 4) ~ 6), 直到得到的残差不能被 EMD 分解, 要么满足 IMF 的条件, 要么其局部极值个数小于 3。最后的残差是:

$$r_1 = D - \sum_{k=1}^K \overline{d_k} \quad (6)$$

K 为 IMF 的总个数, 则信号 D 可以表示为:

$$D = \sum_{k=1}^K \overline{d_k} + r_k \quad (7)$$

该系数 $\beta_k = \varepsilon_k \text{std}(r_k)$ 允许在每次迭代中选择信噪比,其中 $\text{std}(\cdot)$ 为标准差算子。

2 随机森林 (RF)

RF^[15] 是一种基于决策树的集成学习技术。该方法利用大量样本数据集生成的多元分类回归树来确定每个样本属于哪个类别。每个决策树代表一个分类器,通过组合多个弱分类器,最终通过投票或取均值^[16],使得整体模型的结果具有较高的准确度和泛化能力,其构成如图 1 所示。因此 RF 模型的分类和预测能力优于基于传统分类和回归的单个决策树。

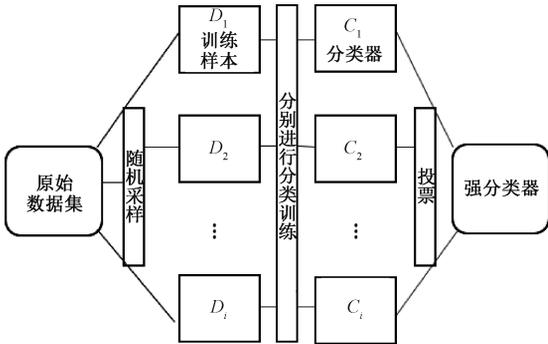


图 1 随机森林构成
Fig. 1 Random forest composition

在 RF 模型中,采用重采样 (bootstrap) 采样方法从原始样本集 T 中提取 n 个样本(一般为原始样本集的 $2/3$),生成一个新的训练样本集。为每个独立的训练样本生成相应的决策树,并从新的训练样本集中生成 n 个决策树构成森林。剩余的样本(通常是原始样本集的 $1/3$)称为袋外 (OOB) 样本,构成测试样本集,用于验证模型的泛化能力。计算林中所有决策树类别,并将样本获得最多投票数的类别表示为其类别,具体流程如图 2 所示。

其中 Bootstrap 基本原理如下:设集合 T 中有 n 个样本,若每次有放回地从该集中抽取 1 个样本,抽取 n 次形成的新集合 T^* 中不包含第 t 个样本的概率如式(8)。

$$p_t = \left(1 - \frac{1}{n}\right)^n \quad (8)$$

当 n 趋于无穷时,有:

$$\lim_{n \rightarrow \infty} p_t = \lim_{n \rightarrow \infty} \left(1 - \frac{1}{n}\right)^n = \exp(-1) \approx 0.368$$

可得,虽然 T^* 和 T 中样本数均为 n ,但 T^* 中包含了重复样本,且 T^* 中包含了约 63.2% 的样本。

RF 具有如下的特点:1)模型的随机性使得该模型不仅可以降低模型的方差,而且具有较好的泛化能力以及

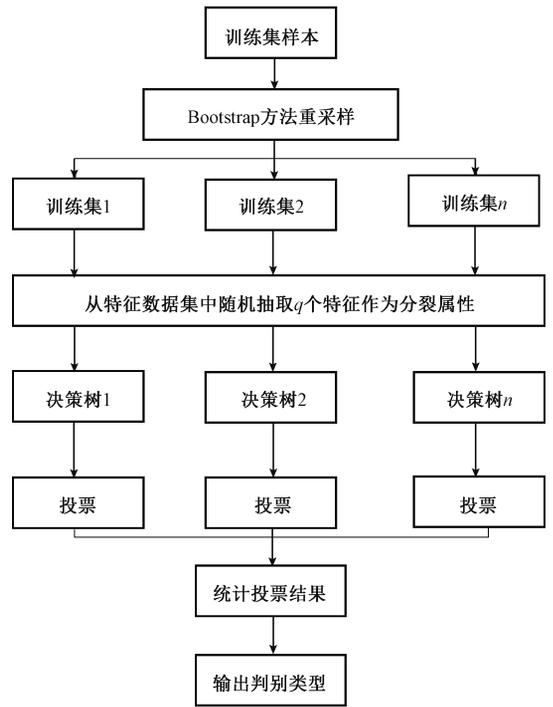


图 2 RF 算法流程
Fig. 2 RF algorithm flow

抗过拟合能力;2)每颗树的生成都是独立的,互不干扰并同时生成,使得该算法具有并行性;3)训练速度快,适用于大规模数据集中且未知情况;4)抗噪能力强,具有较强的工业价值;5)算法简单易于实现。

