

DOI: 10.19650/j.cnki.cjsi.J1905015

# 数据驱动的地铁车门微小故障智能诊断方法\*

施文<sup>1,2</sup>, 陆宁云<sup>1,2</sup>, 姜斌<sup>1</sup>, 支有冉<sup>2</sup>, 许志兴<sup>2</sup>

(1. 南京航空航天大学自动化学院 南京 211106; 2. 南京康尼机电股份有限公司 南京 211106)

**摘要:**车门控制系统是地铁车辆中最重要的子系统之一,其机电部件紧密耦合且存在频繁往复运动,易受环境和乘客干扰,故障率居高不下。为准确检测诊断地铁车门早期故障,本文提出一种大数据驱动的车门故障特征优选方法和基于随机森林(RF)的智能诊断方法。首先,从地铁运营公司累积的大量车门运行状态数据中,提取门扇位置、驱动电机转速和电流信号的多阶段时域特征指标,构建车门运行状态的特征向量;然后,应用距离评估准则,优选对故障敏感度高且对干扰鲁棒性强的车门状态特征,降低特征维度,减少冗余、无关特征的干扰;以优选后的车门状态特征作为RF网络的输入,故障标签作为输出,建立智能故障诊断模型,实现车门系统不同微小故障状态的自动识别。在杭州地铁4号线台架车门上的应用结果表明,所提方法能准确提取早期故障的微弱特征,故障分类模型精度高,故障诊断准确率优于现有其他方法。

**关键词:**微小故障诊断;距离评估准则;随机森林;地铁车门系统。

中图分类号: TP206.3 文献标识码: A 国家标准学科分类代码: 510.8060

## Data-driven intelligent incipient fault diagnosis for subway vehicle door system

Shi Wen<sup>1,2</sup>, Lu Ningyun<sup>1,2</sup>, Jiang Bin<sup>1</sup>, Zhi Youran<sup>2</sup>, Xu Zhixing<sup>2</sup>

(1. College of Automation Engineering, Nanjing University of Aeronautics & Astronautics, Nanjing 211106, China;

2. Nanjing Kangni Mechanical and Electrical Co., Ltd, Nanjing 211106, China)

**Abstract:** Door control system is one of the most important sub-systems in the subway vehicle. Due to the complex mechatronic structure, frequent open and close movement, and crowded passenger flow environment, high failure rate of door system persists. To accurately detect the incipient fault, a big-data-driven optimal feature selection algorithm and a random forests (RF) based incipient fault diagnosis method are proposed in this paper. Firstly, multi-phase time-domain fault features are extracted from door's position, driven-motor's speed and current signals. Secondly, the irrelevant and redundant features are removed and the optimal fault features are retained by using distance evaluation technology. The selected optimal fault features are adopted as the input of RF classifier. The fault labels are utilized to formulate an intelligent fault diagnosis model. Finally, the fault diagnosis model can realize the online automatic recognition of different incipient faults in the door sub-system. Experiments are conducted on the bench testing door system of Hangzhou line 4. Results show that the proposed method can extract the early features of incipient faults. Compared with several existing methods, the diagnostic accuracy and robustness of the proposed method are greatly improved after optimal feature selection.

**Keywords:** incipient fault diagnosis; distance evaluation technology; random forests; door system of subway vehicle

## 0 引言

车门是轨道交通车辆中最重要的部件之一,其健康状态直接影响列车运行效率与乘客人身安全。尤其在地铁车辆运营过程中,由于车门开闭频繁、乘客干扰因素

众多,导致客室车门故障频发<sup>[1]</sup>。据统计,车门故障发生次数约占客车系统总故障的30%以上<sup>[2]</sup>。在严重故障发生之前,车门通常会出现一些微弱的故障征兆。及时发现系统常见故障的早期征兆,准确识别故障类型,提前进行系统维护,可极大降低或避免灾难性故障的发生<sup>[3]</sup>,这对提高车门系统的安全性、可靠性、降低故障率具有重要意义。

收稿日期:2019-04-19 Received Date:2019-04-19

\* 基金项目:国家自然科学基金(61873122)、江苏省轨道交通车辆门系统重点实验室(筹)项目(KN17-26)资助

地铁车辆运营过程中累积了大量的车门状态数据,数据驱动的故障诊断与智能运维已成为行业热点方向<sup>[4]</sup>。但对于微小故障,由于其具有隐蔽性、潜伏性、模糊性,难以被发现和捕捉,且不同微小故障之间难以准确区分,迫切需求更先进、更智能的特征提取和选择方法以及相应的故障诊断方法<sup>[5-7]</sup>。现在比较流行的无监督特征提取常常会出现大量无关特征和冗余特征<sup>[8]</sup>,从而影响故障诊断的准确率。因此,能否有效提取凸显故障信息的稳健特征值是数据驱动故障诊断方法的核心基础<sup>[9]</sup>。此外,选择准确率高且性能稳健的分类器也是确保故障诊断方法性能的关键之一。

近年来,随着机器学习的快速发展,如人工神经网络<sup>[10-12]</sup>、K-means 聚类<sup>[13]</sup>、贝叶斯网络<sup>[14-16]</sup>、支持向量机<sup>[17-18]</sup>等相继被应用于车门状态监测与诊断领域,取得了较好的效果。譬如,文献[13]提出了一种基于 K-means 的微小故障诊断方法,在车门台架试验中取得了良好的诊断效果。但该方法应用于实际车门系统时,对环境和乘客干扰的鲁棒性较差。文献[14]利用贝叶斯网络对车门驱动电机进行异常状态分析;文献[17]应用支持向量机辨识车门设备的实时状态。这2种方法均通过训练数据建立单一分类器,然后对测试样本集进行状态识别实现故障诊断。但对于微小故障,由于其故障征兆微弱、故障模式复杂,单一分类器诊断准确率较低。已有研究表明,与单分类器相比,分类器群在复杂的故障诊断问题中能够达到更高正确率<sup>[19-20]</sup>。

本文针对大数据驱动下的微小故障诊断技术中的2大核心步骤<sup>[21]</sup>,首先运用距离评价准则对特征值进行敏感度评估,选择出最优特征子集,然后采用多分类器的典型代表——随机森林(random forests, RF)<sup>[22]</sup>,实现微小故障的智能诊断。本文所提方法已在杭州地铁4号线的台架门系统中进行了应用验证,结果表明本文方法能准确提取早期故障的微弱特征,故障分类模型精度高,故障诊断准确率优于现有其他方法。

