

引用格式:王子兰,杨瑞.基于随机森林算法的旋转机械齿轮组故障诊断[J].山东科技大学学报(自然科学版),2019,38(5):104-112.

WANG Zilan, YANG Rui. Fault diagnosis of rotating machinery gearbox based on random forest algorithm[J]. Journal of Shandong University of Science and Technology (Natural Science), 2019, 38(5): 104-112.

基于随机森林算法的旋转机械齿轮组故障诊断

王子兰¹, 杨 瑞²

(1. 山东科技大学 数学与系统科学学院, 山东 青岛 266590; 2. 山东科技大学 电气与自动化工程学院, 山东 青岛 266590)

摘 要:针对单一的分类器用于旋转机械故障诊断时存在准确率不高的问题,提出一种基于随机森林算法的旋转机械齿轮组故障诊断方法。该方法利用随机森林多分类器组合决策树的思想,通过多分类器的组合学习提高故障诊断的准确率,并在风力涡轮动力传动系统故障诊断模拟器系统上进行了多工况多故障的实验验证。首先,收集多工况、多故障的齿轮传感器信号,提取传感器信号的时域特征作为随机森林的输入特征量。然后,利用构建好的随机森林模型进行齿轮组的故障诊断,并将随机森林算法的分类结果与支持向量机方法的分类结果进行对比。通过对故障诊断结果的分析,随机森林算法避免了复杂的寻参过程和传统分类器的过拟合现象,能够处理大规模数据集,通过分类器的组合,提高了故障诊断准确率,并缩短了分类模型的预测时间,具有较好的应用前景。

关键词:随机森林;旋转机械;故障诊断;决策树;特征提取

中图分类号:TH132

文献标志码:A

文章编号:1672-3767(2019)05-0104-09

DOI:10.16452/j.cnki.sdkjzk.2019.05.013

Fault diagnosis of rotating machinery gearbox based on random forest algorithm

WANG Zilan¹, YANG Rui²

(1. College of Mathematics and Systems Science, Shandong University of

Science and Technology, Qingdao, Shandong 266590, China;

2. College of Electrical Engineering and Automation, Shandong University of Science and

Technology, Qingdao Shandong 266590, China)

Abstract: As the accuracy rate of single classifier based fault diagnosis is low for rotating machinery, a fault diagnosis method based on the random forest algorithm is proposed in this paper. Hardware experiments were carried out on wind turbine drivetrain diagnosis simulator (WTDDS) to show the effectiveness of the proposed method. Initially, the sensor signals under multiple operation modes with different gear fault types were collected with time domain characteristics and extracted as input features of random forest. Afterwards, the fault diagnosis of the gearbox was carried out using the constructed random forest model, and the classification result using random forest algorithm was compared with that using support vector machine. Through the analysis of the fault diagnosis results, the ran-

收稿日期:2018-11-01

基金项目:国家自然科学基金项目(61603223);山东省自然科学基金项目(ZR2016FB01);青岛市应用基础研究计划项目(16-5-1-7-jch, 17-1-1-1-jch);中央高校基本科研业务费专项资金项目(18CX02093A);青岛市博士后研究人员应用研究项目;山东科技大学人才引进科研启动基金项目;西交利物浦大学重点项目建设专项资金(KSF-E-34)

作者简介:王子兰(1994—),女,山西晋中人,硕士研究生,主要从事故障诊断与容错控制的研究。

杨 瑞(1985—),男,山东济南人,副教授,主要从事数据驱动的故障诊断研究,本文通信作者。

E-mail: r.yang@xjtlu.edu.cn

dom forest algorithm can process large-scale data sets and improve the accuracy of fault diagnosis by avoiding the over-fitting phenomenon of conventional classifier. It can also shorten the prediction time through the combination of weak classifiers with higher prospects as a potential application.

Key words: random forest; rotating machinery; fault diagnosis; decision tree; feature extraction

随着工业复杂性的增加,大型设备日趋复杂化,旋转机械作为现代工业的主要动力源之一,通常运行在较为恶劣的环境中,一旦发生故障轻则影响生产,造成一定的经济损失,重则带来严重的安全事故。为了保障设备的高效安全运行,有必要开展有关旋转机械的故障诊断和监测研究。由于其子系统的复杂性、工作状态的多变性和操作不当等因素的影响,导致在旋转机械的故障诊断中获取到的通常是被污染的信号。因此,提高旋转机械故障诊断的准确率一直是备受关注的问题。齿轮组作为旋转机械中不可缺少的传动和连接部件,在传动过程中损坏比较频发,其能否正常运行与整个旋转机械的工作性能密切相关。因此在旋转机械中,齿轮组的故障诊断一直备受关注。常见的齿轮组故有两种:一种是在传动过程中轮齿发生的故障,主要有点蚀、断齿、磨损等;另一种是齿轮偏心、轴承不平衡等造成的故障。

