

# 基于 SMOTE-IPF 和 SDAE 的行星齿轮箱故障诊断方法研究

赵亚磊,王友仁,钱心筠,孙泽金,张鲁晋  
(南京航空航天大学 自动化学院,江苏 南京 210016)

**摘要:**针对现有故障诊断模型在故障样本缺乏时诊断率低的问题,提出一种基于迭代过滤合成少数类过采样方法(SMOTE-IPF)和堆叠去噪自动编码器(SDAE)的故障诊断模型。该方法利用 SDAE 对样本进行特征提取,使用 SMOTE-IPF 在合成新样本的同时利用多个决策树对新样本进行投票过滤,使数据集达到平衡,最后使用分类器进行故障分类。通过行星齿轮实验平台进行实验,验证了所提方法在故障样本极度缺乏下故障诊断的有效性。

**关键词:**故障诊断;样本不平衡;合成少数类过采样技术;堆叠去噪自动编码器;行星齿轮箱  
**中图分类号:**TH132.425 **文献标志码:**A **文章编号:**1671-5276(2023)05-0042-04

## Research on Fault Diagnosis Method of Planetary Gearbox Based on SMOTE-IPF and SDAE

ZHAO Yalei, WANG Youren, QIAN Xinyun, SUN Zejin, ZHANG Lujin

(College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

**Abstract:** To solve the problem of low diagnosis rate of existing fault diagnosis models when in lack of fault samples, proposes a fault diagnosis model based on iterative filtering synthetic minority class oversampling method (SMOTE-IPF) and stacked denoising autoencoder (SDAE). This method uses SDAE to extract features from samples, applies SMOTE-IPF to synthesize new samples and uses multiple decision trees to vote and filter new samples to balance the dataset, and adopts a classifier to classify faults. Experiments are carried out on the planetary gear experimental platform to verify the effectiveness of the proposed method in fault diagnosis in the extreme lack of fault samples.

**Keywords:** fault diagnosis; sample imbalance; synthetic minority over-sampling technique; stacked denoising autoencoder; planetary gearbox

## 0 引言

行星齿轮箱作为机械传动系统的关键部件,广泛应用于航空航天、风力发电、冶金、化工等领域。复杂的传动结构和恶劣的工况环境极易导致齿轮发生故障。一般的故障诊断模型需要大量的故障信号进行模型训练,而行星齿轮箱在实际运行过程中产生的振动数据主要以正常样本为主,故障样本往往较少。

针对不平衡样本下的故障诊断问题,WANG Z R 等<sup>[1]</sup>建立了 GAN 的生成器和判别器的网络结构,使用 SDAE 构建网络,用来学习原样本的数据分布,实现了行星齿轮箱在不均衡样本下的故障诊断能力。GUO Q W 等<sup>[2]</sup>融合了一维卷积神经网络和辅助分类生成对抗网络(ACGAN),该方法在扩充了高质量样本的同时,也提升了不均衡数据下轴承故障诊断的准确率。RAMENTOL E 等<sup>[3]</sup>在数据不平衡的情况下,利用改进的 SMOTE-FRST-2T 算法扩充建立的特征集使其达到平衡,完成了不平衡样本下齿轮的故障诊断。以上方法中,GAN 需要大量数据进行生成器和判别器的模型训练,而 SMOTE 方法也存在噪声和边界问

题,当故障样本严重缺乏时,以上方法的故障诊断正确率会出现较大幅度的下降。

本文提出了一种基于迭代过滤合成少数类过采样方法(SMOTE-IPF)和堆叠去噪自动编码器(SDAE)的不平衡样本故障诊断方法,该方法利用 SDAE 对样本进行自动特征提取,这种无监督的特征提取方法在降低样本维度的同时,可以减少样本不平衡对故障样本的影响。然后使用 SMOTE-IPF 方法在合成新样本的同时利用多个决策树对样本进行迭代投票过滤,直至数据集达到平衡,最后利用支持向量机(support vector machines, SVM)进行故障诊断。

## 1 理论方法

### 1.1 SMOTE-IPF 方法

SMOTE 是为解决不平衡样本问题而提出的一种过采样方法,它放弃了随机复制样本的方式,通过在两个少数类样本中间插值的方法合成新的少数样本<sup>[4]</sup>。该方法首先从源样本中依次取出每个样本  $x_i$ , 在每一个  $x_i$  样本周围选择最近的  $k$  个同类样本作为近邻样本集;然后从近邻