## 1 双向对开塞拉门控制系统

### 1.1 车门系统简介和常见故障

本文以地铁车辆中最常见的双向对开塞拉门为研究对象<sup>[23]</sup>,车门系统主要由机械控制和电气控制2部分组成。其中,机械控制部分主要通过长、短导柱和螺杆完成门扇的往复运动;电气控制部分主要通过电子门控制器控制电机的旋转进而控制门扇的移动速度。电气和机械控制部件相互作用,完成车门系统的运动控制。

车门结构复杂,安全苛求,零部件众多,且各部件间配合紧密。随着列车的持续运营,由于元部件老化、磨损等因素,车门系统会逐渐出现一些微小故障,影响车门性能,

微小故障的持续演变最终将导致车门严重故障的发生。本文以车门安全体系规范及实际运营维修返回的相关技术统计为依据,研究7种典型微小故障工况,为工作人员提供维修指导,避免系统严重故障的发生。如表1所示。

表1 车门典型微小故障

Table 1 Typical incipient faults of door system

编号	微小故障状态	故障风险
1	对中尺寸异常	开门阻力增大甚至无法关闭
2	V型尺寸异常	障碍检测功能异常,影响门的密封性
3	门扇高度异常	开关门阻力增大,障碍检测功能异常
4	门扇开度减小	影响开门障碍检测
5	门扇开度增加	障碍检测功能异常
6	密封尺寸减小	关门局部阻力增大
7	密封尺寸增加	开门阻力增大甚至无法关闭

### 1.2 数据预处理和特征初选

以杭州4号线为例,目前可提供的车门监测参数包括:驱动电机采集的车门位置、电机转速和电流信号。正常开、关门时间为3.5 s,采样周期为10 ms,即每条开关门数据大约有350个采样点。

微小故障具有极高的隐蔽性,驱动电机采集的原始参数无法全面的体现车门运行状态。而车门系统各部件是否有开裂、破损、润滑不良,部件之间装配是否合理,滑道是否有异物,这些信息都可以体现在车门的运动曲线中,而且不同部件的故障会导致车门曲线在不同运动阶段的不同形式的异常变化。因此,本文首先提出一种多阶段多值域的特征初选方法。根据车门工作原理,以车门速度曲线为模板,将车门运动过程分为升速段、高速段、减速段、缓行段和到位后段5个阶段,如图1所示,然后对系统测量信号分阶段进行特征提取,建立特征向量集。

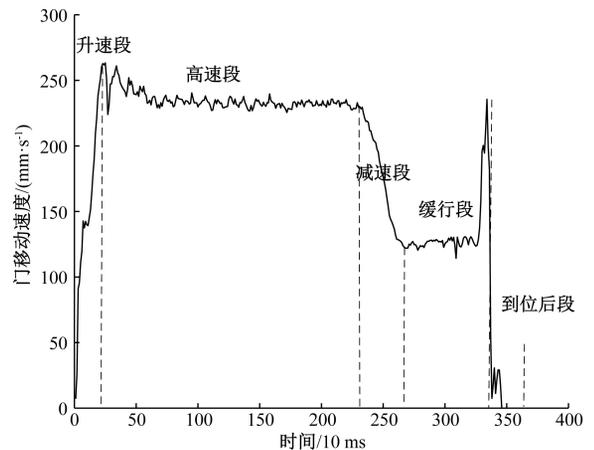


图1 车门运动阶段划分示意图(以开门为例)

Fig.1 Phase segmentation of door movement (e.g. door opening)

目前,基于数据的元特征提取方法主要有时域分析法、频域分析法和时频域分析法<sup>[24]</sup>。时域分析法是最常用的特征提取方法。通过计算各种时域系数,得到时间序列的低阶或高阶统计量,计算简单,物理意义明确。考虑车门测量频域信息匮乏,本文主要提取时域统计特征,即通过计算车门位置、电机转速和电流信号在车门运动各个阶段中的极值、均值、方差、偏度、峰度等指标构建故障特征向量集  $F \in R^{N \times P}$ ,

$$F = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{bmatrix} = \begin{bmatrix} f_{11} & f_{12} & \cdots & f_{1P} \\ f_{21} & f_{22} & \cdots & f_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ f_{N1} & f_{N2} & \cdots & f_{NP} \end{bmatrix} \quad (1)$$

式中:  $N$  表示开关门数据个数;  $P$  表示初选的原始特征数量。

本文一共提取 180 个开关门数据特征,即  $P = 2 \times 5 \times 3 \times 6 = 180$ , 其中,2 表示开门和关门共 2 条曲线,5 表示 5 个运动阶段,3 表示 3 个观测参数,6 表示 6 个统计特征。每一次开关门运动曲线可提取 1 个初始特征向量,  $f = [f_1, f_2, \dots, f_{180}]$ 。

这些特征又可分为有量纲特征和无量纲特征<sup>[25]</sup>, 其中,偏度和峰度属于无量纲特征,不受门系统各种工况的影响,对微小故障有良好的诊断效果。最大值、最小值、平均值和方差属于有量纲特征,可以反映车门阻力大小及信号的波动等情况。

## 2 基于距离评估准则的特征优选算法

特征选择是从 1 组特征中挑选出一些最有效的特征以降低特征空间维数的过程<sup>[26]</sup>。依据其是否独立于后续分类器的学习算法,特征选择方法可以分为过滤式(Filter)和封装式(Wrapper)2种<sup>[27]</sup>。Filter 模式选择效率较高,但 Filter 模式选择出来的特征子集与现有常用分类器算法的结合效率较低,选择效果不理想最终导致分类效果较差。Wrapper 模式利用分类器的训练准确率评估特征子集,与分类器契合度高,因而分类准确率高。

本文采用 Wrapper 模式下的距离评估准则(DET)来进行特征优选,利用特征间距离测度的评价函数筛选出一定数量的特征子集。具体的评估原则是:同一类故障数据的类内特征距离较小,不同类的故障数据的类间特征距离较大。具体计算步骤如下<sup>[28]</sup>。

1) 计算第  $i$  个特征在第  $j$  类故障数据中的类内距离,

$$d_{i,j} = \frac{1}{N(N-1)} \sum_{m,n=1}^N |F_{i,j}(m) - F_{i,j}(n)| \quad (2)$$

$i = 1, 2, \dots, P; j = 1, 2, \dots, M;$   
 $m, n = 1, 2, \dots, N, m \neq n.$

式中:  $N$  表示样本数;  $P$  表示初始特征数;  $M$  表示微小故

障类别数;  $F_{i,j}(m)$  和  $F_{i,j}(n)$  分别表示样本  $m$  和样本  $n$  在第  $j$  类数据中第  $i$  个特征值。本文中  $P = 180, M = 7$ 。

