

计算机集成制造系统
Computer Integrated Manufacturing Systems
ISSN 1006-5911, CN 11-5946/TP

《计算机集成制造系统》网络首发论文

题目: 基于主动生成式过采样和 DSN 的轴承故障诊断
作者: 李慧芳, 徐光浩, 黄双喜
收稿日期: 2021-10-09
网络首发日期: 2022-06-22
引用格式: 李慧芳, 徐光浩, 黄双喜. 基于主动生成式过采样和 DSN 的轴承故障诊断 [J/OL]. 计算机集成制造系统.
<https://kns.cnki.net/kcms/detail/11.5946.TP.20220621.1031.004.html>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式 (包括网络呈现版式) 排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊 (光盘版)》电子杂志社有限公司签约, 在《中国学术期刊 (网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊 (网络版)》是国家新闻出版广电总局批准的网络连续型出版物 (ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

基于主动生成式过采样和 DSN 的轴承故障诊断

李慧芳¹，徐光浩¹，黄双喜²

(1.北京理工大学 自动化学院复杂系统智能控制与决策国家重点实验室，北京 100081；

2.清华大学 自动化系，北京 100084)

摘要：针对数据不平衡分类问题，提出一种基于主动生成式过采样与深度堆叠网络（Deep Stacking Network, DSN）的故障诊断方法。首先，在带有分类器的生成式对抗网络（Auxiliary Classifier Generative Adversarial Network, ACGAN）的训练过程中，将 Wasserstein 距离作为新目标函数，为生成器提供有效梯度，并根据损失值之比自适应地调整迭代过程中生成器与判别器的训练次数，克服训练不协调所导致的模型收敛困难，以提高 ACGAN 的训练稳定性，改善生成样本的质量。其次，采用基于委员会查询（Query By Committee, QBC）的主动学习算法，并设计多样性评价指标 Diversity，对 ACGAN 生成的高信息熵样本进行二次筛选，以保证所挑选样本的多样性；同时利用筛选出的样本训练判别器，引导生成器生成信息量丰富的少数类样本。最后，在平衡数据集的基础上，训练基于 DSN 的故障分类模型。对比实验验证了所提出方法的有效性。

关键词：故障诊断；不平衡数据；生成式对抗网络；深度学习

中图分类号：TP206+.3

文献标识码：A

Active generative oversampling and deep stacking network-based bearing fault diagnosis approach

LI Huifang¹, XU Guanghao¹, HUANG Shuangxi²

(1. Key Laboratory of Intelligent Control and Decision of Complex Systems, School of Automation, Beijing Institute of Technology, Beijing 100081, China; 2. Department of Automation, Tsinghua University, Beijing 100084, China)

Abstract: Aiming at the problem of data imbalance classification, an active generative over-sampling and Deep Stacking Network (DSN) based fault diagnosis method is proposed. Firstly, in the training process of Auxiliary Classifier Generative Adversarial Network (ACGAN), the Wasserstein distance is taken as a new objective function to provide an effective gradient for the generator, and the training counts for the generator and discriminator are adaptively adjust in each iteration, to overcome the difficulty of convergence caused by uncoordinated training, and thus improve the training stability of ACGAN and the quality of generated samples. Secondly, a Query By Committee (QBC)-based active learning algorithm is used and a Diversity evaluation index is designed to filter the samples that is produced from the AGANN generator and also with high information entropy so as to ensure the diversity of selected samples. At the same time, these

收稿日期：2021-10-09；**修订日期：**2022-03-07。**Received** 09 Oct. 2021; **accepted** 07 Mar. 2022.

基金项目：国家自然科学基金资助项目(61836001)；国家重点研发计划资助项目(2018YFB1003700)。

Foundation items: Project supported by the National Natural Science Foundation, China (No.61836001), and the National Key Research and Development Program, China (No.2018YFB1003700).

filtered samples are utilized to train a discriminator to guide the generator produce the minority samples with rich information. Finally, a DSN-based fault classifier is trained based on the balanced dataset. A set of comparative experiments are conducted to verify the effectiveness of the proposed method.

Keywords: fault diagnosis; class imbalance; generative adversarial network; deep learning

1 引言

工业设备的智能化运维与健康管理对于智能制造系统的无故障运行非常重要^[1]。这就要求对关键设备进行持续监控,以提高其安全可靠、降低维护成本。随着工业互联网的发展和广泛应用,海量多源异构的设备运行状态数据也变得容易收集,并逐渐涌现出许多数据驱动的故障诊断模型^[2]。这些模型能够利用自身的深层结构自主学习并挖掘故障特征,建立从运行数据到故障类型的非线性映射,实现端到端的准确故障识别。可以说,深度学习方法推动了故障诊断技术的发展,并逐步成为最有潜力的主流方法^[3,4]。现有的深度学习算法大多假设各类故障的训练样本数目几近相同,即数据的类别基本平衡。可是,实际收集到的故障数据往往存在类别不平衡的情况,对容易发生的故障,可收集的数据较多;对发生频率很低的故障,可收集的数据也相对较少。通常,将样本数目较多的故障称为多数类故障,而数目较少的称为少数类故障。在面向不平衡数据时,标准深度学习模型容易将不常发生的少数类故障识别为多数类故障,导致少数类故障的检测率较低甚至无法检测,影响模型的实用性^[5-7]。一旦故障发生却不能及时、准确地诊断,或者潜在的故障即将发生却不能正确地预测或预警,“轻”则影响产品质量、延误生产,“重”则可能酿成事故、威胁人身安全。

关于不平衡数据驱动的故障分类问题,已经出现了大量研究。从数据层面看,典型的是重采样方法,主要针对数据稀少问题,通过改变训练数据的分布减少少数类样本的不平衡比例,以尽量使数据适合于标准机器学习算法。按采样对象不同,现有的重采样方法可分为欠采样和过采样方法。

欠采样方法,通过减少多数类样本的数量和平衡数据集来提高少数类的分类精度。最简单的欠采样方法是随机欠采样(Random Under Sampling, RUS),通过随机删除部分多数类样本来平衡样本数量,例如方昊等^[8]采用多次随机欠采样代替单次随机欠采样来解决软件故障检测中的数据不平衡问题。随机欠采样方法操作简单,但是随机删除可能丢失潜在的样本信息,从而导致分类器性能下降。为此,出现了一些启发式欠采样方法,例如 MANI 根据 K 最邻近法(K-nearest neighbors, KNN)测量的多数类和少数类之间的距离移除样本^[9]。与此类似,BARANDELA 等人基于 KNN 规则,从类边界中删除容易被错误分类的多数类样本^[10]。这些方法能够在一定程度上解决随机欠采样容易丢失重要样本信息的问题,但其本质是舍弃多数类样本,很有可能破坏样本集的分布,而且当少数类样本过少时,通过大量删除多数类样本来平衡数据集也不现实。