基金项目:航空发动机及燃气轮机重大专项基础研究项目(J2019-IV-0018-0086)

第一作者简介:赵亚磊(1994—),男,安徽阜阳人,硕士研究生,研究方向为旋转机械故障诊断,gss95533@163.com。

样本集中随机选取一个作为辅助样本;最后通过在辅助样本和源样本之间进行线性插值的方式合成新样本,插值如式(1)所示。

$$x_{\text{new},a} = x_{i,a} + (x_{j,a} - x_{i,a}) \times \gamma \quad (1)$$

式中: $x_{j,a}$ 中的 $j$ 代表源样本的第 $j$ 个同类样本, $a$ 代表第 $a$ 个属性值; $\gamma$ 是 $[0, 1]$ 之间的随机数; $x_{\text{new},a}$ 表示合成的新样本。

传统的SMOTE方法对源样本的依赖性较高,当少数类样本中出现噪声样本时,会造成大量的生成样本处于错误的空间分布,同时由边界样本生成的样本也会模糊类别间的界限。而SMOTE-IPF在原有方法上进行了改进。它在利用SMOTE生成少数类样本数据的同时,使用多个决策树作为判别器对生成的样本进行判断,然后根据投票结果对合成的噪声样本进行过滤删除,这种方法避免了由噪声样本和边界样本对诊断结果的影响,SMOTE-IPF方法如图1所示。

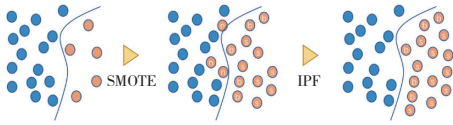


图1 SMOTE-IPF示意图

## 1.2 堆叠去噪自动编码器

自动编码器是一种无监督的神经网络,典型的自动编码器由对称的3层网络组成,分别是输入层、隐含层和输出层,通过隐含层实现对输入数据的编码和重构<sup>[5]</sup>。在训练过程中,自动编码器通过将样本进行编码和解码的方式来生成原始样本的近似样本,然后通过求解原始样本与近似样本的相似程度构造损失函数,再通过最小化损失函数进行网络训练。在完成训练以后,将隐含层的输出值作为包含所有原始样本信息重构表达。

由于3层网络的表达能力有限,为了提升模型的特征表达能力,SDAE将前一个编码器隐藏层的编码特征作为后一个编码器的输入,通过堆叠多个编码器来加深网络的深度<sup>[6]</sup>。SDAE的网络结构如图2所示。

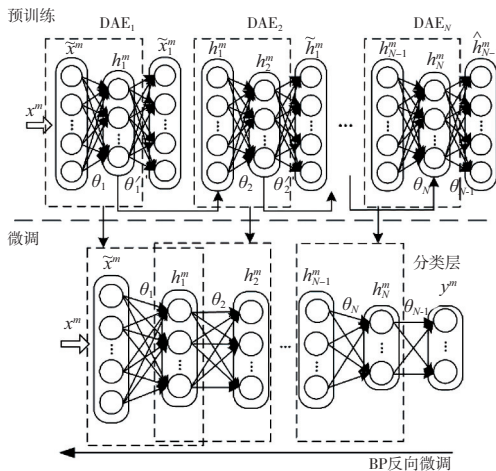


图2 堆叠自动编码器的网络结构

1个SDAE网络包括1个输入层、1个输出层和 $N$ 个隐含层构成。在前向传播过程中,将原始输入样本 $x$ 输入

到 $\text{DAE}_1$ 中,通过加噪、编码、解码的方式进行计算,然后取出隐藏编码 $h_1$ 作为第2个 $\text{DAE}_2$ 的网络输入,重复以上步骤进行 $N$ 次堆叠训练,最后1个编码器的编码输出为

$$h_N^m = f_{\theta,N}(W_N h_{N-1}^m + b_N) \quad (2)$$

在完成多层自动编码器的重构训练后,通过层层加深提取每个编码器的深层特征,最终将最后1个编码器的隐藏层编码作为SDAE网络提取到的特征。

## 2 基于SMOTE-IPF和SDAE的故障诊断方法

针对实际工程中行星齿轮箱故障样本少,故障标签不易获取的情况,本章提出了基于SMOTE-IPF和SDAE的不平衡样本故障诊断方法,该方法主要分为故障特征提取、样本过采样和故障诊断3个阶段,实验流程如图3所示。