然后,计算第  $i$  个特征值在  $M$  类故障中的类内距离的平均值:

$$D_i = \frac{1}{M} \sum_{j=1}^M d_{i,j} \quad (3)$$

2) 计算  $N$  个样本的第  $i$  个特征在第  $j$  类故障数据中的平均值:

$$q_{i,j} = \frac{1}{N} \sum_{n=1}^N F_{i,j}(n) \quad (4)$$

然后,计算第  $i$  个特征值在  $M$  类故障的类间距离的平均值:

$$D'_i = \frac{1}{M(M-1)} \sum_{u,w=1}^M |F_{i,u} - F_{i,w}| \quad (5)$$

$u, w = 1, 2, \dots, M, u \neq w$

式中:  $q_{i,u}$  和  $q_{i,w}$  分别表示第  $u$  类和第  $w$  类故障数据中  $N$  个样本的第  $i$  个特征的平均值。

3) 计算第  $i$  个特征的评估因子:

$$\alpha_i = \frac{D'_i}{D_i} \quad (6)$$

当特征  $i$  的类间距离越大,类内距离越小时,评估因子  $\alpha_i$  的值越大。也就是说,越能够满足评估原则的特征,其评估因子的值就越大,该特征也就越敏感,对微小故障正确分类的作用也越大。

4) 计算完所有特征的评估因子  $\alpha_i (i = 1, 2, \dots, 180)$  后,按照  $\alpha_i$  值的大小对特征进行降序排列,逐一增加特征的个数作为后续的 RF 算法的输入。本文使用的特征输入终止条件为:(1) 故障分类准确率达到 95%;(2) 特征连续增加  $T$  个(一般选择  $T = 5$ <sup>[29]</sup>),分类准确率没有提高。当满足上述任一终止条件时,即停止输入特征,此时选择的特征作为对故障进行分类的最优特征子集  $S$ 。

为了更清楚地解释该特征优选方法,图 2 以特征 a 和特征 b 在 3 类故障数据中的特征评估为例进行解释。将特征 a 和特征 b 进行敏感度对比,其中每类数据各 5 个样本,即  $M = 3, P = 2, N = 5$ 。从图中可以看出,特征 a 可以准确地将数据分为 3 类,特征 a 的类内距离小于特征 b 的类内距离,即  $D_a < D_b$ , 且特征 a 的类间距离大于特征 b 的类间距离,即  $D'_a > D'_b$ , 由式(6)可得出,  $\alpha_a > \alpha_b$ 。因此在对这 3 类故障数据进行分类时,特征 a 比特征 b 更加敏感。

## 3 基于 RF 的智能故障诊断方法

### 3.1 RF

由 Breiman 于 2001 年提出的 RF<sup>[17]</sup> 是一种利用随机重采样技术(bootstrap)和节点随机分割技术构造多

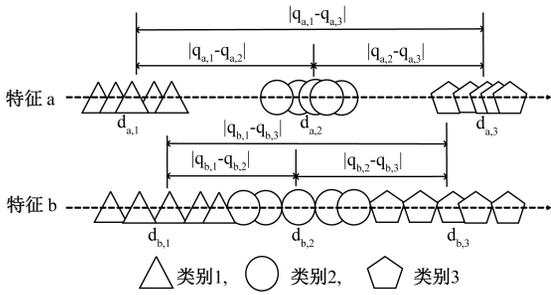


图 2 特征敏感度

Fig.2 Feature sensitivity

个决策树 (decision tree, DT) 并通过投票获得最终分类结果的集成机器学习方法。RF 是由众多 DT 分类器组成的综合分类器  $\{h(x, \Theta_k), k = 1, 2, \dots, K\}$ , 其中参数集  $\{\Theta_k\}$  是服从独立同分布的随机向量,  $K$  是随机林中决策树的个数, 在给定的自变量  $x$  下, 每个 DT 分类器通过投票确定最优分类结果<sup>[30]</sup>, 图 3 所示为 RF 方法的原理。

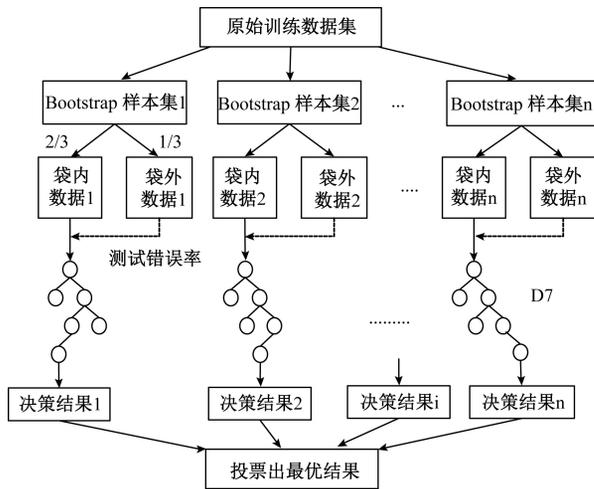


图 3 RF 方法原理

Fig.3 Scheme of the RF

本文生成 RF 的步骤如下。

1) 通过 bootstrap 重采样技术, 从故障数据训练集中有放回地重复随机抽取  $c$  个样本, 重复  $K$  次 (即生成树的数目为  $K$ ), 每次抽取的 bootstrap 样本集为 DT 生长的训练集。

根据文献[31]可知, RF 的诊断错误率在前期随着树数量的增加, 分类错误率大幅降低, 当 DT 数量达到 40 以后, 错误率趋于稳定 (大约为 0.001 6), 但是, 模型的复杂度和训练时间会随着树的数量变多而增加。因此, 综合考虑模型的准确率与复杂度, 本文 RF 分类树的数量  $K=40$ 。

2) 随机选取候选分割特征, 使分类器在结构上有所不同。本文按照  $\alpha_i$  值的大小逐一增加特征的个数作为 RF 算法的输入, 每次从  $i$  个输入特征中随机选取  $I$  个特征, 且  $I \leq 180$ 。根据文献[22]给出的经验公式, 通常取  $I = \text{int}(\sqrt{i})$ , 然后从选出的  $I$  个特征中采用信息增益的策略, 选择一个特征作为该节点的分裂特征。重复上述过程构建每棵决策树分类器的分支, 直到单棵 DT 能够描述样本分类或遍历所有特征属性, 在上述过程中  $I$  的值保持不变。