3 特征筛选

3.1 特征筛选的原因

滚动轴承提取出来的特征数量与分类模型之间存在着一定的关系即开始时特征数量不断地增加同时分类器性能也不断地提高,在某个位置分类器性能到达最优;然而当特征数量继续不断增加,分类器性能反而减少。如图 3 所示,表明特征数量过多或者过少都会对分类器性能产生巨大的影响^[17],因此在故障诊断中,当面临庞大的特征数量时,需要对这些特征进行组合优化。即判断哪些是相关特征,哪些不是相关特征,从候选特征中选出“优秀”的特征^[18],提升分类器性能以及提高轴承故障诊断的准确性和效率。

3.2 mRMR

在许多分类问题中,如何在相关特征中筛选一个有效的特征数据集是非常重要的^[19]。特征的筛选具有减少特征数量,减少处理时间和噪声、检测出更好、更显著的特征等优点。通常,特征的筛选要考虑到过滤方法或分类性能(包装器方法)。虽然特征筛选算法根据其分

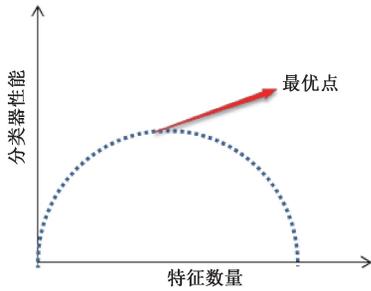


图 3 特征数量与分类模型性能关系

Fig. 3 Relationship between feature number and classification model performance

类性能得到了有效的结果,但它们往往需要大量的样本和相当长的时间。mRMR^[20]特征筛选算法根据特征与类标签的相关性列出特征。此外,它还旨在利用所选特征集来选择最不相关的特征向量。受益于互信息量 $I(X; Y)$ 的量来计算特征-特征和特征标签之间的相似性。

$$I(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)}$$

$$I(X, Y) = I(Y, X) \quad (9)$$

其中, $p(x, y)$ 为 x 和 y 的联合概率密度函数, $p(x)$ 、 $p(y)$ 分别为 x 和 y 的空间和边际概率密度函数。当两个随机变量完全独立时,互信息为 0。

在互信息定义的基础上,利用特征筛选方法 mRMR 得到内部特征间冗余度最小、稳定性结果相关性最大的最优特征子集。

对于数据集 $R = \{x_1, x_2, \dots, x_M | y\}$, 有 M 组输入特征,且具有稳定性标签向量 y , 假设 S 是 R 的一个子集,该子集的冗余度为:

$$V_s = \frac{1}{|S|^2} \sum_{i, j \in S} I(x_i; x_j) \quad (10)$$

式中: $|S|$ 为子集 S 中包含的特征组的个数。

子集 S 与目标向量 y 的相关性为:

$$W_s = \frac{1}{|S|} \sum_{i \in S} I(x_i; y) \quad (11)$$

以式(12)为优化目标,寻找冗余 V 更少、相关性 W_s 更强的输入特征的最优子集。

$$\max_s \left(\frac{W_s}{V_s} \right) \quad (12)$$

获取最优特征子集的计算负担是巨大的。因此, mRMR 技术采用增量搜索算法对所有特征组进行排序,然后选择最优的特征子集。具体流程如下。

1) 定义所选特征组的集合为 S 。

2) 计算每组输入特征 x_i 与目标 y 的相关性,然后根据式(13)选择与目标最相关的一组输入特征 $x^{(1)}$ 。将选定的输入特征 $x^{(1)}$ 组加入集合 S , 作为第 1 个输入特