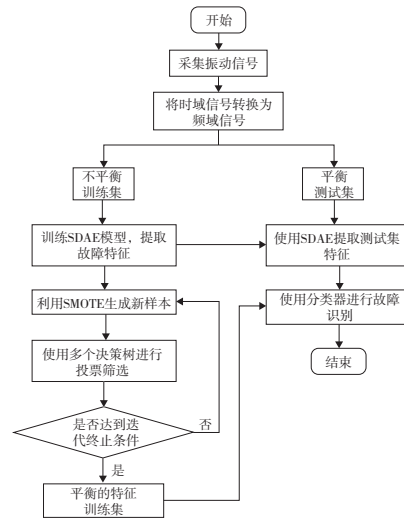


图3 基于SMOTE-IPF和SDAE的诊断流程图

## 3 故障诊断实验

### 3.1 实验数据集

使用行星齿轮模拟实验平台进行试验验证,试验平台如图4所示。试验中共设置了正常模式和5种故障模式,故障模式包括太阳轮点蚀、太阳轮裂纹、太阳轮磨损、行星轮裂纹和行星轮磨损。使用传感器采集转速为1200 r/min,载荷为15 N·m的加速度振动信号,采样频率为40960 Hz,每种模式共采集了256个长度为8192的样本数据。

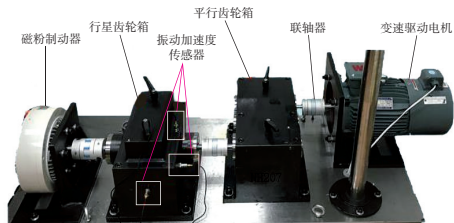


图4 行星齿轮箱故障模拟实验平台

为了验证本文提出方法在不平衡数据集上的故障诊断效果,将振动信号转换为频域信号,将故障数据构造成

多个不平衡比例的训练集进行实验。训练集中不平衡比例设置分别为 2 : 1, 4 : 1, 8 : 1, 10 : 1, 20 : 1 和 40 : 1, 每个样本集中正常样本的个数都为 200, 故每个数据集的故障样本分别为 100, 50, 25, 20, 10 和 5, 测试集中故障样本和正常样本数量相同, 每个模式下的样本都为 100 个。

### 3.2 生成样本的可视化分析

将训练集中的样本不平衡数据输入到 SDAE 模型中进行自动特征提取。其中, 本文中 SDAE 模型由 3 个去噪编码器组成, 编码器的隐藏层节点数依次为 800、400 和

200。经过 SDAE 特征提取后的特征维数为 200, 然后将特征训练集输入到 SMOTE-IPF 模型中, 利用 SMOTE 方法对故障样本进行过采样, 同时利用决策树对生成的样本进行投票, 删除生成样本中的噪声样本。设置决策树个数为 50, 迭代次数为 3, 迭代终止比例为 1%。

为了验证 SMOTE-IPF 方法生成特征样本的有效性, 使用 t-SNE 对原始样本和生成样本进行特征可视化, 观察生成特征的分布情况, 特征散点图如图 5 所示(选择后 3 个数据集进行展示)(本刊黑白印刷, 相关疑问请咨询作者)。