传统的旋转机械齿轮组故障诊断方法有基于解析模型的方法^[1]、基于专家系统的方法^[2]、基于人工智能的方法^[3-4]等。其中,基于解析模型的方法需要对复杂的机械结构进行数学建模及大量的推导计算,虽然准确率较高,但对复杂机械的原理等专业知识的要求较高,因此在旋转机械的故障诊断中普适性较弱。基于专家系统的方法不需要知道系统的精确模型,故障诊断过程便于理解,但实际诊断中需要大量的专家知识及经验,知识的数量和质量对诊断结果的准确性有一定的影响。在基于人工智能的故障诊断方法中,主要有基于神经网络的方法^[3]、基于支持向量机的方法^[4]等。其中,基于神经网络的方法在旋转机械故障诊断中应用较广,其结构的选择、收敛速度较慢等都制约着诊断结果的准确性。支持向量机采用结构风险最小化的原理,在训练数据集较小的情况下诊断效果较好。但实际工业中,通常存在大量训练样本,对于大量输入数据支持向量机的诊断性能相对较差。

机器学习算法在故障诊断领域越来越受到人们的重视,随机森林(random forest, RF)作为其中的一种也在不断发展。RF 最早由 Leo Breiman 和 Adele Culter 提出,该算法是一个包含多个弱分类器的组合学习算法^[5]。大量研究^[6-8]表明,RF 相比其他算法有着明显的优势。能够处理大规模数据,并且在数据缺失的情况下也具有很好的预测精度,实现过程比较简单,既能处理离散数据也能处理连续数据,因此近 10 年来,RF 算法在多个领域都有较好的发展。目前国内将 RF 算法用于故障诊断的研究相对较少,本研究将 RF 算法应用于旋转机械齿轮组的故障诊断中,并将实验结果与旋转机械故障诊断中应用较广的支持向量机方法进行比较。结果显示,与支持向量机相比,RF 算法模型参数较少,实现过程较为简单,并且能够处理大规模数据,明显提高了故障诊断的准确率。因此,基于 RF 算法的故障诊断有着重要的研究价值。

将随机森林算法应用于旋转机械齿轮组的故障诊断中,针对多工况、多故障同时存在的情况,用一种新的人工智能方法进行故障诊断,取得了比传统旋转机械齿轮组故障诊断更高的分类准确率。该方法不仅拓展了随机森林算法的应用领域,同时也给旋转机械齿轮组的故障诊断提供了新思路。

1 基于 RF 算法的旋转机械齿轮组故障诊断

1.1 RF 算法介绍

RF 是一种有监督的集成学习算法,广泛应用于分类问题和回归问题。该算法综合了 1996 年 Leo^[9]提出的 Bagging 算法和 1998 年 Ho^[10]提出的随机子空间方法,利用两种随机的思想,很好地抑制了传统分类器的过拟合现象。其基本思想是将分类能力较弱的多个分类器进行组合,形成分类能力较强的分类器群。该算法选取分类回归树(classification and regression tree, CART)^[11]作为基分类器进行相关问题的分类。

RF 算法是一种建立在特征集合上的算法,特征的优劣在一定程度上制约着算法的分类准确率。因此,特征的选取在 RF 中具有重要的意义。假定 RF 由 k 个分类器 $h_1(T), h_2(T), \dots, h_k(T)$ 集合而成,表示为:

$$H = \{h_1(T), h_2(T), \dots, h_k(T)\} \quad (1)$$

其中： T 是输入特征集， $h_q(T), q=1,2,\cdots,k$ 是基分类器，每个分类器均为利用 CART 算法生成的决策树，综合 k 棵决策树的分类结果按少数服从多数原则进行投票，投票结果即为 RF 的分类结果。

RF 算法的实现包括以下几个部分：

1) 利用 bootstrap 重抽样方法^[12]从含有 n 个样本的原始数据集中进行 k 次随机且有放回的抽样，形成 k 个 bootstrap 子样本集，每个子样本集中均含有 n 个样本。