过采样方法则将重点放在少数类样本上,通过增加少数类样本的数量来降低数据间的不平衡程度。与欠采样相比,过采样无需担心多数类的信息丢失,模型依然可以获得完整的多数类故障特征,

在一定程度上保证了多数类故障的识别率。最容易实现的过采样方法是随机过采样(Random Over Sampling, ROS),通过随机复制或简单旋转来增加少数类样本的数量,易于实现,但是“复制”会增加过拟合的风险。为此,CHAWLA 等人提出了一种合成少数类过采样技术(Synthetic Minority Oversampling Technique, SMOTE),以解决因随机插入合成样本而导致的过拟合问题^[11]。作为一种经典的过采样方法,SMOTE 已经应用于不同领域^[12],但它可能引入噪声和导致过度泛化。因此,出现了一系列 SMOTE 的变体。ZHU 等人在选择最近邻样本时,通过引入不同样本的权值,有效地处理过度泛化问题^[13]。Boderline-SMOTE 和 Safe-level-SMOTE^[14, 15]是在考虑了领域多数类样本的基础上,对原始 SMOTE 进行改进。Boderline-SMOTE 将过采样操作限制在类边界附近,而 Safe-level-SMOTE 则定义了安全区和噪声区,以防止增加新的样本重叠。与此类似,韩志艳等人提出了一种新的过采样方法,加权合成少数类过采样技术(Weighted Synthetic Minority Oversampling Technique,WSMOTE),将样本划分为几个重要性不同的组,再基于重要性进行样本合成^[16]。

尽管上述方法在不平衡数据分类方面取得了一定的效果,但在实际应用时仍然面临挑战,主要在于其很难自动学习原始数据样本的分布,容易引入噪声,且对不同的数据集适应性较差。深度学习方法具备自主学习的能力,为这一问题提供了新的契机。例如 ANDO 等人提出了一种深度过采样(Deep Over-sampling, DOS)模型,通过在特征空间对少数类样本进行重采样,提高了少数类的分类性能^[17]。在基于深度学习的生成方法中,2014 年 GOODFELLOW 等人提出的生成式对抗网络(Generative Adversarial Networks, GANs),已被广泛应用于不同领域^[18],也出现了一些 GANs 的变体。GANs 通过学习原始样本的数据分布特征,生成具有相似数据分布的新合成样本^[19]。LEE 等利用原始 GANs 从经验模态分解能谱数据中生成样本,在故障诊断中的表现优于典型的过采样技术^[20]。DOUZAS 等则利用条件生成式对抗网络(Conditional Generative Adversarial Network, CGAN)生成少数类样本,有效地改善了数据的不平衡程度^[21]。WU 等人构建了半监督信息最大化生成对抗网络(semi-supervised information maximizing Generative Adversarial Network, ss-InfoGAN),利用对抗结构生成少数样本,通过增加数据来解决数据不平衡问题^[22]。此外,还有研究将 GANs 与自动编码器相结合。WANG 等将 GANs 与堆叠降噪自动编码器(Stacked Denoising Autoencoders, SDAE)结合,提出了一种基于深度特征提取的故障分类方法,进行齿轮箱故障诊断^[23]。HAN 等人提出了一种 GAN-SAE 故障诊断方法,首先基于 GAN 对样本数据的不平衡性进行补偿,再利用堆叠自动编码器(Stacked Auto Encoder, SAE)方法提取信号特征^[24]。HAN 等人将对抗学习引入卷积神经网络(Convolutional Neural Network, CNN),提高了特征表示的鲁棒性,增强了其模型的泛化能力^[25]。XIE 等人提出了一种深度卷积 GAN 模型,通过从少数类样本中模拟原始数据分布并生成新样本来改善不平衡性^[26]。包萍等人提出了一种基于 GANs 和 CNN 的故障诊断方法,在平衡数据集之后,引入 Focal loss 损失函数来提高故障识别的准确性^[27]。

现有面向类别不平衡的故障识别方法中,大多数试图从数据层面解决类别不平衡带来的少数类

故障检测不出或检测率太低的问题，通过基于过采样的数据增强方法，在一定程度上减少或者弥补这种不平衡性的影响，再使用标准深度学习模型进行故障分类。作为目前流行的数据增强技术，基于 GAN 的生成式过采样方法获得了越来越多的关注。但是现有的方法在生成样本中存在以下问题：

（1）由于 GAN 固有的模式坍塌与模型训练稳定性问题，导致生成样本的质量较差，甚至可能引入噪声样本；（2）在模型训练过程中，缺乏对生成样本个体差异性的关注，使生成器生成的样本多样性不足，影响模型的泛化能力，最终影响模型的少数类故障识别率。

为此，本文考虑数据不平衡对诊断准确率的影响，提出了一种基于主动生成式过采样与深度堆叠网络的故障分类方法，采用 ACGAN 和主动学习算法，学习原始样本分布，在保证生成样本与原始样本分布相似的前提下，生成多样性好的高质量样本。在对数据集进行平衡化处理后，设计基于 DSN 的故障分类器，提高少数类识别率。主要贡献如下：

1) 为了改善 ACGAN 训练的稳定性，并解决因原目标函数 JS 散度的突变性所造成的生成器梯度难以获取问题，首先，以连续变化的 Wasserstein 距离为目标函数，搭建 W_ACGAN 模型，使生成器能够更好地获取有效梯度，在一定程度上缓解 ACGAN 的训练困难。其次，设计基于损失值之比的模型参数更新策略，自适应地调整生成器与判别器训练的每轮迭代次数，即生成器和判别器的训练步伐，克服因任何一方训练效果太差或太好，即生成器和判别器训练不协调而引起的模型收敛困难，确保 W_ACGAN 训练的稳定性，提升生成少数类样本的质量。