征组。

$$\max_{x^{(1)} \in R} I(x^{(1)}; y) \quad (13)$$

3) 利用 S 中先前记录的特征 $x^{(i)}$, 根据式(11)选择下一组输入特征 $x^{(j)}$ 。

$$\max_{x^{(j)} \in R-S} \left[\frac{I(x^{(j)}; y)}{\frac{1}{|S|} \sum_{x^{(i)} \in S} I(x^{(i)}; x^{(j)})} \right] \quad (14)$$

4) 将步骤 3) 中选择的特征组 $x^{(j)}$ 加入到集合 S 中, 然后重复步骤 3)。

5) 直到所有输入特征都排序完毕。最终的排序结果表明,如果选择 M_1 ($M_1 \leq M$) 个特征组的子集作为分类器的输入,那么集合 S 中第 1 个 M_1 特征组就是最优子集,与目标的相关性更强,内部冗余更少。

4 实验

4.1 实验准备

本文采用的振动原始数据集来自美国凯斯西储大学的开放数据。在电机驱动端 SKF 6205 上进行诊断,该实验数据分别提取了选取滚动轴承的 10 种状态,为正常、内圈点蚀直径 0.177 8 mm、内圈点蚀直径 0.355 6 mm、内圈点蚀直径 0.533 4 mm、滚珠点蚀直径 0.177 8 mm、滚珠点蚀直径 0.355 6 mm、滚珠点蚀直径 0.533 4 mm、外圈点蚀直径 0.177 8 mm(中心方向)、外圈点蚀直径 0.355 6 mm(中心方向)、外圈点蚀直径 0.533 4 mm(中心方向)。采样频率为 12 kHz,转速为 1 797 r/min。将采集到的原始振动信号进行分割,每个样本包含 1 024 个采样点。每种状态下的样本数量为 115 组,10 类故障总共 1 150 组数据集。首先从 10 类故障中每类随机抽取 70 组总共 700 组作为训练集,剩下的 450 组数据作为测试集,两个互斥数据集作为本次实验的数据来源。分类情况如表 1 所示,其中 RF、IF、OF 分别为滚珠、内圈、外圈故障(中心方向)。

表 1 轴承数据集描述

Table 1 Bearing dataset description

滚动轴承状态	训练/测试	类别标签
正常	70/45	1
IF(0.177 8 mm)	70/45	2
IF(0.355 6 mm)	70/45	3
IF(0.533 4 mm)	70/45	4
RF(0.177 8 mm)	70/45	5
RF(0.355 6 mm)	70/45	6
RF(0.533 4 mm)	70/45	7
OF(0.177 8 mm)	70/45	8
OF(0.355 6 mm)	70/45	9
OF(0.533 4 mm)	70/45	10

为了解决原始信号中存在的大量噪声致使故障特征提取不易的问题,将原始信号进行 CEEMDAN 分解, ϵ_1 取 0.2, 设置加噪次数和最大迭代次数分别为 500 和 5 000, 最后分解得到一系列 IMF。分解效果如图 4~7 所示。