3) 每棵树都在没有任何剪裁的情况下最大限度地生长。

4) 将生成的树组为 RF, 根据 RF 的投票数对新数据进行分类, 投票规则如下, 根据投票规则, 最终判断出车门的微小故障状态。

$$H(x) = \arg \max_{i=1}^k Z(h_k(x) = M) \quad (7)$$

式中:  $H(x)$  为组合分类模型;  $h_k(x)$  是单个 DT 分类模型;  $M$  是故障类别归属;  $Z(\cdot)$  为指示函数, 它的取值范围是 0 和 1, 当括号中的条件成立时值为 1, 否则为 0。最终选择投票最多的故障类别作为样本数据的最终类别。

### 3.2 微小故障诊断模型

基于特征优选和 RF 分类器的微小故障诊断建模过程流程图如图 4 所示。

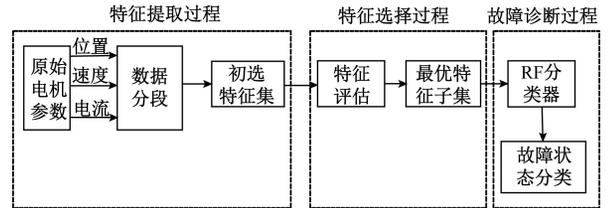


图 4 微小故障诊断模型

Fig.4 Diagnosis model for incipient faults

详细建模步骤如下。

1) 数据采集和预处理。采集 7 种微小故障工况下的电机原始参数, 每种工况下采集  $N$  组样本, 并对数据进行预处理, 包括对数据的对齐和劣点数据去除。

2) 数据分段和特征提取。依据开关门的速度曲线和工作原理采用多阶段多值域的特征初选方法, 逐段计算时域特征组成系统特征集  $[f_1, f_2, \dots, f_{180}]$ 。

3) 每种故障状态数据集  $N$  中选择  $n$  个样本为训练样本, 其余样本作为测试样本。对初选的特征集进行特征选择过程如下: 首先, 计算出全部特征的类内距离  $D$  和类间距离  $D'$ , 并根据式(6)计算出评估因子  $\alpha_i (i = 1, 2, \dots, 180)$ 。然后, 按照  $\alpha_i$  值的大小对特征进行降序排列, 逐一增加特征的个数输入 RF 分类器进行训练并测试, 得出分类准确率, 最后, 根据终止条件选择出最优特征子集  $S$ 。

4) 将最优特征子集  $S$  作为 RF 分类器的输入建立预测模型  $C = [1, 2, 3, 4, 5, 6, 7]$ , 用训练好的预测模型对测试样本进行分类, 最终识别出车门系统的微小故障类型。

### 4 实验与结果分析

#### 4.1 实验简介

为验证本文方法的有效性和工程应用可行性, 本文以杭州地铁 4 号线的台架车门为应用测试平台。首先检查台架车门的相关尺寸和功能是否符合检验标准并确认该门未发生任何形式的故障, 然后通过人为有意调整车门部件的尺寸、施加全程或局部阻力、安装正线列车的磨损部件等手段, 模拟单因子影响下的 7 种微小故障工况, 执行数周的开关门循环操作, 获取相关的实验数据。

#### 4.2 基于特征优选和 RF 分类器的微小故障诊断结果

首先, 采集 7 类典型的微小故障类型车门数据各 200 组, 样本总容量为 1 400 组。微小故障类型如表 1 所示, 分别为: 对中异常、V 型异常、高度异常、门扇开度减小、门扇开度增加, 密封尺寸减小, 密封尺寸增加。为 7 种微小故障状态分别标上标签  $C = [1, 2, 3, 4, 5, 6, 7]$ , 即故障类别  $M = 7$ , 每类故障样本数  $N = 200$ 。每种工况下, 随机选择其中 160 组作为训练数据, 剩余 40 组作为测试数据。

然后, 采用多阶段多值域的特征初选方法对电机原始参数进行特征提取, 共提取到 180 个开关门特征向量, 即  $P = 180$ , 并建立原始特征集。

对 180 个初选特征进行特征评估, 计算各特征的评估因子  $\alpha_i (i = 1, 2, \dots, 180)$ , 如图 5 所示。

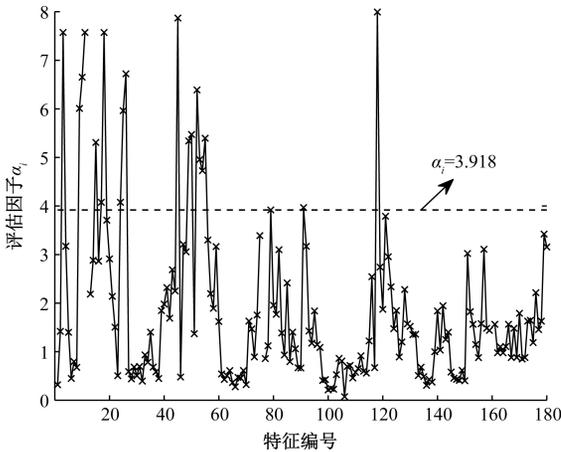


图 5 特征集的评估因子

Fig.5 Evaluation factors for all features

根据从大到小的顺序, 按照  $\alpha_i$  值的大小对特征进行降序排列, 逐一增加特征的个数输入 RF 分类器进行训练和测试。当达到终止条件时, 输入 RF 分类器的特征

增加到 20, 即选择出 20 个敏感特性组成最优特征子集  $[S_1, S_2, \dots, S_{20}]$ , 即位于图 5 中水平线的上方的特征, 水平线的高度对应于第 20 个敏感特征评估因子的大小  $\alpha_{20} (\alpha_{20} = 3.918)$ 。表 2 所示为所选敏感特征的序号和相应的评估因子  $\alpha_i$ 。

表 2 敏感特征

Table 2 Sensitive features

敏感特征编号	特征序号	$\alpha_i$	敏感特征编号	特征序号	$\alpha_i$
1	118	7.994	11	50	5.396
2	45	7.574	12	55	5.340
3	3	7.574	13	49	5.310
4	11	7.574	14	15	4.957
5	18	6.722	15	53	4.727
6	26	6.653	16	54	4.076
7	10	6.390	17	17	4.076
8	52	6.010	18	24	4.076
9	9	5.962	19	91	3.968
10	25	5.474	20	79	3.918