2) 针对每个 bootstrap 子样本集，利用 CART 算法进行决策树的构建。该算法使用“基尼指数”(Gini index)作为划分最优特征及最优分裂值的指标。假设当前所选取的特征 t_i 中包含 C 个类别，其 $Gini$ 值如下所示：

$$Gini(t_i) = \sum_{j=1}^C \sum_{j' \neq j} p_j p_{j'} = 1 - \sum_{j=1}^C p_j^2 \tag{2}$$

式中， p_j 为第 j 个类别出现的概率。从式(2)可以看出， $Gini$ 值越小分类的纯度越高。因此，遍历该决策树的所有特征，选取 $Gini$ 值最小的特征作为当前节点的分裂特征。

3) 用 k 个 bootstrap 子样本集生长 k 棵决策树形成随机森林。每棵决策树在生长时，随机地从含有 M 个特征的集合 T 中抽取 $m=\lfloor \sqrt{M} \rfloor$ ^[13] 个特征作为一个特征子集，利用特征子集作为决策树生长过程中的划分属性，按照 $Gini$ 值最小的原则进行决策树的生长。

4) 综合 k 棵决策树的分类结果^[14]，利用少数服从多数的原则进行投票，投票结果即为 RF 的分类结果。

1.2 基于 RF 算法的旋转机械齿轮组故障诊断方法

针对主要的两种基于人工智能的故障诊断方法存在的不足，本研究拟采用基于 RF 算法进行齿轮组的故障诊断，其故障诊断流程图如图 1 所示。

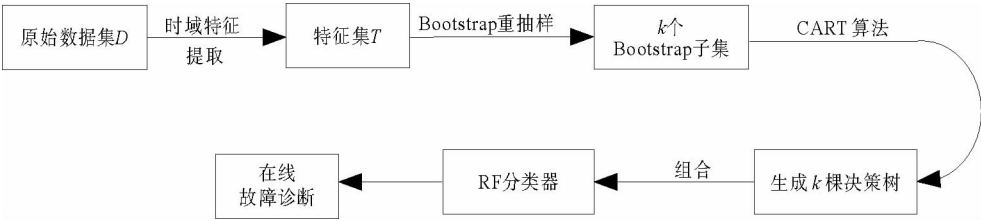


图 1 基于 RF 算法的故障诊断流程图

Fig. 1 Flowchart of RF based fault diagnosis

基于 RF 算法的旋转机械齿轮组故障诊断具体步骤为：

1) 原始数据集 D 的获取：选取齿轮组裂纹、缺齿、断齿三类故障进行数据采集，针对每类故障，分别从旋转机械的振动传感器、扭力传感器测量中选取 l 个数据，并记录每个数据对应的故障标签 $y_i(i=1,2,\cdots,3l)$ ：

$$y_i \in \{y_1,y_2,y_3\}, i=1,2,\cdots,3l, \tag{3}$$

$$D=\{D_1 \cup D_2\}, \tag{4}$$

$$D_1=\{(x_i,y_i)\}, i=1,2,\cdots,3l, \tag{5}$$

$$D_2=\{(x_i^{prime},y_i)\}, i=1,2,\cdots,3l. \tag{6}$$

其中： y_1, y_2, y_3 分别表示裂纹、缺齿、断齿三类故障， D_1 为振动传感器测量输出、 D_2 为扭力传感器测量输出。

2) 特征提取：结合旋转机械故障检测系统的特点和文献[15]的时域特征参数表，选取表 1 中的时域特征作为决策树生长过程中的分裂属性。

表 1 时域特征描述
Tab. 1 Description of time domain characteristics

时域指标	表达式	符号表示	时域指标	表达式	符号表示
最大值	$t_1 = \max(x_i)$	x_1	峰值指标	$t_5 = \frac{t_1}{t_2}$	x_5
均方根	$t_2 = \sqrt{\frac{\sum_{i=1}^n (x_i)^2}{n}}$	x_2	裕度指标	$t_6 = \frac{t_1}{t_3}$	x_6
方根幅值	$t_3 = \left(\frac{\sum_{i=1}^n \sqrt{ x_i }}{n} \right)^2$	x_3	绝对平均值	$t_7 = \sum_{i=1}^n x_i / n$	x_7
标准差	$t_4 = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$	x_4			