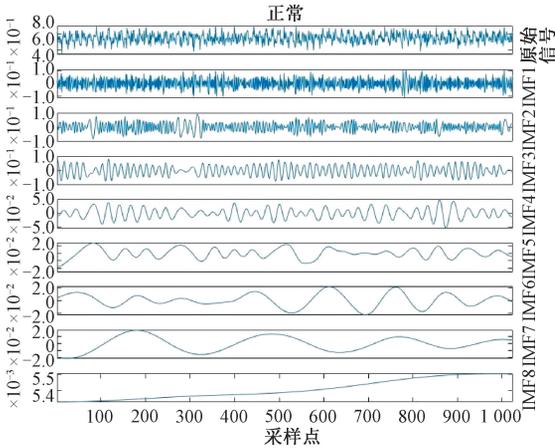


图 4 正常信号 CEEMDAN 分解

Fig. 4 CEEMDAN decomposition of normal signal

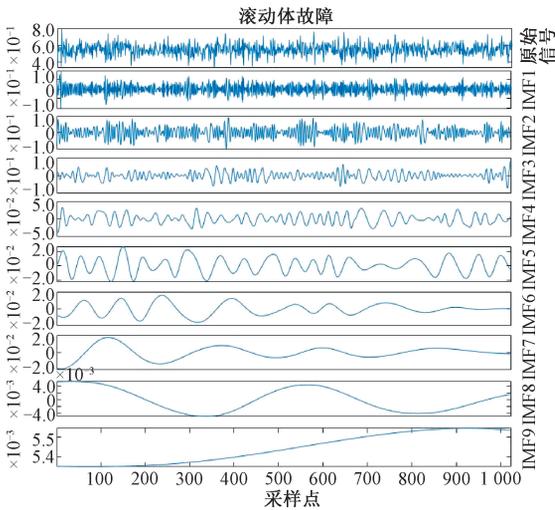


图 5 滚动体故障 CEEMDAN 分解

Fig. 5 CEEMDAN decomposition of rolling element fault

由图 4~7 可知,第 1 个分量信号振幅较其他分量明显较小、振动频率较快,因此可以断定该分量为随机噪声,不作为本次实验的分量选择,第 8、9 分量含有的特征信息较少,与原始信号相关性较小,缺乏物理分析意义,属于虚假分量,将之剔除。第 2~7 个分量波动较慢,时域波形图大部分由高次谐波组成,频谱能量较为集中,且大部分时域波形接近正弦波,是理想的信号分析时间序列,因此将此部分分量进行重构。其次分别提取重构信号的 17 个时域特征,组成本文实验的特征分析原始数

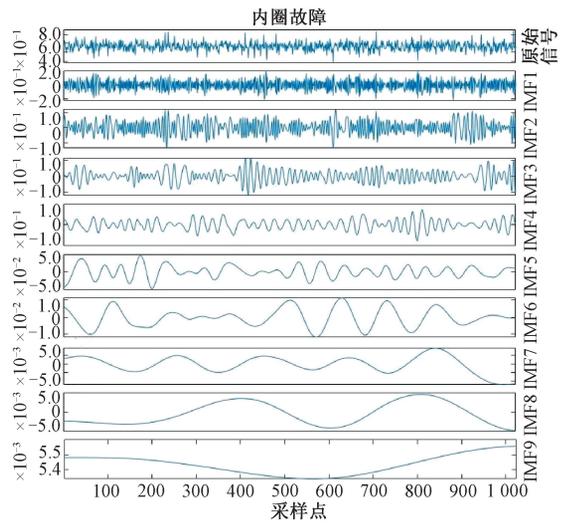


图 6 内圈故障 CEEMDAN 分解

Fig. 6 CEEMDAN decomposition of inner ring fault

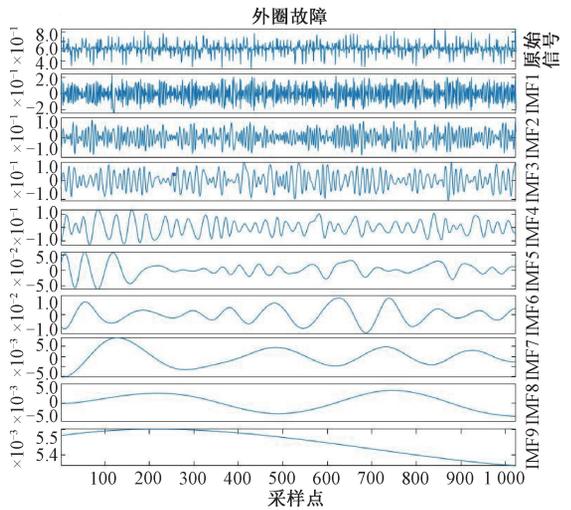


图 7 外圈故障 CEEMAN 分解

Fig. 7 CEEMDAN decomposition of outer ring fault

据集。

4.2 特征分析

“优秀”的特征不仅能达到表征原始数据的目的,还能降低学习任务的难度^[21]。本次实验将对提取的 17 个时域特征在不同维度进行冗余性和相关性分析,获取各个维度的特征知识,以分析其对本次分类实验结果的影响。