最后, 将最优特征子集  $[S_1, S_2, \dots, S_{20}]$  输入 RF 建立预测模型, 用训练好的预测模型对测试样本进行分类, 图 6 所示为使用原始特征集时 RF 分类器对微小故障的诊断情况, 图 7 所示为特征选择后 RF 分类器的诊断情况, 对比可以看出, 特征选择前 RF 对于门扇开度减小和门扇开度增加这 2 种微小的错误故障诊断较多, 特征选择后有了明显的改善。计算出全部特征集  $[f_1, f_2, \dots, f_{180}]$  和最优特征子集  $[S_1, S_2, \dots, S_{20}]$  的平均分类准确率

$$\bar{r}\% = \frac{1}{c} \sum_{i=1}^c r_i\%, \quad c = 7, \text{ 并将二者进行比较, 如表 3 所示。}$$

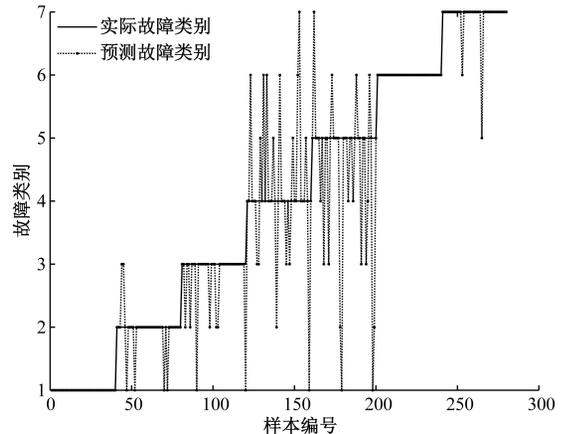


图 6 基于原始特征和 RF 分类器的微小故障诊断结果

Fig.6 Diagnosis results based on original feature set and the RF classifier-based method

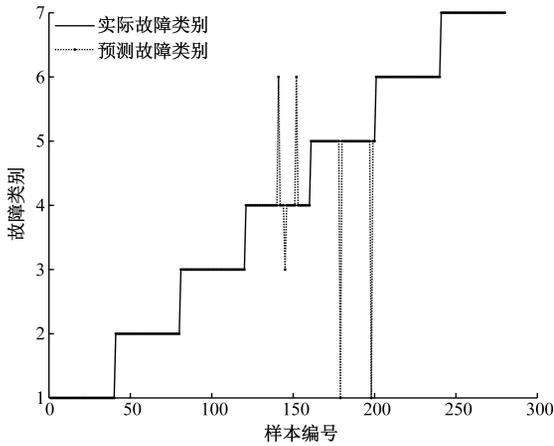


图 7 基于优选特征和 RF 分类器的故障诊断结果

Fig.7 Diagnosis results for RF based method after optimal feature selection

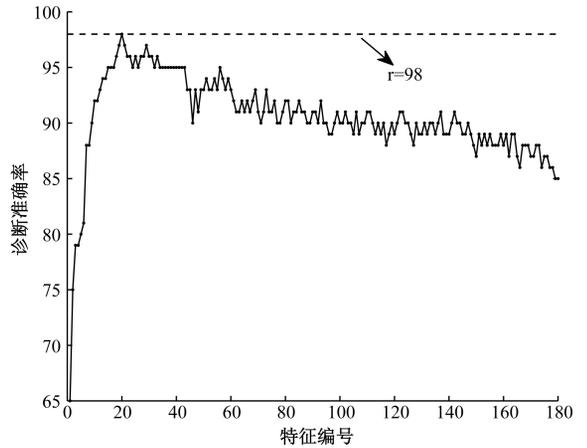


图 8 诊断准确率与输入特征个数之间的关系

Fig.8 The relationship between diagnosis accuracy and the number of selected features

表 3 不同特征集用于诊断的准确率比较

Table 3 Diagnosis accuracy for RF based method with the original feature set and the optimal feature set

特征集	样本数量	诊断错误数	$\bar{r}/\%$
原始特征集	280	42	85
最优特征子集	280	6	98

通过表 3 的对比结果可知,分类平均准确率  $\bar{r}\%$  从 85% 提高至 98%。原始特征集分类准确度相对较低的原因是某些特性对门系统的微小故障状态不敏感。也就是说,提取的 180 个特征不能准确识别不同状态之间的差异,从而影响了分类的准确性。然而,通过对原始特征集  $[f_1, f_2, \dots, f_{180}]$  进行特征评估,选择出包含 20 个敏感特征的最优特征子集  $[S_1, S_2, \dots, S_{20}]$  作为 RF 分类器的输入,可以显著提高故障诊断准确率  $\bar{r}\%$ 。

逐步将 RF 分类器的输入特征数从 1 个增加到 180 个,并对其进行训练和测试。为了使平均分类准确率  $\bar{r}\%$  与输入特征的个数  $i$  之间的关系更清晰,绘制了它们之间的关系曲线,如图 8 所示。可以发现,开始增加的特征的敏感性比较强,因此  $\bar{r}\%$  随着特征数量的增加而提高。但当输入 RF 分类器的特征数量在一定程度上继续增加时,分类准确率  $\bar{r}\%$  开始减少。这是由于随后增加的特征对车门系统 7 种微小故障状态的分类不敏感。随着特征数量的增加,门系统的几种微小故障状态之间的差异越来越模糊,因此  $\bar{r}\%$  随着特征数量的增加而减小。结合表 3 和图 8 可以看出,原始特征集中包含大量的冗余特征和无关特征。因此,去除不利于分类甚至影响分类准确率的不敏感特征,不仅可以减小数据集规模,而且可以提高 RF 分类器的分类准确率。

### 4.3 RF 分类器和常用分类算法的诊断结果和对比分析

为了验证 RF 分类器在微小故障诊断问题上的优势,本文选择了 3 种对比性方法,包括:应用较为广泛的贝叶斯网络,SVM 算法,以及作者早期开发并已在广州地铁正线上成功应用的基于 K-means 聚类的故障诊断算法。其中,SVM 采用 C-SVM, K-means 中选择类簇数  $K=7$ 。图 9~11 所示为 3 种方法基于原始特征集的诊断结果,图 12~14 所示为 3 种方法基于最优特征子集的诊断结果,其中,RF 分类器、贝叶斯、SVM 和 K-means 的诊断准确率如表 4 所示。结果可知,使用最优特征集后,几种算法的准确率都得到提高,其中 RF 分类器的诊断准确率达到到了 98%,而其余 3 种算法的准确率均没有达到 90%。通过和其余 3 种算法的诊断结果进行对比可知,在同样样本数据的条件下,无论是否进行特征选择,RF 分类器都具有明显的优越性,使用 RF 分类器进行微小故障诊断可以获得更精确的诊断结果。