2) 为了引导 W_ACGAN 生成丰富多样的少数类样本，首先，将主动学习算法 QBC 嵌入到 W_ACGAN 训练过程中，从生成样本中挑选一批信息熵较高的样本，并采用多样性评价指标 Diversity 对所选样本进行二次挑选，以去除冗余、保证所挑选样本的多样性。其次，将所选样本当作真实样本，连同剩余样本一起训练判别器。通过判别器与生成器的对抗式学习，在保证生成样本与真实样本相似的基础上，引导生成器生成类似于所选样本并具有丰富信息的少数类样本，为少数类识别提供更多的故障知识。

3) 为了减少深度网络训练参数微调阶段因随机梯度下降可能带来的参数优化问题，在对故障数据进行平衡化处理的基础上，设计基于 DSN 的故障分类模型，即将依次训练好的多个基本功能组件或分类器相互堆叠，构建复杂的故障分类模型。通过跨机器并行训练与特征学习训练故障分类器，在确保模型参数优化性能的同时，大大提高故障诊断模型的训练效率。

2 相关理论基础

2.1 主动学习算法

主动学习算法旨在提高模型性能的同时，尽可能减少样本标记的成本。作为典型的数据驱动方法，深度学习需要学习大量的参数，且参数优化离不开大量的注释样本。早期的主动学习算法未得到足够重视，主要因为传统机器学习方法对标记样本的要求相对较少。随着工业互联网的深入应用，海量数据的收集变得相对容易。然而，大多深度学习的研究依赖于大量公开可用的注释数据集，且

高质量的标注数据集往往需要耗费大量的人力。特别地，当数据标注需要高水平的领域专家知识时，数据集的获取就变得异常艰难，而主动学习算法开始逐渐发挥自身的价值^[28]。

主动学习的基本思想是：从没有标记的数据集中选择最有用的样本，并经督导者进行标注后将其加入到标记样本集，用以训练机器学习模型，从而在保持模型性能的同时尽可能地降低样本标记成本。通过精心设计查询规则，主动学习方法从无标记的数据集中选择最好的样本，向督导者查询其标签，试图最大限度地降低标记成本。图 1 给出了一个基本的主动学习模型。查询规则的设计对主动学习方法的性能至关重要，因此出现了不同的查询策略。例如，在一个给定未标记数据集上的主要查询策略包括基于不确定性的方法和基于多样性的方法，以及同时考虑查询样本不确定性和多样性的混合查询策略，试图在这两者之间寻找平衡。基于不确定性的单独采样往往会带来采样偏差，选取的样本未必代表未标记数据集的分布。另一方面，只追求抽样多样性的策略会增加标注的成本，因为可能选择相当数量信息含量低的样本。主动学习已有大量研究，但仍存在高维数据扩展的问题。因此，大多数人工智能研究更倾向于低维问题。此外，主动学习算法经常根据预先提取的特征查询高值样本，不具备提取特征的能力。

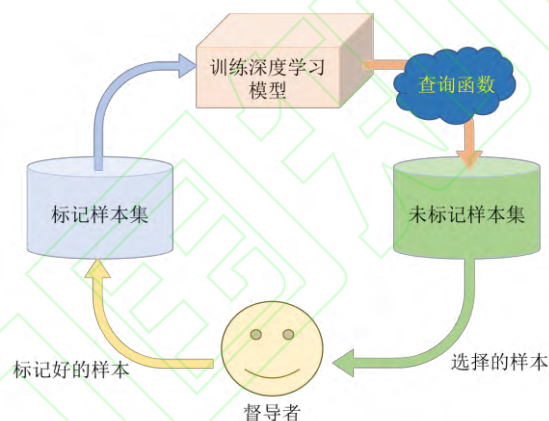


图 1 主动学习模型

2.2 ACGAN

带有辅助分类器的生成式对抗网络^[29]（Auxiliary Classifier Generative Adversarial Network, ACGAN）起源于条件生成对抗网络（Conditional GAN, CGAN）。CGAN 通过在生成器的输入中添加样本标签信息来控制生成样本的类别。ACGAN 是对 CGAN 的拓展，除了在输入中添加标签信息外，ACGAN 采用一个分类器以辅助判别器。因此，ACGAN 不仅可以判断一个样本来自于真实分布还是生成分布，还可以判断该生成样本的类别。也就是说，在 ACGAN 的判别器中增加了一个分类功能，其结构如图 2 所示。同时，ACGAN 的实验结果似乎说明了通过在 GAN 的潜在空间中添加更多的结构和/或专门的损失函数，就可以产生更高质量的样本。

与原始 GAN 生成的样本相比，ACGAN 的每一个生成样本 x_{fake}^i 均有一个相应的类别标签，可以表示为 $x_{fake}^i = G_{acgan}(label_i, z_i)$ 。其中， $label_i$ 表示第 i 个样本对应的标签， z_i 表示第 i 个噪声输入。由

于 ACGAN 的分类功能，其损失函数也包括判别损失 L_d 和分类损失 L_c 两个部分，且 L_d 和 GAN 的判别器损失函数相同，如式(1)所示：

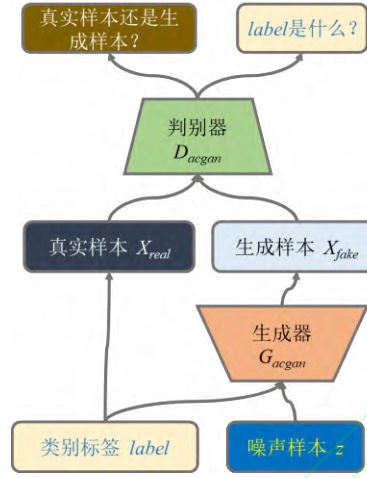


图2 ACGAN 结构

$$L_d = E \left[\log P(\text{预测标签} = \text{真实} | X_{real}) \right] + E \left[\log P(\text{预测标签} = \text{生成} | X_{fake}) \right] \quad (1)$$

其中， $P(\text{预测标签} = \text{真实} | X_{real})$ 表示“输入为真实样本、判别器给出的预测标签也为真实样本”的概率， $P(\text{预测标签} = \text{生成} | X_{fake})$ 表示“输入为生成样本、判别器给出的预测标签也为生成样本”的概率。分类损失 L_c 是 ACGAN 的特点，其计算如式(2)所示：

$$L_c = E \left[\log P(\text{分类标签} = \text{真实标签} | X_{real}) \right] + E \left[\log P(\text{分类标签} = \text{真实标签} | X_{fake}) \right] \quad (2)$$

其中， $P(\text{分类标签} = \text{真实标签} | X_{real})$ 、 $P(\text{分类标签} = \text{真实标签} | X_{fake})$ 分别表示“输入为真实样本和生成样本，而判别器给出的分类标签和真实的类别标签一致”的概率。