注： x_i 是传感器测量， $i = 1, 2, \cdots, n$ 是样本数

3) bootstrap 重抽样及训练数据集和测试数据集的划分:根据 1.1 节 bootstrap 重抽样方法对原始数据集进行 k 次抽样,每个数据未被抽中的概率为 $\left(1 - \frac{1}{k}\right)^k$,则当 k 趋于无穷大时,其值将收敛于 $\frac{1}{e}$ 。因此, k 个样本集共包含了原始数据集中 $\frac{2}{3}$ 的数据,该部分用于 RF 中作训练数据集。未被抽中的数据称为袋外数据(out of bag, OOB),约占原始样本集的 $\frac{1}{3}$,用于 RF 中作测试数据集。并且利用 OOB 误差可以很好地对 RF 的分类准确率^[16]进行评估。

4) CART 决策树的生长和分类结果:根据 1.1 节 CART 决策树的构建方法选取最优的分裂特征和最优分裂值 a 进行节点的分裂。假定选取的最优分裂特征为 t_i ,最优分裂值 a 的选取如下:

$$\min_a Gini(t_i, a) = \frac{N_1}{N} Gini(t_{i1}) + \frac{N_2}{N} Gini(t_{i2}). \tag{7}$$

其中: t_{i1} 、 t_{i2} 为枚举得到的最优分裂特征 t_i 的两个子样本集, N_1 、 N_2 为两个子样本集的样本数, N 为最优分裂特征为 t_i 时的样本数。按照公式(7)的规则进行 CART 决策树的生长,直到 m 个特征全部用作分裂节点,则完成了该决策树的生长。

5) RF 的分类结果:将构建好的 k 棵决策树作为 RF 的基分类器群,对特征数据进行故障类别的划分,综合 k 棵决策树的分类结果,利用少数服从多数的原则进行投票得到 RF 的分类结果,从而实现故障诊断。

上述基于 RF 算法的旋转机械齿轮组故障诊断过程描述如表 2、表 3 所示。

2 实验验证

2.1 实验平台

实验平台选取美国 SpectraQuest 公司的风力涡轮动力传动系统故障诊断模拟器,其结构图如图 2 所示,其中 1 是单相电机、2 是二级平行轴变速箱、3 是行星齿轮箱变速器、4 是可编程磁力制动器、5 是传感器(从左到右分别为扭力传感器、振动传感器)。该平台可以在多工况、多故障同时存在的情况下对齿轮组的运行状态进行检测。

提出的故障诊断方法是针对旋转机械齿轮组的裂纹、缺齿、断齿三类故障进行的诊断。图 3 列出了齿轮组在正常情况下及三种故障情况下的实物图,其中:1 为正常、2 为缺齿、3 为裂纹、4 为断齿。

表 2 基于 RF 算法的旋转机械齿轮组故障诊断方法(离线过程)

Tab. 2 RF based fault diagnosis method on rotating machinery gearbox (offline process)

诊断过程	操作方法
预采集过程	1. 数据的采集:振动传感器测量及故障类型 D_1 , 扭力传感器测量及故障类型 D_2 ; 2. 特征提取:根据上文中步骤(2)进行故障特征的提取; 3. 利用 bootstrap 重抽样方法重复抽样 k 次,将每次抽样生成的 bootstrap 子集进行训练集和测试集的划分。
训练过程	1. 输入训练数据集; 2. 根据 1.2 节公式(7),进行 k 棵 CART 决策树的生长,并输出每棵决策树的分类结果; 3. 综合 k 棵决策树的分类结果,利用少数服从多数原则投票得到 RF 的分类结果; 4. 输出训练数据集的故障类型。
测试过程	1. 输入测试数据集; 2. 利用测试数据集进行每棵决策树的生长,并对每棵决策树的分类结果进行投票得到分类结果; 3. 输出:测试数据集的故障类型。

裂纹主要发生在齿根表面。由于齿面所受作用力超出材料所能承受的最大作用力,导致齿轮发生裂纹,在啮合时裂纹闭合促使缝隙中的油压增高,加速了裂纹的扩散,最终齿轮表面的金属一小块一小块的剥落形成点蚀。断齿是齿轮最常见的故障类型,由于齿轮的作用力主要集中在齿根处,因此断齿常发生于齿根部位。通常有三种情况:一是由于齿轮长时间处于工作状态,在各种力的作用下,产生疲劳裂纹并逐渐扩散,最终导致齿轮发生疲劳断齿。二是当齿轮处于超负荷的工作环境时,齿轮受到严重的冲击,导致齿根处的作用力超出所能承受的最大范围而发生断裂。三是由于齿轮本身质量较差或工作过程中出现偏载,使得齿根处受力不均匀而发生局部断齿(缺齿),局部断齿通常发生在轮齿端部。此外,受齿轮箱中未被清理的物质混入齿轮间隙、润滑效果不好等因素的影响,常常导致齿轮被磨损。齿轮磨损后,齿的厚度变薄,加大了整个设备的振动和噪声,严重时也可能引发断齿。