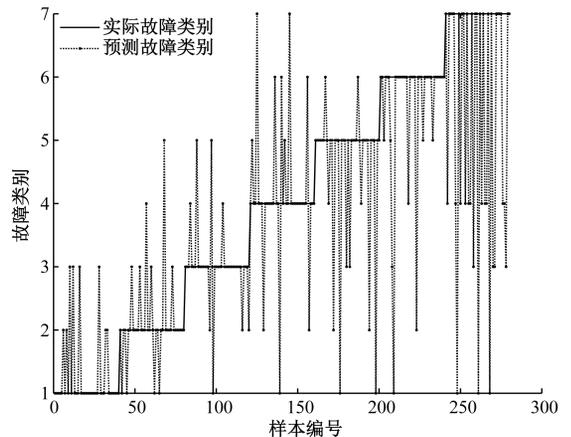


图 9 基于原始特征和贝叶斯网络的微小故障诊断结果

Fig.9 Diagnosis results based on the original feature set and Bayesian network based method

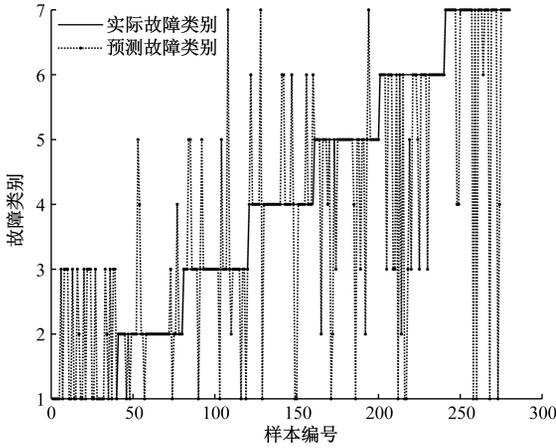


图 10 基于原始特征和 SVM 的微小故障诊断结果  
Fig.10 Diagnosis results based on the original feature set and SVM based method

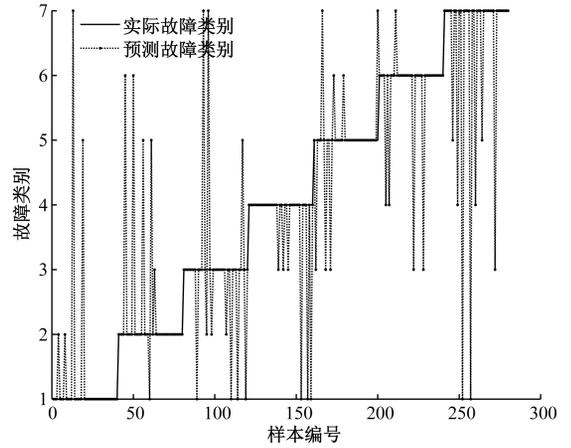


图 13 基于优选特征和 K-means 的微小故障诊断结果  
Fig.13 Diagnosis results for K-means based method after optimal feature selection

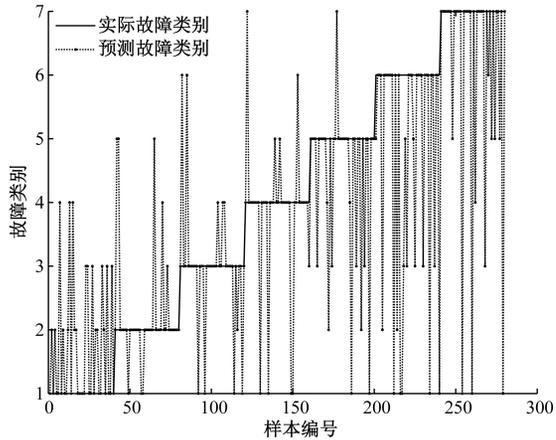


图 11 基于原始特征和 K-means 的微小故障诊断结果  
Fig.11 Diagnosis results based on the original feature set and K-means based method

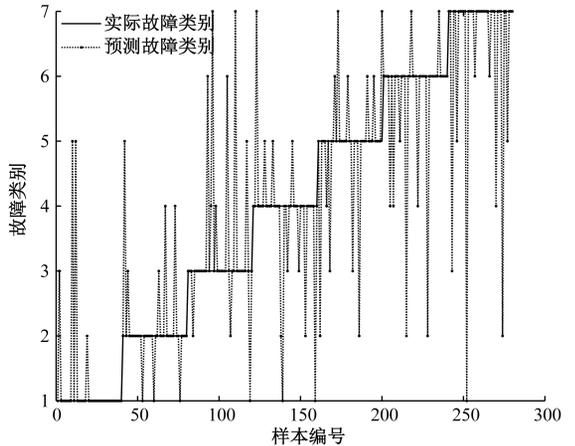


图 14 基于优选特征和 SVM 的微小故障诊断结果  
Fig.14 Diagnosis results for SVM-based method after optimal feature selection

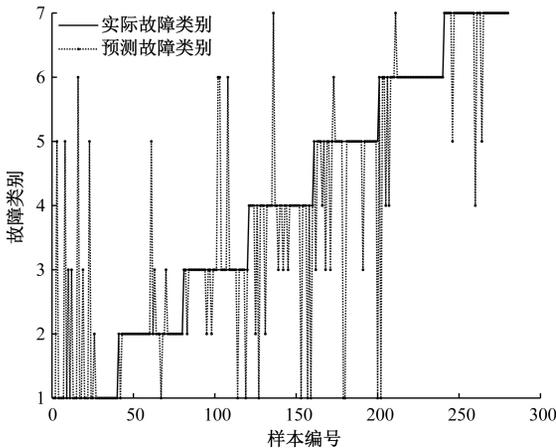


图 12 基于优选特征和贝叶斯网络的微小故障诊断结果  
Fig.12 Diagnosis results for Bayesian network based method after optimal feature selection

表 4 不同学习算法诊断准确率比较

Table 4 Comparison of diagnostic accuracy of different learning algorithms

学习算法	样本数量	诊断错误数量	原始特征集的 诊断准确率/%	最优特征集的 诊断准确率/%
RF	280	42	85	98
贝叶斯	280	66	76	85
SVM	280	70	75	85
K-means	280	75	73	80