在 ACGAN 中，判别器通过模型训练使 $L_c + L_d$ 最大化，即既要能够判断样本是属于真实样本还是生成样本，又要能够判断生成样本以及真实样本的类别。生成器则通过模型训练使 $L_c - L_d$ 最大化，也就是说，除了使生成样本能够“愚弄”判别器，生成器还要使各个生成样本的类别与其相应的原始类别样本接近，使判别器在真、假判别中判断错误，但在识别样本类别时分类正确。从结构上看，ACGAN 与标准 GAN 无太大差别，但这种对标准 GAN 的修改，似乎能够在一定程度上使 GAN 的训练更加稳定，从而产生更高质量的样本。

2.3 DSN

深度堆叠网络(Deep stacking network, DSN)是在深度神经网络(Deep Neural Network, DNN)的基础上发展起来的一种新型网络结构，具有良好的可扩展性^[30]。其结构类似于深度置信网络(Deep

Belief Network, DBN)，但有以下几点不同：(1) DSN 的基本模块是浅层的全连接神经网络，DBN 的组件为受限玻尔兹曼机（Restricted Boltzmann Machines, RBM）；(2) DSN 后续模块的输入同时包含前一模块的输出和输入；(3) DSN 的每一个模块都采用有监督的训练。图 3 是一个 DSN 网络结构，包含四个模块，且每个模块由三层网络组成。

类似于 DBN，深度堆叠网络的训练过程同样可分为两个步骤：模块单独训练和模块约束训练。在模块单独训练过程中，假设 $X = \{x_1, \dots, x_i, \dots, x_N\}$ 表示一组训练样本， x_i 表示第 i 个训练样本， N 为训练样本的总个数。那么，对于 x_i ，一个 DSN 基本模块的输出 $y_i = (W_{dsn}^U)^T \times h_i$ ，其中 $h_i = \Gamma \left[(W_{dsn}^L)^T \times x_i \right]$ 为隐含层输出， W_{dsn}^L 为输入层和隐含层之间的权值矩阵， W_{dsn}^U 为隐含层与输出层之间的权值矩阵。同理，可得所有样本的隐含层输出 $H = \{h_1, \dots, h_i, \dots, h_N\}$ 和最终输出 $Y = \{y_1, \dots, y_i, \dots, y_N\}$ 。

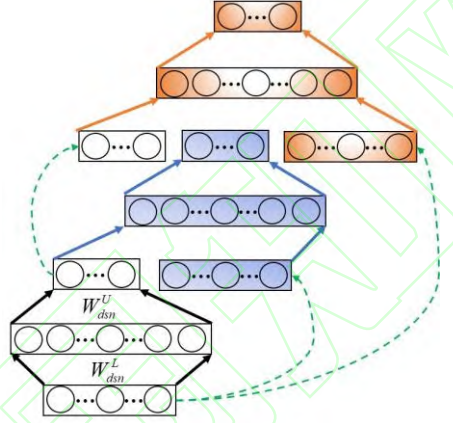


图 3 DSN 结构

给定样本 $X = \{x_1, \dots, x_i, \dots, x_N\}$ 相应的标签集 $\tilde{Y} = \{\tilde{y}_1, \dots, \tilde{y}_i, \dots, \tilde{y}_N\}$ ，其中 \tilde{y}_i 表示第 i 个样本 x_i 的标签，则 DSN 的模块单独训练过程为：通过学习调整参数 W_{dsn}^U 和 W_{dsn}^L ，使 Y 和 \tilde{Y} 总平方误差 E_{dsn} 的平均值最小， E_{dsn} 的计算如式(3)所示：

$$E_{dsn} = \|Y - \tilde{Y}\| = \text{Tr} \left[(Y - \tilde{Y})(Y - \tilde{Y})^T \right] \quad (3)$$

一旦权重 W_{dsn}^L 的值固定（例如通过随机数确定 W_{dsn}^L ），则隐含层的输出 H 也将唯一确定。因此，可通过设置梯度 $\frac{\partial E_{dsn}}{\partial W_{dsn}^U}$ 为 0 来确定权重 W_{dsn}^U 。 $\frac{\partial E_{dsn}}{\partial W_{dsn}^U}$ 的计算如式(4)所示：

$$\begin{aligned} \frac{\partial E_{dsn}}{\partial W_{dsn}^U} &= \frac{\partial \text{Tr} \left\{ \left[(W_{dsn}^U)^T \times H - \tilde{Y} \right] \times \left[(W_{dsn}^U)^T \times H - \tilde{Y} \right]^T \right\}}{\partial W_{dsn}^U} \\ &= 2H \left[(W_{dsn}^U)^T \times H - \tilde{Y} \right]^T \end{aligned} \quad (4)$$

令 $\frac{\partial E_{dsn}}{\partial W_{dsn}^U} = 0$ ，可得 $W_{dsn}^U = (HH^T)^{-1} H \tilde{Y}^T$ 。

在模块约束训练(Module-bound fine tuning)阶段, 可使用批量模式梯度下降法, 进一步学习每个模块的权重 W_{dsn}^L , 具体过程如式(5)所示:

$$\begin{aligned}\frac{\partial E_{dsn}}{\partial W_{dsn}^L} &= -\frac{\partial \text{Tr}\left\{\left[\left(W_{dsn}^U\right)^T \times H - \tilde{Y}\right] \times \left[\left(W_{dsn}^U\right)^T \times H - \tilde{Y}\right]^T\right\}}{\partial W_{dsn}^L} \\ &= -\frac{\partial \text{Tr}\left(HH^T\right)^{-1} H \tilde{Y}^T \tilde{Y} H^T}{\partial W_{dsn}^L}\end{aligned}\quad (5)$$

3 基于主动生成式过采样和深度堆叠网络的故障诊断

针对标准 GAN 固有的训练困难、模式坍塌等所导致的生成样本质量以及多样性较差问题, 本文提出了一种基于 QBC 和 ACGAN 的主动生成式过采样方法。通过更加平滑的 Wasserstein 距离代替原有的 JS 散度, 衡量真假样本分布之间的距离, 并构建 ACGAN 模型, 即 W_ACGAN, 以提高其训练的稳定性。其次, 采用 QBC 以及多样性评价指标 (Diversity), 从 ACGAN 生成的样本中挑选最具代表性且多样性好的样本, 以引导 ACGAN 生成更多样且利于后续故障分类的少数类样本。同时, 提出了一种基于损失值的自适应模型训练方法, 通过调整生成器与判别器的训练周期, 提升生成器与判别器之间的对抗效果, 改善所生成样本的质量。最后, 构建基于 DSN 的故障分类器。基于主动生成式过采样的故障诊断方法流程如图 4 所示。