首先利用单个特征进行分类,分类器采用 RF,统计每个特征分类结果如表 2 所示。

分析表 2 可知,均值、绝对平均值、方根幅值、方差、标准差、有效值、波形因子的实验分类效果较好,记为有效特征,共计 7 个,可得出该组重要性程度较高,对分类

表 2 单个特征分类准确率图

Table 2 Single feature classification accuracy chart

序号	特征	准确率/%	序号	特征	准确率/%
1	均值	85.33	10	峰峰值	26.22
2	绝对平均值	84.89	11	裕度因子	25.78
3	有效值	82.67	12	最小值	25.56
4	方根幅值	81.11	13	余隙因子	25.33
5	波形因子	70.44	14	脉冲因子	24.89
6	方差	69.56	15	峰值	24.44
7	标准差	68.67	16	最大值	23.56
8	峭度	43.33	17	偏度	14.00
9	峰值因子	32.11			

有重要影响,具有较高的分类价值。剩余特征共 10 个,记为无效特征,单个特征分类准确率较低,对分类结果影响较小。

为了比较 mRMR 优异的特征筛选能力,分别采用 MIC、Filter 过滤、皮尔逊相关系数以及原始时域特征矩阵随机抽取进行特征筛选,分类器选择随机森林算法进行分类。其中皮尔逊相关系数和 MIC 以相关系数最小原则进行升序排列,Filter 过滤以单个特征方差最大原则进行降序排列。实验部分按照特征顺序进行抽取,初始特征个数为 1,设置特征个数增长步长为 1,最大特征分类个数为 17,最终分类效果如图 8 所示。

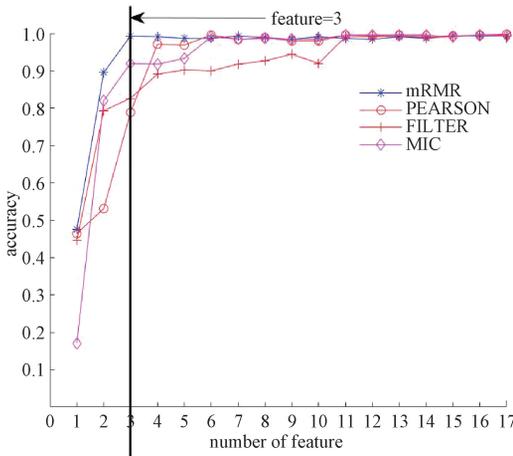


图 8 特征收敛图

Fig. 8 Characteristic convergence graph

从图 8 可以看出,基于 mRMR 特征筛选方法收敛效果要好于其余特征筛选方法,仅需 3 个特征便可以达到较好收敛性能,由于 mRMR 从冗余性和相关性方面拥有较好的选择策略,所选择特征与目标变量的相关关系度量效果较理想,因此获得了较好的预测精度和特征排序序列,分类准确率达到 99.03%,且收敛后保持了较高的分类效果,具有较强的鲁棒性。基于 Filter 的特征过滤方法准确率最低,该方法只考虑单个特征的变化,导致了较

低的分类精度,该方法选择的特征不能很好的拟合原始数据。

为了证明基于 mRMR 的有效特征保留效果,图 9~13 所示为特征最优排序以及各个特征选择算法的最终特征排列情况,排列顺序为算法结果从上到下依次排列。有效特征的重要程度较高,因此红色特征越靠上代表选取的特征具有最快的收敛性。同时为了更好说明 mRMR 特征筛选的效果,对各个特征筛选方法依据实验结果取前 8 个特征,统计该部分有效特征和无效特征的个数,统计结果如表 3 所示。