2.2 实验数据采集

实际工作中,转速的改变,噪声、环境的改变,齿轮、轴承等元件的损坏都可能引起工况的改变。但实际的转

表 3 基于 RF 算法的旋转机械齿轮组故障诊断方法(在线过程)

Tab. 3 RF based fault diagnosis method on rotating machinery gearbox (online process)

诊断过程	操作方法
输入	在线获得的数据集
分类过程	1. 根据上文中步骤(2)进行时域特征的提取; 2. 将提取特征之后的数据输入训练好的 RF 模型; 3. 输出每棵决策树的分类结果,投票得到 RF 的分类结果。
输出	在线数据集的故障类型

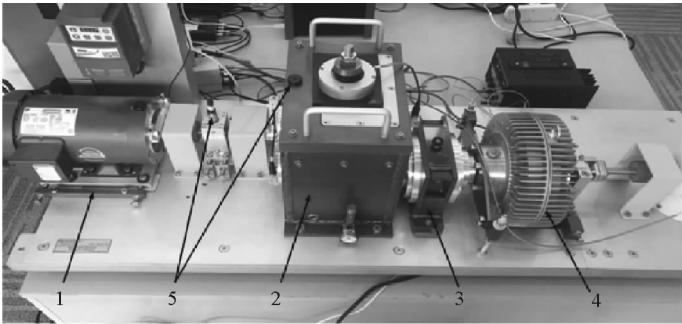


图 2 风力涡轮动力传动系统故障诊断模拟器
Fig. 2 Wind turbine drivetrain diagnosis simulator

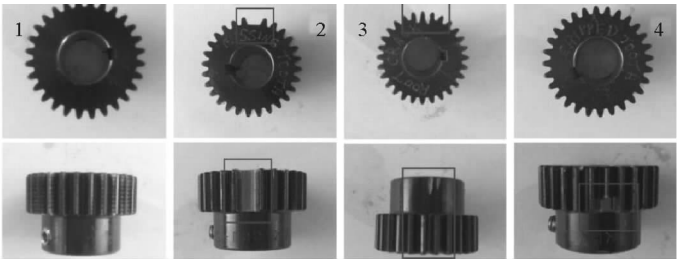


图 3 齿轮故障类型实物图
Fig. 3 The different classes of gear fault

速是不可预测的,在实验过程中转速的变化可以通过调节电机转动频率和负载电压来实现,假设转子转速是匀速的,而数据是在不同转速下采集的,在一定程度上也保证了工况的不同,具有一定的普适性。齿轮部分可选择分布形式为直齿或斜齿等不同类型的轮齿,还可以选择齿轮裂纹、断齿、缺齿等多种故障类型的齿轮。

现将工况的处理表示如下:借助风力涡轮动力传动系统故障诊断方法模拟器,选取四种工况(转动频率为 6、10 Hz,负载电压为 8、5 V)同时采集齿轮裂纹、缺齿、断齿三种故障的振动传感器和扭力传感器输出。其中,每个传感器采集 2 000 个数据,按照传感器的不同,将得到的故障数据按列进行整合,生成一个 $(2\,000 \times 3) \times 2$ 的矩阵。按照同样的整合方法将剩余三种工况的数据进行整合,共得到四个 $(2\,000 \times 3) \times 2$ 的矩阵。

根据表 1 所示的时域特征分别对四种工况下传感器测得的数据进行故障特征的提取。为了降低提取特征之后数据的维度,实验中将每 4 个数数据划分为一组进行特征提取。提取过程如下:分别对 500×3 组振动传感器的测量输出 D_1 和 500×3 组扭力传感器的测量输出 D_2 进行多维度特征向量的提取,记为集合 T 。 T 包含如下 14 个变量:最大值 $x_1(x_8)$ 、均方根 $x_2(x_9)$ 、方根幅值 $x_3(x_{10})$ 、标准 $x_4(x_{11})$ 、峰值指标 $x_5(x_{12})$ 、裕度指标 $x_6(x_{13})$ 、绝对平均值 $x_7(x_{14})$ 。利用 bootstrap 重抽样方法分别对每种工况下提取特征后的 500×3 行数据集进行重复抽样生成 k 个 bootstrap 子集,并对每种工况进行训练集和测试集的划分。