3.1 基于 QBC 和 W_ACGAN 的主动生成式过采样方法

现有基于 GAN 的数据采样方法, 大多直接使用原始 GAN 拟合少数类样本分布, 并采用训练好的生成器生成样本对少数类样本进行扩充。然而, GAN 固有的模式坍塌问题使训练过程倾向于生成较为安全的样本, 以期更好地欺骗判别器, 这样生成器所生成的样本大多为相似的安全样本, 即多样性较差。如果使用多样性较差的样本对原始少数类样本进行补充, 易于引起样本的类内分布不平衡, 那么模型将无法学习到完整的少数类样本分布, 也使部分少数类样本更难识别。此外, GAN 固有的训练局限性, 需要小心平衡生成器和判别器的训练程度。如果判别器训练得太好, 生成器无法愚弄判别器, 则判别器就无法从对抗训练中获得进步; 同时, 判别器越好, 生成器的梯度消失越严重, 生成样本的质量也就较差。

为了使 GAN 能够生成多样化的高质量少数类故障样本, 本文设计了一种基于 QBC 和 W_ACGAN 的主动生成式过采样方法。首先, 通过修改 ACGAN 的模型结构, 引入基于 Wasserstein 距离的生成式对抗网络^[31] (Wasserstein Generative Adversarial Network, WGAN) 的思想, 并使用更加平滑的 Wasserstein 距离代替原有的 Jensen-Shannon (JS) 散度, 来衡量真假样本分布的距离。其次, 使用 QBC 算法从生成器所生成的样本中挑选利于提升分类效果的样本, 并利用所设计的多样性评价指标对所选样本进行二次筛选, 以过滤部分相对冗余的样本。通过将所挑选的样本标签设为与真实样本一致的标签来训练判别器, 以引导生成器生成多样化的、有助于改善分类性能的样本。最

后，基于不同迭代批次的损失值比值，计算生成器与判别器的训练周期，以自适应地调整生成器与判别器的训练程度，达到平衡训练的目的。

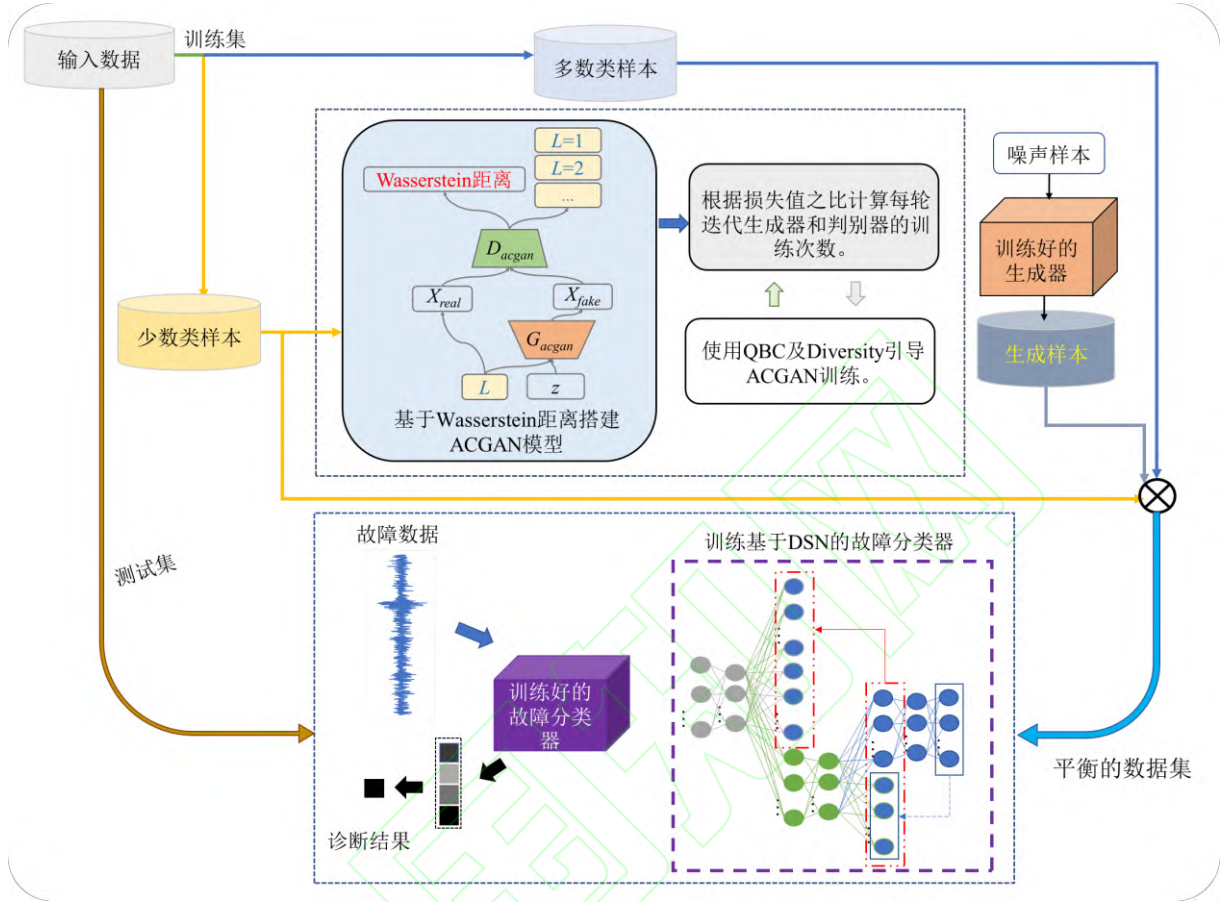


图4 本文方法的总体流程

3.1.1 W_ACGAN 模型构建

为了解决 ACGAN 的训练困难问题，本文在 ACGAN 中引入 WGAN 的 Wasserstein 距离，设计 W_ACGAN，减少因 JS 散度带来的梯度消失，提高模型训练的稳定性。