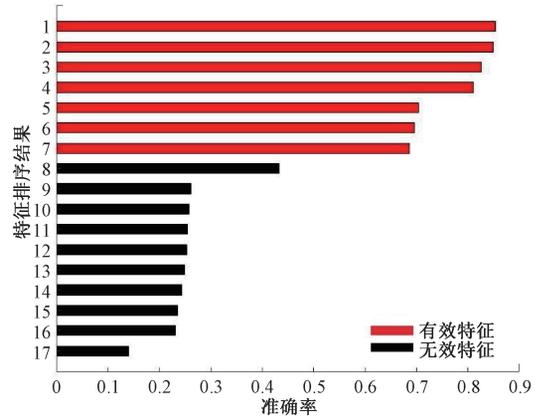


图 9 最优特征排序

Fig. 9 Optimal feature ranking graph

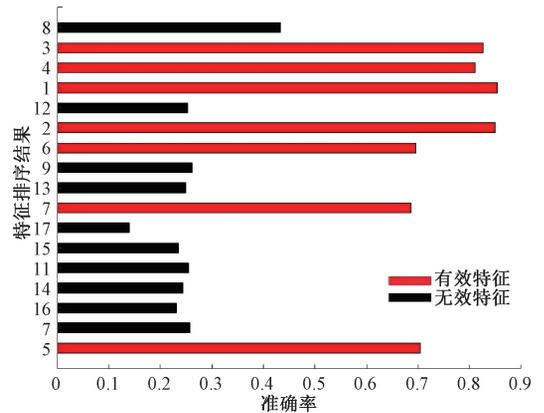


图 10 基于 mRMR 的特征筛选结果

Fig. 10 mRMR based feature selection result graph

表 3 前 8 个有效/无效特征保留结果

Table 3 Top 8 valid/invalid feature retention results

特征提取方法	mRMR	Pearson	MIC	Filter
有效特征	5	4	2	0
无效特征	3	4	6	8

由图 10 和表 3 可知,mRMR 在衡量特征重要程度方面具有较大优越性,充分考虑了冗余性和相关性,较好判

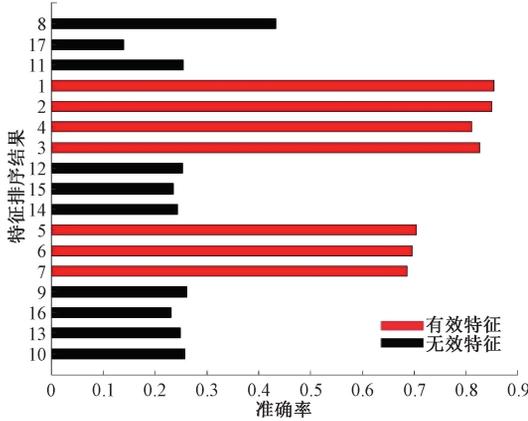


图 11 基于 Pearson 相关系数的特征筛选结果
Fig. 11 Feature selection result graph based on Pearson correlation coefficient

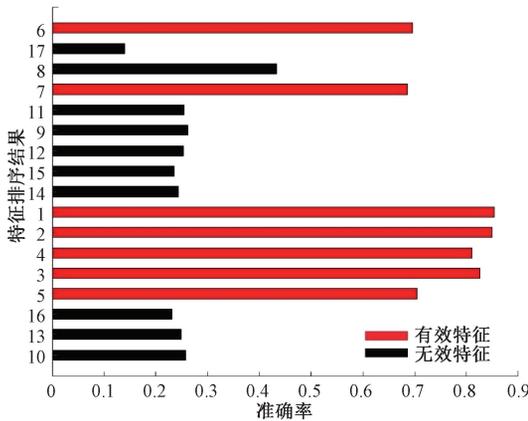


图 12 基于 MIC 的特征筛选结果
Fig. 12 Feature selection result graph based on MIC

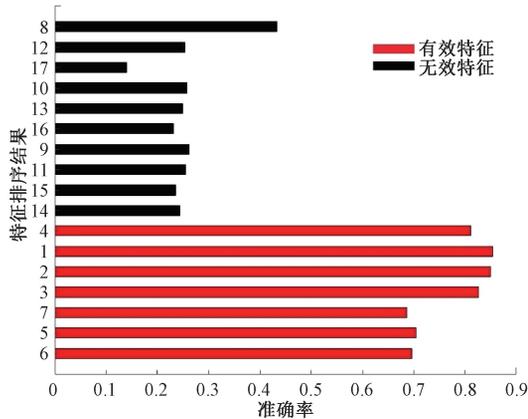


图 13 基于 Filter 过滤的特征筛选结果
Fig. 13 Feature selection result graph based on filter

的分类,并保持较高的分类精度。其次是皮尔逊相关系数,该特征提取方法对特征间相关性进行了分析,能较好的保留主要特征,但冗余性问题并未解决,因此特征筛选效果稍差。基于 Filter 过滤的特征选择方法效果较差,重要特征均未得的较好选择,该方法的特征选择规律拟合效果较差,与原始信号规律产生较大偏差。