## 5 结 论

车门系统微小故障征兆微弱、不确定性强且不同故

障之间难以区分,迫切需求更先进更智能的特征选择算法和故障诊断分类器。本文提出一种大数据驱动的车门故障特征优选方法和基于随机森林的智能诊断方法,得出以下结论。

1) 根据车门运动特性,分段提取监测参数的时域特征,能提取大量有效特征,但其中也包括了許多无用特征和冗余特征。

2) 通过 DET 特征选择方法对提取到的特征进行敏感度评估,可选择出最优特征子集作为 RF 分类器的输入。

3) 在相同样本数据的条件下,RF 诊断模型比其他几种经典模型的诊断准确率高,更加适用于微小故障诊断。

实验结果表明,该智能诊断方法能够有效捕捉微小故障信息,准确诊断出故障类型,并克服了基于 K-means 算法的诊断方法鲁棒性差等缺陷,应用前景明确。

## 参考文献

[1] 高文明,李志保,唐谦,等.轨道车辆门远程监控与故障诊断系统软件设计[J].机械设计与制造工程,2016,45(10):49-51.  
GAO W M, LI ZH B, TANG Q, et al. Development of remote monitoring and fault diagnosis system for the doors of railway vehicle [J]. Machine Design and Manufacturing Engineering, 2016, 45(10): 49-51.

[2] FINK O, ZIO E, WEIDMANN U. Fuzzy classification with restricted Boltzman machines and echo-state networks for predicting potential railway door system failures[J]. IEEE Transactions on Reliability, 2015, 64(3): 861-868.

[3] MIAO Q, MAKIS V. Condition monitoring and classification of rotating machinery using wavelets and hidden Markov models[J]. Mechanical Systems & Signal Processing, 2007, 21(2): 840-855.

[4] 姜斌,吴云凯,陆宁云,等.高速列车牵引系统故障诊断与预测技术综述[J].控制与决策,2018,33(5):841-855.  
JIANG B, WU Y K, LU N Y, et al. Review of fault diagnosis and prognosis techniques for high-speed railway traction system [J]. Control and Decision, 2018, 33(5): 841-855.

[5] 周东华,纪洪泉,何潇.高速列车信息控制系统的故障诊断技术[J].自动化学报,2018,44(7):1153-1164.  
ZHOU D H, JI H Q, HE X. Fault diagnosis techniques for the information control system of high-speed trains [J]. Acta Automatica Sinica, 2018, 44(7): 1153-1164.

[6] 杨宇,潘海洋,程军圣.基于特征选择和 RRVMCD 的滚动轴承故障诊断方法[J].振动工程学报,2014,27(4):629-636.  
YANG Y, PAN H Y, CHENG J SH. The rolling bearing fault diagnosis method based on the feature selection and RRVMCD [J]. Journal of Vibration Engineering, 2014, 27(4): 629-636.

[7] 胡金海,谢寿生,骆广琦,等.基于核函数 Fisher 鉴别分析的特征提取方法[J].振动、测试与诊断,2008,28(4):322-326.  
HU J H, XIE SH SH, LUO G Q, et al. Feature extraction method based on kernel-based fisher discriminant analysis [J]. Journal of Vibration, Measurement & Diagnosis, 2008, 28(4): 322-326.

[8] 苏祖强,汤宝平,赵明航,等.基于多故障流形的旋转机械故障诊断[J].振动工程学报,2015,28(2):309-315.  
SU Z Q, TANG B P, ZHAO M H, et al. Rotating machinery fault diagnosis based on multiple fault manifold [J]. Journal of Vibration Engineering, 2015, 28(2): 309-315.

[9] 苗强, MAKIS V. 基于隐马尔科夫模型的故障诊断系统研究[J].航空学报,2005,26(5):641-646.  
MIAO Q, MAKIS V. Condition monitoring of rotating machinery using hidden Markov models [J]. Acta Aeronautica et Astronautica Sinica, 2005, 26(5): 641-646.

[10] 李建伟,程晓卿,秦勇,等.基于 BP 神经网络的城市轨道交通车辆可靠性预测[J].中南大学学报(自然科学版),2013(s1):42-46.  
LI J W, CHENG X Q, QIN Y, et al. Reliability prediction of urban rail transit vehicle based on BP neural network [J]. Journal of Central South University (Science and Technology), 2013(s1): 42-46.

[11] 姜洪开,邵海东,李兴球.基于深度学习的飞行器智能故障诊断方法[J].机械工程学报,2019,55(7):27-34.  
JIANG H K, SHAO H D, LI X Q. Deep learning theory with application in intelligent fault diagnosis of aircraft [J]. Journal of Mechanical Engineering, 2019, 55(7): 27-34.

[12] 张宁,魏秀业,郭小勇,等.基于改进鱼群算法优化神经网络的轴承故障诊断研究[J].轴承,2019(5):44-48.  
ZHANG N, WEI X Y, GUO X Y, et al. Study on fault diagnosis for bearings based on improved fish swarm algorithm to optimize neural network [J]. Bearing, 2019