原始 ACGAN 判别器着重于将真实样本分为正例、生成样本分为负例，其损失函数为：

$$-E_{x \sim P_r} [\log D(x)] - E_{x \sim P_g} [\log (1 - D(x))] \quad (6)$$

其中， P_r 表示真实样本的分布， P_g 表示生成样本的分布， $D(x)$ 表示判别器的表达式。当固定生成器的参数，并训练判别器时，任意样本 x （真实样本或者生成样本）对判别器损失函数的贡献可以表达为式(7)：

$$-P_r(x) \log D(x) - P_g(x) \log [1 - D(x)] \quad (7)$$

为了获得最优的判别器表达式，令式(7)中 $D(x)$ 的导数为 0，可得式(8)：

$$-\frac{P_r(x)}{D(x)} + \frac{P_g(x)}{1 - D(x)} = 0 \quad (8)$$

其中, $P_r(x)$ 表示样本 x 来自真实分布的概率, $P_g(x)$ 表示样本 x 来自生成分布的概率。将式(8)化简, 可得最优判别器 $\tilde{D}(x)$ 的表达式, 如式(9)所示:

$$\tilde{D}(x) = \frac{P_r(x)}{P_r(x) + P_g(x)} \quad (9)$$

也就是说, 当 $P_r(x)=0$ 且 $P_g(x) \neq 0$ 时, 最优判别器给出样本 x 来自真实分布的可能性为 0。当 $P_r(x)=P_g(x)$ 时, 样本 x 来自真实分布和生成分布的可能性相等, 即最优判别器给出样本 x 来自真实分布的概率为 0.5。

考虑一种极端情况, 即当判别器训练为最优判别器时, 生成器的损失函数为:

$$\begin{aligned} & E_{x \sim P_r} [\log \tilde{D}(x)] + E_{x \sim P_g} [\log (1 - \tilde{D}(x))] \\ &= E_{x \sim P_r} \log \frac{P_r(x)}{\frac{1}{2} \times (P_r(x) + P_g(x))} + E_{x \sim P_g} \log \frac{P_g(x)}{\frac{1}{2} \times (P_r(x) + P_g(x))} - 2 \log 2 \\ &= 2 \times JS(P_r \parallel P_g) - 2 \log 2 \end{aligned} \quad (10)$$

其中, $JS(P_r \parallel P_g)$ 为真实分布 P_r 与生成分布 P_g 之间的 JS 散度 (Jensen-Shannon divergence), 其计算如下:

$$\begin{aligned} JS(P_r \parallel P_g) &= \frac{1}{2} E_{x \sim P_r} \log \frac{P_r(x)}{\frac{1}{2} \times (P_r(x) + P_g(x))} \\ &\quad + \frac{1}{2} E_{x \sim P_g} \log \frac{P_g(x)}{\frac{1}{2} \times (P_r(x) + P_g(x))} \end{aligned} \quad (11)$$

由以上推导可得: 在判别器为最优时, 原始 GAN 定义的生成器损失值转化为最小化真实样本分布和生成样本分布之间的 JS 散度。但是, 当真实分布 P_r 与生成分布 P_g 之间没有不可忽略的重叠部分时, 其 JS 散度恒为常数 $\log 2$ 。此时, 由式 (10) 可知: 生成器损失函数的梯度恒为 0, 即生成器无法再得到优化。事实上, 很大可能是生成分布与真实分布间没有不可忽略的重叠部分, 这正是 GAN 难以训练的原因^[32]。

为此, 本文以 Wasserstein 距离代替原有的 JS 散度, 衡量真实分布 P_r 与生成分布 P_g 间的距离并进行最小化。Wasserstein 距离 $W(P_r, P_g)$ 可按式(12)计算:

$$W(P_r, P_g) = \inf_{\delta \in \Pi(P_r, P_g)} E_{(x,y) \sim \delta} \|x - y\| \quad (12)$$

其中, $\Pi(P_r, P_g)$ 表示真实分布 P_r 与生成分布 P_g 组合成的所有可能联合分布集合, δ 为属于 $\Pi(P_r, P_g)$ 的任一联合分布, x 、 y 分别表示真实样本和生成样本。从式(12)可以看出, 对于 JS 散度, Wasserstein

距离的优势在于：无论真实分布 P_r 与生成分布 P_g 间是否有重叠部分，Wasserstein 距离总是能够反映两个分布之间的距离。也就是说，Wasserstein 距离相对连续、平滑，能产生 JS 散度无法提供的梯度，即引入 Wasserstein 距离可以使 GAN 的训练更加稳定，且 GAN 生成器的损失函数变为 $-E_{x \sim P_g} [D(x)]$ ，判别器的损失函数变为 $E_{x \sim P_g} [D(x)] - E_{x \sim P_r} [D(x)]$ 。

3.1.2 基于 QBC 和 Diversity 的 W_ACGAN 训练引导

针对原始 GAN 生成样本多样性不足问题，本文在 W_ACGAN 训练过程中引入 QBC 算法及 Diversity 评价指标，以期引导 W_ACGAN 生成多样化的、利于提升故障分类效果的样本。QBC 算法旨在从生成样本中，选出委员会分类结果最不一致的样本，即更容易被分错的样本，对这些样本的学习更有必要，且学习它们对提升分类器性能的贡献也会更大。通过将 QBC 嵌入到 W_ACGAN 训练过程中，从 QBC 挑选的样本中选出与其它样本更不一致的样本，提升生成样本的多样性。具体流程如下：首先，采用 QBC 算法从生成样本中选出 M 个熵值较高的样本，再借助 Diversity 评价指标从中筛选出 \hat{M} ($\hat{M} < M$) 个样本。其次，将该 \hat{M} 个样本标签设置为真实样本标签，即作为真实样本加入原始真实样本集。最后，利用得到的真实样本集与剩余生成样本一起训练判别器。

使用 QBC 算法挑选样本时，先以装袋的形式从包含所有样本的训练集中选取 K 个训练集 T_1, T_2, \dots, T_K ，并使用 K 个训练集分别训练 K 个独立的分类模型 C_1, C_2, \dots, C_K ，以构成一组委员会 $C = \{C_1, C_2, \dots, C_K\}$ 。其次，将生成器对 r 个噪声样本 $z_1, z_2, \dots, z_j, \dots, z_r$ 所生成的样本 $G(z_1), G(z_2), \dots, G(z_j), \dots, G(z_r)$ 分别输入到 K 个分类模型，每个样本将获得 K 个分类模型给出的 K 个预测标签。最后，按照式 (13) 并利用 K 个预测标签分别计算 r 个样本的熵值 $H[G(z_1)], \dots, H[G(z_j)], \dots, H[G(z_r)]$ ：

$$H[G(z_j)] = \frac{\sum_{\varsigma=1}^U P_j^{\varsigma} \log[P_j^{\varsigma}]}{\log U} \quad (13)$$