4 结 论

本文针对原始信号驳杂,含有较多噪声的情况,对进行 CEEMDAN 的分解与重构,去除噪声等无关数据,再对时域等特征进行有条件筛选,有效提高了算法分类精度。针对时域特征向量存在“多余”特征,提出一种基于 mRMR 的特征提取方法,根据策略剔除冗余和相关性较大的特征,实验结果表明,该方法能较好的拟合原始信号特征规律,并且能够选取到优秀特征,使得算法能够快速收敛。

参考文献

[1] 杨菊花,刘洋,陈光武,等. 基于改进 EMD 的微机械陀螺随机误差建模方法 [J]. 仪器仪表学报, 2019, 40(12): 196-204.
YANG J H, LIU Y, CHEN G W, et al. Modeling method for random error of micromechanical gyroscope based on improved EMD [J]. Chinese Journal of Scientific Instrument, 2019, 40(12): 196-204.

[2] ZHAO D, LI K, LI H. A new method for separating EMI signal based on CEEMDAN and ICA [J]. Neural Processing Letters, 2021, 53(3): 2243-2259.

[3] MIN Y, YE M, TIAN L, et al. Unsupervised feature selection via multi-step markov probability relationship [J]. Neurocomputing, 2021, 453 : 241-253.

[4] 余博,田福庆,梁伟阁. 基于深度卷积变分自编码网络的故障诊断方法 [J]. 仪器仪表学报, 2018, 39(10): 27-35.
SHE B, TIAN F Q T, LIANG W G. Fault diagnosis based on a deep convolution variational autoencode network [J]. Chinese Journal of Scientific Instrument, 2018, 39(10): 27-35.

[5] BABAJANIAN B H, AMIRI G G, NEKOOEI M, et al. Damage detection of a cable-stayed bridge based on combining effective intrinsic mode functions of empirical mode decomposition using the feature selection technique [J]. Inverse Problems in Science and Engineering, 2021, 29(6): 861-881.

[6] 雷海锐,高秀峰,刘辉. 基于机器学习的混合式特征选择算法 [J]. 电子测量技术, 2018, 41(16): 42-46.