- (5): 44-48.
- [13] HAN G W, ZHANG Y, LU N Y, et al. Incipient anomaly detection for railway vehicle door system based on adaptive mean shift clustering [ C ]. Chinese Automation Congress. IEEE, 2017:1297-1302.
- [14] 周巧莲, 金碧筠, 冒玲丽, 等. 基于贝叶斯网络地铁车门系统可靠性分析和故障诊断 [ J ]. 沈阳工业大学学报, 2014, 36(4):441-445.
- ZHOU Q L, JIN B Y, MAO L L, et al. Reliability analysis and fault diagnosis of metro door system based on Bayesian network [ J ]. Journal of Shenyang University of Technology, 2014, 36(4):441-445.
- [15] 徐霖. 城轨列车客室车门系统故障诊断方法研究——基于改进的 TOPSIS 法与贝叶斯网络 [ D ]. 北京: 北京交通大学, 2015.
- XU L. Research on fault diagnosis method of urban rail vehicle passenger compartment door system: Based on improved TOPSIS method and Bayesian network [ D ]. Beijing: Beijing Jiaotong University, 2015.
- [16] 王金鑫, 王忠巍, 马修真, 等. 基于贝叶斯网络的柴油机润滑系统多故障诊断 [ J ]. 控制与决策, 2019, 34(6): 1187-1194.
- WANG J X, WANG ZH W, MA X ZH, et al. Diagnosis of multiple faults of diesel engine lubrication system based on Bayesian networks [ J ]. Control and Decision, 2019, 34(6): 1187-1194.
- [17] 梁治华, 曹江涛, 姬晓飞. 基于 EEMD 和 CS-SVM 的滚动轴承故障诊断研究 [ J ]. 机电工程, 2019 ( 6 ): 622-627.
- LIANG ZH H, CAO J T, JI X F. Fault diagnosis of rolling bearing based on EEMD and CS-SVM [ J ]. Journal of Mechanical & Electrical Engineering, 2019 ( 6 ): 622-627.
- [18] 洪杰, 韩磊, 苗学问, 等. 基于支持向量机的滚动轴承状态寿命评估 [ J ]. 北京航空航天大学学报, 2010, 36(8):896-899.
- HONG J, HAN L, MIAO X W, et al. Assessment based on support vector machine for rolling bearing grade-life [ J ]. Journal of Beijing University of Aeronautics and Astronautics, 2010, 36(8):896-899.
- [19] 张钰, 陈珺, 王晓峰, 等. 随机森林在滚动轴承故障诊断中的应用 [ J ]. 计算机工程与应用, 2018, 54(6): 100-104, 114.
- ZHANG Y, CHEN J, WANG X F, et al. Application of random forest on rolling element bearings fault diagnosis [ J ]. Computer Engineering and Applications, 2018, 54(6): 100-104, 114.
- [20] 胡青, 孙才新, 杜林, 等. 核主成分分析与随机森林相结合的变压器故障诊断方法 [ J ]. 高电压技术, 2010, 36(7): 1725-1729.
- HU Q, SUN C X, DU L, et al. Transformer fault diagnosis method using random forests and Kernel principle component analysis [ J ]. High Voltage Engineering, 2010, 36(7): 1725-1729.
- [21] 李远军, 孙继炫. 基于特征提取与识别两阶段的汽车电机轴承故障诊断 [ J ]. 电子测量与仪器学报, 2019, 33(2): 56-63.
- LI Y J, SUN J X. Car motor bearing fault diagnosis based on fault feature extraction and recognition stages [ J ]. Journal of Electronic Measurement and Instrumentation, 2019, 33(2): 56-63.
- [22] BREIMAN L. Random forest [ J ]. Machine Learning, 2001, 45:5-32.
- [23] 贡智兵, 王祖进. 城轨塞拉门综合评价方法研究 [ J ]. 新型工业化, 2018, 8(6): 77-82, 89.
- GONG ZH B, WANG Z J. The Journal of new industrialization [ J ]. The Journal of New Industrialization, 2018, 8(6): 77-82, 89.
- [24] CHENG X, XING Z, QIN Y, et al. Reliability analysis of metro door system based on FMECA [ J ]. Journal of Intelligent Learning Systems & Applications, 2014, 5(4):216-220.
- [25] 徐国权, 罗倩, 郭鹏飞. 多特征评估筛选的滚动轴承故障诊断算法 [ J ]. 计算机仿真, 2018, 35(12): 446-450, 455.
- XU G Q, LUO Q, GUO P F. Rolling bearing fault diagnosis algorithm based on multi feature evaluation and selection [ J ]. Computer Simulation, 2018, 35(12): 446-450, 455.
- [26] 郑宝芬, 苏宏业, 罗林. 无监督特征选择在时间序列数据挖掘中的应用 [ J ]. 仪器仪表学报, 2014, 35(4): 834-840.
- ZHENG B F, SU H Y, LUO L. Application of unsupervised feature selection in time series data mining [ J ]. Chinese Journal of Scientific Instrument, 2014, 35(4): 834-840.
- [27] 姚旭, 王晓丹, 张玉玺, 等. 特征选择方法综述 [ J ]. 控制与决策, 2012, 27(2):161-166, 192.
- YAO X, WANG X D, ZHANG Y X, et al. Summary of feature selection algorithms [ J ]. Control and Decision, 2012, 27(2):161-166, 192.
- [28] YANG B S, HAN T, AN J L. ART-KOHONEN neural

network for fault diagnosis of rotating machinery [J]. Mechanical Systems and Signal Processing, 2004, 18(3):645-657.

- [29] 雷亚国,何正嘉,訾艳阳,等.基于特征评估和神经网络的机械故障诊断模型[J].西安交通大学学报,2006,40(5):558-562.

LEI Y G, HE ZH J, ZI Y Y, et al. Mechanical fault diagnosis model based on feature evaluation and neural networks [J]. Journal of Xi'an Jiaotong University, 2006, 40(5):558-562.

- [30] 贺思捷,刘大同,彭宇.遥测数据驱动的无人机飞行状态识别方法[J].仪器仪表学报,2016,37(9):2004-2013.

HE S J, LIU D T, PENG Y. Flight mode recognition method of the unmanned aerial vehicle based on telemetric data [J]. Chinese Journal of Scientific Instrument, 2016, 37(9):2004-2013.

- [31] CUTLER K, BREIMAN L. Random forests[J]. Machine Learning, 2004, 45(1):157-176.

## 作者简介



**施文**,2017年于安徽大学获得学士学位,现为南京航空航天大学硕士研究生,主要研究方向为数据驱动的故障诊断与健康管理等。

E-mail: 18225602360@163.com

**Shi Wen** received her B. Sc. degree from Anhui University in 2017. She is currently a M. Sc. candidate at Nanjing University of Aeronautics and Astronautics. Her main research interests include data-driven fault diagnosis and health management.



**陆宁云**(通信作者),1998年于东北大学获得学士学位,2000年于东北大学获得硕士学位,2004年于东北大学获得博士学位,现为南京航空航天大学教授,主要研究方向为数据驱动的故障诊断、故障预测与健康管理等。

E-mail: luningyun@nuaa.edu.cn

**Lu Ningyun** (Corresponding author) received her B. Sc. degree, M. Sc. degree, and Ph. D. degree all from Northeastern University in 1998, 2000, and 2004, respectively. She is currently a professor at Nanjing University of Aeronautics and Astronautics. Her main research interests include data-driven fault diagnosis, fault prediction and health management.



**许志兴**,1993年于南京工业大学获得学士学位,1997年于南京工业大学获得硕士学位,2001年于南京航空航天大学获得博士学位,现为南京康尼机电股份有限公司高级工程师,主要研究方向为轨道车辆门系统的远程监测、故障智能诊断、车门的智能运维系统等。

系统等。

E-mail: 1018481966@qq.com

**Xu Zhixing** received his B. Sc. degree and M. Sc. degree both from Nanjing Tech University in 1993 and 1997 respectively, and received his Ph. D. degree from Nanjing University of Aeronautics and Astronautics in 2001. He is currently a senior engineer at Nanjing Kangni Mechanical and Electrical Co., Ltd. His main research interests include remote monitoring, intelligent fault diagnosis and intelligent maintenance system of railway vehicle door system.