其中， $P_j^{\varsigma} = p(\hat{y}_j = \varsigma | G(z_j))$ 表示第 j 个生成样本被 K 个分类模型预测为第 ς 类少数类的概率，即

$p(\hat{y}_j = \varsigma | G(z_j)) = \frac{K_{\varsigma}}{K}$ ， K_{ς} 代表 $G(z_j)$ 的 K 个预测标签中属于第 ς 类标签的个数， U 为少数类的类别个数。当委员会对样本 $G(z_j)$ 的预测标签相同时，其熵值 $H[G(z_j)] = 0$ ，即对于委员会的每个分类器来说，样本 $G(z_j)$ 都属于容易正确分类的样本。那么，该样本的加入对提升分类器性能的贡献

相对较少。反之，样本的熵值越大，表明其越容易被分错，也就意味着该样本提供的信息量越大。因此，本文从 r 个生成样本中选出 M 个熵值较高的样本，即 $\{G^1, G^2, \dots, G^k, \dots, G^M\}$ ，且

$G^k \in \{G(z_1), G(z_2), \dots, G(z_j), \dots, G(z_r)\}$ ，以提高分类器的性能。其中， G^k 为被选中的第 k 个样本。

尽管 QBC 算法已选出 M 个信息量较高的样本，可是为确保生成的样本能够均匀分布，避免因单一引导带来的生成样本多样性不够好的问题，本文设计一个评价指标 Diversity，对 M 个样本 $\{G^1, G^2, \dots, G^k, \dots, G^M\}$ 进行二次筛选。具体步骤如下：

首先，从 $\{G^1, G^2, \dots, G^k, \dots, G^M\}$ 中任意挑选一个样本 G^k ，根据公式(14)依次计算 G^k 与其它 $M-1$ 个样本之间的欧氏距离 $Ed_{k,1}, Ed_{k,2}, \dots, Ed_{k,k-1}, Ed_{k,k+1}, \dots, Ed_{k,M}$ ：

$$Ed_{k,u} = \sqrt{|G^k - G^u|^2} \quad (14)$$

其中， $Ed_{k,u}$ 表示 G^k 和第 u 个样本 G^u 之间的欧氏距离。

其次，将求得的所有欧氏距离，即 $Ed_{k,1}, Ed_{k,2}, \dots, Ed_{k,k-1}, Ed_{k,k+1}, \dots, Ed_{k,M}$ 求和，得到样本 G^k 的 Diversity 值 D_k ，即：

$$D_k = \sum_{s=1}^{k-1} Ed_{k,s} + \sum_{s=k+1}^M Ed_{k,s} \quad (15)$$

同理，依次计算其余 $M-1$ 个样本的 Diversity 值 $D_1, D_2, \dots, D_{k-1}, D_{k+1}, \dots, D_M$ 。

最后，从 $\{G^1, G^2, \dots, G^k, \dots, G^M\}$ 中挑选出 \hat{M} 个 Diversity 值较大的样本 $\hat{G}^1, \hat{G}^2, \dots, \hat{G}^{k'}, \dots, \hat{G}^{\hat{M}}$ ，作为最终选择的样本。其中， $\hat{G}^{k'} \in \{G^1, G^2, \dots, G^k, \dots, G^M\}$ 。同时，将这些挑选得到的样本的标签设置为真实标签，即 *True*，剩余未被挑选的样本标签仍然是 *False*，并将所挑选样本当作真实样本参与判别器训练，以引导生成的样本既朝真实样本的方向靠近，也朝被挑选样本的方向逼近。

为了避免所挑选的生成样本与原始样本差别过大，并导致生成器朝着偏离真实样本的方向发展，本文在被挑选样本的损失中引入衰减因子项 $\sigma (0 < \sigma < 1)$ ，以调整被挑选样本在判别器损失函数中的贡献。该衰减因子的取值将随着迭代次数的增加而增大，这是因为训练开始时，生成样本与原始样本差别较大，应以真实样本为主导来引导训练；随着迭代次数的增加，生成样本与原始样本开始逐渐接近，我们希望生成样本朝着被挑选样本的方向逼近。因此，最终判别器的损失函数应包含两个部分：一是未被挑选样本与真实样本的损失即原有损失，二是添加了衰减因子项的被挑选样本的损失，如式(16)所示：

$$d_{loss} = \frac{1}{L} \sum_{l=1}^L D(x_l) - \frac{1}{Q} \sum_{j=1}^Q D(G(z_j)) + \frac{\sigma}{\hat{M}} \sum_{k'=1}^{\hat{M}} D(\hat{G}^{k'}) \quad (16)$$

其中， x_l 表示第 l 个真实样本， L 为真实少数类样本的总个数， Q 为未被选择的生成样本总个数。

3.1.3 基于损失值的自适应模型参数更新

为了使 W_ACGAN 的训练更加稳定, 本文从控制判别器与生成器的训练步伐出发, 在优化生成器时, 假定判别器的判别能力优于当前生成器的生成能力, 以使判别器指导生成器朝着更好的方向学习。具体地: 先将判别器的参数更新一次或者多次, 再更新生成器的参数。与 WGAN 采用固定模式, 即“先更新判别器 5 次、再更新生成器 1 次”不同, 本文提出了一种自适应训练方法。通过计算上次迭代损失值与本次迭代损失值的比值, 得到下轮迭代判别器与生成器的参数更新次数。假设上轮迭代判别器的损失值为 d_{Loss}^{pre} , 本轮迭代判别器的损失值为 d_{Loss}^{curr} , 上轮迭代生成器的损失值为 g_{Loss}^{pre} , 本轮迭代生成器的损失值为 g_{Loss}^{curr} , 则本文提出的自适应训练方法步骤如下:

步骤一: 训练判别器与生成器各一次, 并将第一轮迭代中判别器和生成器的损失值分别赋给 d_{Loss}^{pre} 和 g_{Loss}^{pre} 。然后, 再训练判别器与生成器各一次, 并将第二轮迭代中判别器和生成器的损失值分别赋给 d_{Loss}^{curr} 和 g_{Loss}^{curr} 。