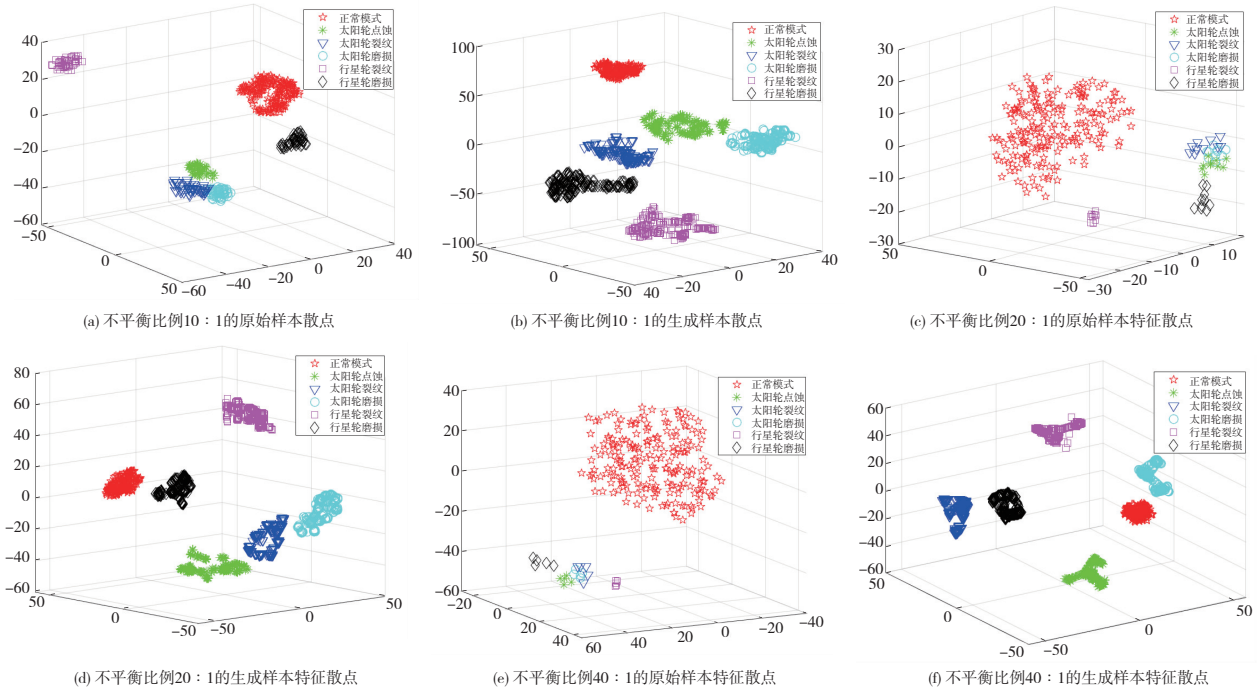


图 5 SMOTE-IPF 合成前后的特征散点图

根据图 5 中生成样本的散点图可知, 随着样本不平衡比例的增大, 原始故障样本之间的界限越来越模糊, 故障模式之间发生重叠现象明显增多。而通过 SMOTE-IPF 方法生成的新样本分布较为集中, 不同类别之间样本的距离分隔较远, 类别间的混叠现象明显减少。

### 3.3 故障诊断结果分析

将 SMOTE-IPF 方法生成的样本输入到 SVM 分类器

中进行模型训练, 然后将测试集样本输入到训练好的模型中进行诊断, 不同样本平衡比例下的故障诊断结果如图 6 所示。为了更全面地反映不平衡样本的故障诊断效果, 使用 F-measure 和 G-mean 两个指标来评价模型的诊断效果。使用每个数据集进行 10 次样本生成试验, 统计生成样本在 SVM 分类器中的诊断结果, 具体结果如图 7 和图 8 所示。