2.3 实验结果分析

图 4 以转动频率 6 Hz、负载电压 8 V 为例,给出了齿轮裂纹故障下的部分扭力传感器测量提取特征之后的数据。

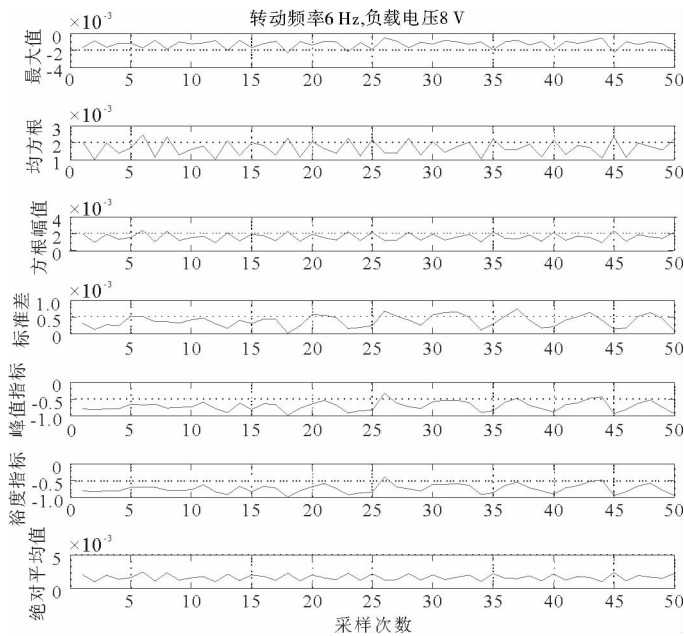


图 4 齿轮裂纹情况下的特征数据

Fig. 4 Extracted feature data under root crack fault

2.3.1 四种工况下的实验结果

基于 RF 算法的旋转机械齿轮组故障诊断实现过程较为简单,整个实验过程涉及到两个参数的选择,分别为最优决策树的棵数 k 和最优特征子集 m 。其中, k 值根据实验过程中的袋外误差来确定,通过优化,选择 $k=320$ 作为 RF 模型中最优决策树的棵数。对于工况 1,按照 1.2 节步骤(2)介绍的特征提取方法进行时域特征的提取,将提取后的特征向量 $\mathbf{x}_1\sim\mathbf{x}_{14}$ 和类别标签 Y 作为 RF 的输入,实验结果如表 5 所示。

表 5 训练数据集和测试数据集的混淆矩阵
Tab. 5 Confusion matrix of training and testing data set

实际输出	预测输出			分类错误率	实际输出	预测输出			分类错误率
	y_1	y_2	y_3			y_1	y_2	y_3	
y_1	334	5	0	0.015	y_1	159	2	0	0.012
y_2	1	341	0	0.003	y_2	2	155	1	0.019
y_3	0	0	348	0	y_3	0	0	152	0

注:左侧为训练数据集的混淆矩阵,右侧为测试数据集的混淆矩阵

由表 5 的混淆矩阵可以看出,利用 RF 进行分类时, y_1 被错分为 y_2 的有 2 个,被错分为 y_3 的有 0 个,分类错误率为 1.2%。 y_2 被错分为 y_1 的有 2 个,被错分为 y_3 的有 1 个,分类错误率为 1.9%。 y_3 的分类结果全部正确。总体来说,利用 RF 进行分类,预测准确率达到了 98.84%。

为衡量每个特征在 RF 的所有决策树中所做贡献的平均值,采用基于 Gini 值的特征重要性计算方法,将特征 t_i 在节点 m 处的重要性定义为节点 m 分枝前后 Gini 值的变化量,并通过计算特征 t_i 在每棵决策树中的重要性得到该特征在随机森林中的重要性。由图 5 可以得出, x_8 、 x_{11} 、 x_{12} 、 x_{13} 四个特征在 RF 的构建过程中贡献率较大,为此 RF 中最重要的四个特征,保证了所构建的 RF 的分类准确率。

在其他三种工况下的故障诊断结果如表 6 所示,由混淆矩阵可以看出,每种工况下 RF 的分类准确率都比较高,并且 RF 对 y_3 的分类效果最好,对 y_1 和 y_2 的分类准确率相差不大。由于 RF 算法在构建过程中只涉及到两个参数(k 和 m) 的选择,算法的实现过程较为简单,并能根据 Gini 值给出变量的重要性排序,方便进行最优特征的选取。