定各个特征的重要程度,重要特征均靠前得到选取,这使得模型能够进行快速收敛,选取的特征极大有利于最终

- LEI H R, GAO X F, LIU H. Mixed feature selection method based on machine learning [J]. *Electronic Measurement Technology*, 2018, 41 (16):42-46.
- [7] SALMANPOUR M R, SHAMSAEI M, RAHMIM A. Feature selection and machine learning methods for optimal identification and prediction of subtypes in Parkinson's disease [J]. *Computer Methods and Programs in Biomedicine*, 2021, 206: 106131.
- [8] 卞华棋, 王育辉, 王栋. 基于 PCA 的空战机动高阶重构与评估 [J]. *电子测量与仪器学报*, 2020, 34 (6): 191-197.
- BIAN H Q, WANG Y H, WANG D. High order reconstruction and evaluation of air combat maneuver based on PCA [J]. *Journal of Electronic Measurement and Instrumentation*, 2020, 34(6):191-197.
- [9] MALEKI N, ZEINALI Y, NIAKI S T A. A k-NN method for lung cancer prognosis with the use of a genetic algorithm for feature selection [J]. *Expert Systems with Applications*, 2021, 164: 113981.
- [10] 颜弋凡, 安路达, 吕志民. 基于最大互信息系数属性选择的冷轧产品机械性能预测 [J]. *中南大学学报(自然科学版)*, 2020, 51(1):68-75.
- YAN G F, AN L D, LYU ZH M. Prediction of mechanical properties of cold rolled products based on maximal information coefficient attribute selection [J]. *Journal of Central South University (Science and Technology)*, 2020, 51(1):68-75.
- [11] 张俐, 袁玉宇, 王枫. 基于最大相关信息系数的 FCBF 特征选择算法 [J]. *北京邮电大学学报*, 2018, 41(4): 86-90.
- ZHANG L, YUAN Y Y, WANG Z. FCBF feature selection algorithm based on maximum information coefficient [J]. *Journal of Beijing University of Posts and Telecommunications*, 2018, 41(4):86-90.
- [12] 孙曙光, 纪学玲, 杜太行. 机械振动下交流接触器电寿命预测失效特征量提取 [J]. *仪器仪表学报*, 2019, 40(3):114-125.
- SUN SH G, JI X L, DU T H. Failure feature extraction for electrical life prediction of AC contactor under mechanical vibration [J]. *Chinese Journal of Scientific Instrument*, 2019, 40(3):114-125.
- [13] CHEN P T, LI F, WU C W. Research on intrusion detection method based on Pearson correlation coefficient feature selection algorithm [C]. *Journal of Physics: Conference Series*, 2021:12-54.
- [14] 杨彦茹, 温杰, 史元浩, 等. 基于 CEEMDAN 和 SVR 的锂离子电池剩余使用寿命预测 [J]. *电子测量与仪器学报*, 2020, 32(12):197-205.
- YANG Y R, WEN J, SHI Y H, et al. Remaining useful life prediction for lithium-ion battery based on CEEMDAN and SVR [J]. *Journal of Electronic Measurement and Instrumentation*, 2020, 32(12):197-205.
- [15] 原渊. 基于 ORB 特征描述子与随机森林的条码缺陷识别算法 [J]. *国外电子测量技术*, 2019, 38 (7): 61-64.
- YUAN Y. Barcode defect recognition algorithm based on ORB feature descriptor and random forest [J]. *Foreign Electronic Measurement Technology*, 2019, 38 (7): 61-64.
- [16] 丁家满, 吴晔辉, 罗青波, 等. 基于深度森林的轴承故障诊断方法 [J]. *振动与冲击*, 2021, 40(12):107-113.
- DING J M, WU Y H, LUO Q B, et al. Fault diagnosis method of bearing based on deep forest [J]. *Vibration and Shock*, 2021, 40 (12): 107-113.
- [17] YUAN C, CLARENCE W S, BING L, et al. Application of feature extraction through convolution neural networks and SVM classifier for robust grading of apples [J]. *Instrumentation*, 2019, 6(4):59-71.
- [18] PIRI J, MOHAPATRA P. An analytical study of modified multi-objective Harris Hawk optimizer towards medical data feature selection [J]. *Computers in Biology and Medicine*, 2021, 135: 104558.
- [19] 易可夫, 胡锦, 崔凯阳, 等. 基于二次特征筛选的覆铜板表面缺陷检测 [J]. *现代制造工程*, 2021 (6): 95-100.
- YI K F, HU J, CUI K Y, et al. Surface defect detection of copper clad laminate based on secondary feature selection [J]. *Modern Manufacturing Engineering*, 2021 (6): 95-100.
- [20] 韩鹏, 郭天, 汪晋宽, 等. 基于 mRMR-ESN 的单变量光伏功率预测 [J]. *东北大学学报(自然科学版)*, 2021, 42(2):174-179.
- HAN P, GUO T, WANG J K, et al. Univariate photovoltaic power prediction based on mRMR-ESN [J]. *Journal of Northeast University (Natural Science Edition)*, 2021, 42 (2): 174-179.
- [21] 陈瑞娟, 邓光华, 刁小飞, 等. 基于 MIC 心率变异性特征选择的情感识别研究 [J]. *电子测量与仪器学报*, 2020, 34(12):57-65.
- CHEN R J, DENG G H, DIAO X F, et al. Research on emotion recognition based on MIC heart rate variability feature selection [J]. *Journal of Electronic Measurement*

and Instrumentation, 2020,34 (12): 57-65.

作者简介



常梦容, 2018 年于曲靖师范学院获得学士学位, 现为昆明理工大学硕士研究生, 主要研究方向为人工智能以及嵌入式研究。
E-mail: 1414959247@qq.com.

Chang Mengrong received her B. Sc. degree from Qujing Normal University in 2018.

She is currently a M. Sc. candidate at Kunming University of Science and Technology. Her main research interests include artificial intelligence and embedded research.



王海瑞, 教授, 博士, 现为教育部高等学校计算机科学与技术专业教学指导分委员会专家成员, 主要从事计算机应用基础和工程技术的研究。

E-mail: hrwang88@163.com

Wang Hairong, Ph. D. degree, professor. Now he is an expert member of the Teaching Guidance Subcommittee of Computer Science and Technology in Higher Education Ministry of Education. His main research interests include computer application foundation and engineering technology.