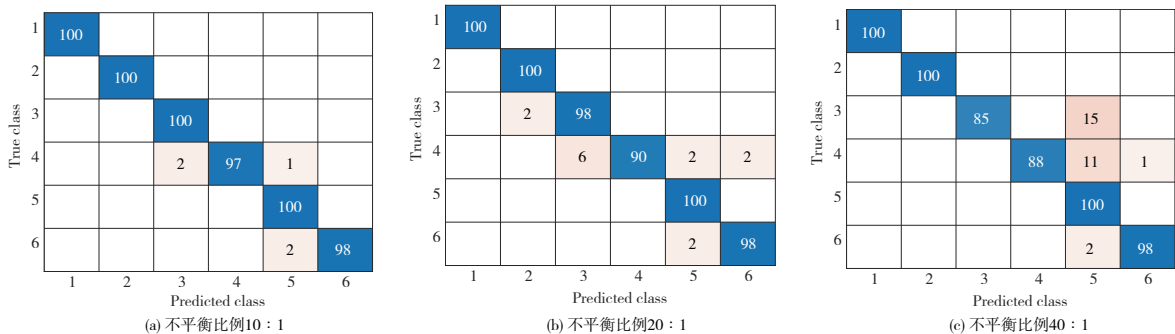


图 6 SMOTE-IPF 生成样本在 SVM 中的分类结果

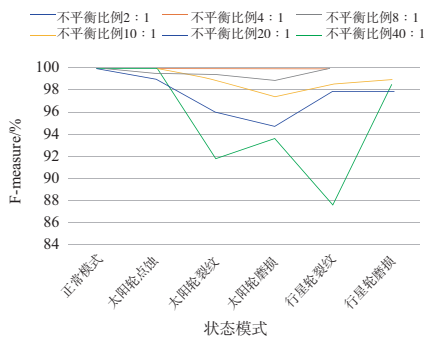


图 7 诊断结果的 F-measure 值

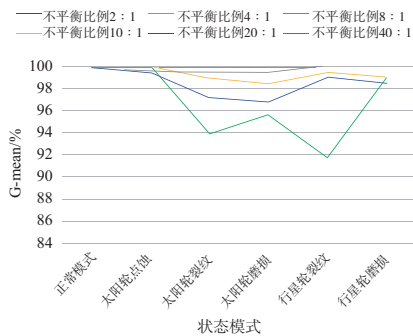


图 8 诊断结果的 G-mean 值

从图 7—图 8 分析可知,当样本不平衡比例小于 10 : 1 时,使用 SMOTE-IPF 方法生成的样本在 SVM 分类器中的分类准确率接近 100%,而当样本不平衡比例增加到 40 : 1 时,分类准确率仍然可以达到 95.17%;另外根据 F-measure 和 G-mean 值可以看出,除了行星轮裂纹的误诊率稍高外,其他模式的生成样本在 SVM 中的漏检率和误诊率都较低。

为了验证本文提出方法的有效性,使用多种常见的样本不平衡故障诊断模型进行对比试验,其中 GAN 的生成器和判别器都使用 3 层卷积神经网络组成。在诊断过程中,每个模型首先在不同样本不平衡比例下的数据进行处理,然后将处理后的样本送入 SVM 分类器进行故障诊断,最终的诊断结果如图 9 所示。

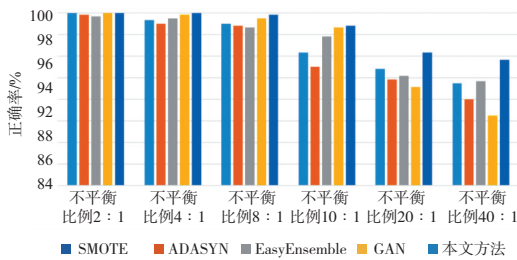


图 9 不同方法的诊断正确率

从图 9 的对比可知,当样本不平衡比例小于 10 : 1 时,5 种方法都能达到较高的诊断正确率,之后随着不平衡比例地增加,模型的诊断正确率开始有较大幅度地下降,而本文所提方法下降幅度最慢,在样本不平衡比例为 40 : 1 时,依然可以达到 95.17% 的诊断正确率,远高于其他 4 种模型。

### 4 结语

本文提出了一个基于 SDAE 和 SMOTE-IPF 的样本不平衡行星齿轮箱故障诊断方法,该方法使用 SDAE 对不平衡训练集进行自动故障特征提取,利用自动编码器不需要标签的特性,避免了多数类在特征提取时的影响,然后使用 SMOTE-IPF 合成少数类样本,使样本达到平衡,最后将生成的样本输入到 SVM 分类器中进行故障分类,通过 t-SNE 的可视化和诊断结果的对比分析验证了所提方法的有效性。

#### 参考文献:

- [1] WANG Z R, WANG J, WANG Y R. An intelligent diagnosis scheme based on generative adversarial learning deep neural networks and its application to planetary gearbox fault pattern recognition[J]. Neurocomputing, 2018, 310:213-222.
- [2] GUO Q W, LI Y B, SONG Y, et al. Intelligent fault diagnosis method based on full 1-D convolutional generative adversarial network[J]. IEEE Transactions on Industrial Informatics, 2020, 16(3):2044-2053.
- [3] RAMENTOL E, GONDRES I, LAJES S, et al. Fuzzy - rough imbalanced learning for the diagnosis of high voltage circuit breaker maintenance: the SMOTE - FRST - 2T algorithm [J]. Engineering Applications of Artificial Intelligence, 2016, 48: 134-139.
- [4] 石洪波,陈雨文,陈鑫. SMOTE 过采样及其改进算法研究综述[J]. 智能系统学报, 2019, 14(6):1073-1083.
- [5] 张鲁晋,陈广艳,孙国栋,等. 基于 SSAE 的输入级弧齿锥齿轮自动特征提取及故障诊断[J]. 机械制造与自动化, 2022, 51(2):161-164.
- [6] VINCENT P, LAROCHELLE H, LAJOIE I, et al. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion[J]. Journal of Machine Learning Research, 2010, 11: 3371-3408.

收稿日期:2022 - 03 - 31