2.3.2 方法对比

为了验证 RF 算法能有效地提高模型的预测准确率,在相同的工况下,将 2.2 节提取特征之后的数据用于本文提出的基于 RF 的故障诊断方法和传统的基于支持向量机(support vector machine, SVM)的故障诊断方法^[17]。

表中首先给出了四种不同的工况,每种工况所采集的数据均包含旋转机械齿轮组裂纹、缺齿、断齿三类故障,并且在每类故障信号中选取 2 000 个数据用作实验所需的原始数据集。然后对采集到的 $2\,000\times 3$ 个故障数据按照文中表 1 的时域特征参数进行时域特征的提取。最后将提取特征之后的数据用于随机森林算法

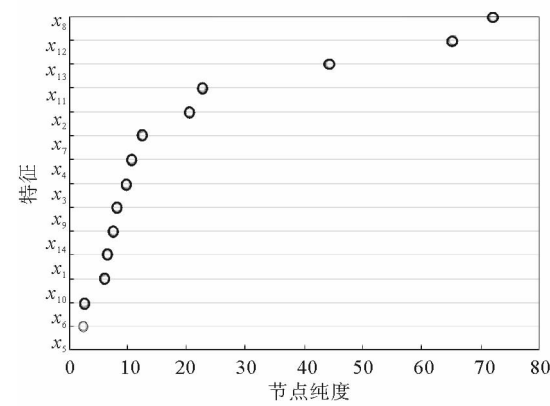


图 5 RF 特征重要性

Fig. 5 Contribution of the RF features

表 6 工况 2、3、4 测试数据集的混淆矩阵

Tab. 6 Confusion matrix of test data set for operation conditions 2, 3 and 4

工况	分类错误率		
	y_1	y_2	y_3
2	0.019	0	0
3	0.020	0.039	0
4	0.100	0.102	0

和文献[17]中介绍的支持向量机方法进行分类准确率的对比。实验对比结果如表 7 所示。

由上述对比实验可以看出,在相同的工况下,应用 RF 算法进行故障分类的分类准确率明显高于使用 SVM 方法的分类准确率。对于不同的工况而言,RF 算法的分类准确率较 SVM 算法明显提高。主要原因在于使用 SVM 算法进行分类得到的是单一的分类器,而 RF 算法采用 bootstrap 重抽样方法生成了多个分类器,且其训练集互不相同,并采用随机子空间的方法对特征集进行抽样,使得每个分类器的特征子集也互不相同,最后利用组合分类器的思想将每个分类器的分类结果进行组合,得到整个森林的分类结果。同时,在建立 RF 模型时只涉及到两个参数的确定,算法的时间复杂度较低。因此,与 SVM 方法相比,RF 算法的模型预测准确率有了明显的提高。

3 结论

针对传统的旋转机械故障诊断中使用单一分类器、诊断准确率不高的问题,提出了一种基于多个分类器组合而成的 RF 算法旋转机械齿轮组故障诊断。该方法通过基分类器组合学习,提高了森林预测准确率,大大缩短了模型的预测时间。利用风力涡轮动力传动系统故障诊断模拟器,进行了多工况、多故障的实验验证。结果显示,RF 算法能够很好地对齿轮组故障进行分类,避免复杂的寻参过程,能够处理大规模数据集,并且通过分类器组合,提高了模型的预测准确率,避免了传统分类器的过拟合现象。同时,应用随机森林算法进行旋转机械的故障诊断过程可看作一个循环过程。针对系统齿轮组出现的新故障,或者旋转机械其他部位出现的故障,首先需要明确知道该部分故障数据的类别标签,然后对这部分故障数据进行时域特征提取,最后将提取数据用作随机森林算法的输入进行随机森林的重新构建,并进行故障类别的划分。如再出现新的故障,可循环操作此过程完成故障类别的划分。

参考文献:

[1]周东华,胡艳艳.动态系统的故障诊断技术[J].自动化学报,2009,35(6):748-758.
ZHOU Donghua,HU Yanyan.Fault diagnosis techniques for dynamic systems[J].Acta Automatica Sinica,2009,35(6):748-758.

[2]YANG Z L,WANG B,DONG X H,et al.Expert system of fault diagnosis for gear box in wind turbine[J].Systems Engineering Procedia,2012,4:189-195.

[3]LI Y,CHENG G,LIU C,et al.Study on planetary gear fault diagnosis based on variational mode decomposition and deep neural networks[J].Measurement,2018,130:94-104.

[4]XING Z Q,QU J F,CHAI Y,et al.Gear fault diagnosis under variable conditions with intrinsic time-scale decomposition-singular value decomposition and support vector machine[J].Journal of Mechanical Science and Technology,2017,31(2):545-553.

[5]LEO B.Random forests[J].Machine Learning,2001,45(1):5-32.

[6]PARKHURST D F,BRENNER K P,DUFOUR A P,et al.Indicator bacteria at five swimming beaches-analysis using random forests[J].Water Research,2005,39(7):1354-1360.

[7]LEE S L A,KOUZANI A Z,HUB E J.Random forest based lung nodule classification aided by clustering[J].Computerized Medical Imaging and Graphics,2010,34(7):535-542.

[8]CHEN X W,LIU M.Prediction of protein-protein interactions using random decision forest framework[J].Bioinformatics,2005,21(24):4394-4400.

[9]LEO B.Bagging predictors[J].Machine Learning,1996,24(2):123-140.

[10]HO T K.The random subspace method for constructing decision forests[J].IEEE Transactions on Pattern Analysis and Machine Intelligence,1998,20(8):832-844.

- [11] 苏锦, 杨明, 王春香, 等. 一种基于分类回归树的无人车汇流决策方法[J]. 自动化学报, 2018, 44(1): 35-43.
SU Ti, YANG Ming, WANG Chunxiang, et al. Classification and regression tree based traffic merging for method sel-driving vehicle[J]. Acta Automatica Sinica, 2018, 44(1): 35-43.
- [12] STROBL C, MALLEY J, TUTZ G. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests[J]. Psychological Methods, 2009, 14(4): 323-348.
- [13] 方匡南, 吴见彬, 朱建平, 等. 随机森林方法研究综述[J]. 统计与信息论坛, 2011, 26(3): 32-38.
FANG Kuangnan, WU Jianbin, ZHU Jianping, et al. A review of technologies on random forests[J]. Statistics and Information Forum, 2011, 26(3): 32-38.
- [14] ZHUANG J F, LUO J, PENG Y Q, et al. Fault diagnosis method based on modified random forest[J]. Computer Integrated Manufacturing Systems, 2009, 15(4): 777-785.
- [15] LEI Y G, ZHENG J H, YAN Y Z. Fault diagnosis based on novel hybrid intelligent model[J]. Chinese Journal of Mechanical Engineering, 2008, 44(7): 112-117.
- [16] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.
- [17] 荆双喜, 华伟. 基于小波-支持向量机的齿轮故障诊断研究[J]. 山东科技大学学报(自然科学版), 2008, 27(4): 31-36.
JING Shuangxi, HUA Wei. Study on the gear fault diagnosis based on wavelet-support vector machine[J]. Journal of Shandong University of Science and Technology(Natural Science), 2008, 27(4): 31-36.

(责任编辑: 李 磊)

(上接第 71 页)

- [18] SHAO S, CHEN M, YAN X. Adaptive sliding mode synchronization for a class of fractional-order chaotic systems with disturbance [J]. Nonlinear Dynamics, 2015: 1-12.
- [19] TAVAZOEI M S, HAERI M. Synchronization of chaotic fractional-order systems via active sliding mode controller[J]. Physica A Statistical Mechanics & Its Applications, 2011, 387(1): 57-70.
- [20] LIN T C, LEE T Y. Chaos synchronization of uncertain fractional-order chaotic systems with time delay based on adaptive fuzzy sliding mode control[J]. IEEE International Conference on Fuzzy Systems, 2011, 19(4): 623-635.
- [21] GENESIO R, TESI A. Harmonic balance methods for analysis of chaotic dynamics in nonlinear systems[J]. Automatica, 1992, 28(3): 531-548.
- [22] LU J G. Chaotic dynamics and synchronization of fractional-order Genesio-Tesi systems[J]. Chinese Physics, 2005, 14(8): 1517-1521.
- [23] FAIEGHI M R, DELAVARI H. Chaos in fractional-order Genesio-Tesi system and its synchronization[J]. Communications in Nonlinear Science & Numerical Simulation, 2012, 17(2): 731-741.
- [24] SHAHRI E S A, BALOCHIAN S. Analysis of fractional-order linear systems with saturation using Lyapunov's second method and convex optimization [J]. International Journal of Automation & Computing, 2015, 12(4): 1-8.

(责任编辑: 傅 游)