步骤二: 采用公式(17)计算下一轮迭代中判别器的参数更新次数 d_{ts} 。为了避免因上轮迭代损失值 d_{Loss}^{pre} 可能为 0 而导致更新次数无穷大的情况, 在分母中添加了一项极小的浮点数 ε :

$$d_{ts} = \begin{cases} d_{ts}^m, & \text{如果 } \frac{d_{Loss}^{curr}}{d_{Loss}^{pre} + \varepsilon} \leq d_{ts}^m \\ \text{ceil} \frac{d_{Loss}^{curr}}{d_{Loss}^{pre} + \varepsilon}, & d_{ts}^m < \frac{d_{Loss}^{curr}}{d_{Loss}^{pre} + \varepsilon} < d_{ts}^M \\ d_{ts}^M, & \frac{d_{Loss}^{curr}}{d_{Loss}^{pre} + \varepsilon} \geq d_{ts}^M \end{cases} \quad (17)$$

其中, d_{ts}^m 和 d_{ts}^M 分别为预先设置的每轮迭代判别器参数更新最小与最大次数, 以避免更新次数过多或过少而导致模型训练效果不佳。可以看出, 判别器的训练是负反馈过程, 即, 本轮判别损失 d_{Loss}^{curr} 越大, d_{ts} 将会越大。相应地, 下一轮迭代中判别器的判别损失将会减小, 从而增强判别器的判别能力。

步骤三: 通过公式(18)计算下一轮迭代中生成器的参数更新次数 g_{ts} , 类似于式(17), 在分母中添加了一项浮点数 ε :

$$g_{ts} = \begin{cases} g_{ts}^m, & \frac{g_{Loss}^{pre}}{g_{Loss}^{curr} + \varepsilon} \leq g_{ts}^m \\ \text{ceil} \frac{g_{Loss}^{pre}}{g_{Loss}^{curr} + \varepsilon}, & g_{ts}^m < \frac{g_{Loss}^{pre}}{g_{Loss}^{curr} + \varepsilon} < g_{ts}^M \\ g_{ts}^M, & \frac{g_{Loss}^{pre}}{g_{Loss}^{curr} + \varepsilon} \geq g_{ts}^M \end{cases} \quad (18)$$

其中, g_{ts}^m 和 g_{ts}^M 分别为预先设置的每轮迭代生成器参数更新最小与最大次数。与判别器相反, 本轮生成损失 g_{Loss}^{curr} 越大, g_{ts} 将会越小, 即生成器的生成损失将会越大, 生成能力也就不会变强。公式(17)

和 (18) 共同作用，能够确保判别器的判别能力总是优于生成器的生成能力，从而引导生成器生成更高质量的样本。4.2 节的实验结果验证了本文自适应参数更新策略的有效性。

步骤四：将 d_{Loss}^{curr} 和 g_{Loss}^{curr} 的值分别赋给 d_{Loss}^{pre} 和 g_{Loss}^{pre} ，按照上述计算所得的判别器和生成器参数更新次数 d_{ts} 、 g_{ts} ，分别训练判别器和生成器。

步骤五：计算判别器和生成器的损失值，并分别赋给 d_{Loss}^{curr} 和 g_{Loss}^{curr} 。

步骤六：重复步骤三至步骤五，直至迭代次数达到预设值 Δ 。

按照上述步骤训练本文的主动生成式过采样网络模型后，向生成器中输入一组与训练所使用的噪声样本分布相同的噪声样本，并将其相应生成器的输出视为少数类样本的补充样本注入原始数据集，以达到平衡数据集的效果。

3.2 基于 DSN 的故障分类模型设计

在经平衡处理的数据集上，构建一种基于 DSN 的故障分类器，即 n 个基本模块，每个模块包含三个网络层。通过将所有前层基本模块的输出预测与原始输入向量进行拼接，作为新模块的输入向量，实现多个模块的“堆叠”。也就是说，基本模块 2 的输入维数等于基本模块 1 的输入与输出维数之和。依此类推，得到包含 n 个基本模块的 DSN 网络，其训练过程可分为两个阶段：基本模块训练和模块约束微调。

在基本模块训练阶段，对所有的基本模块单独进行训练。设 $X = \{x_1, x_2, \dots, x_q, \dots, x_{total}\}$ 表示经过类别平衡后的训练样本集，其中 $x_q = (x_q^1, x_q^2, \dots, x_q^v)^T$ 表示第 q 个样本向量， v 为输入向量的维数， $total$ 为训练样本的总个数。以基本模块 1 为例说明训练过程：假设隐层神经元的数目为 hid^1 、输出向量的维数为 out ，那么第 1 个基本模块的输出为 $y_q^1 = W_2^1 h_q^1$ 。其中， $h_q^1 = \text{sig}(W_1^1 x_q)$ 为基本模块 1 隐含层的输出， W_1^1 为基本模块 1 的输入层与隐含层之间的 $hid^1 \times v$ 的权值矩阵； W_2^1 为基本模块 1 的隐含层与输出之间的 $out \times hid^1$ 的权值矩阵， $\text{sig}(\cdot)$ 表示 sigmoid 激活函数。经过 softmax 函数之后，基本模块 1 输出层的输出变为 $\hat{y}_q^1 = [a_{q,1}^1, a_{q,2}^1, \dots, a_{q,s}^1, \dots, a_{q,out}^1]$ ，其中， $a_{q,s}^1$ 表示基本模块 1 的第 s 个神经元的输出值。与原有 DSN 使用均方误差损失函数不同，本文选用更适合分类的交叉熵损失函数，并按式 (19) 计算、更新梯度 $g_{\theta_1}^1$ ：

$$g_{\theta_1}^1 \leftarrow \nabla_{\theta_1}^1 \left[-\frac{1}{total} \sum_{q=1}^{total} \sum_{s=1}^{out} 1(y_q = \xi) \log(a_{q,s}^1) \right] \quad (19)$$

其中， $1(\cdot)$ 是指示函数，当条件“.”为真时取“1”；当条件“.”为假时取“0”。 $y_q = \xi$ 表示第 q 个样本属于第 ξ 类。同样，利用反向传播法和梯度下降法依次训练其余基本模块。

在模块约束的微调过程中，通过批量模式梯度下降法进一步学习每个模块的权值矩阵。当微调

迭代次数达到预先定义的上界值 Δ_{fin} ，训练过程终止并得到最终的 DSN 模型。整个训练和测试流程如图 5 所示。

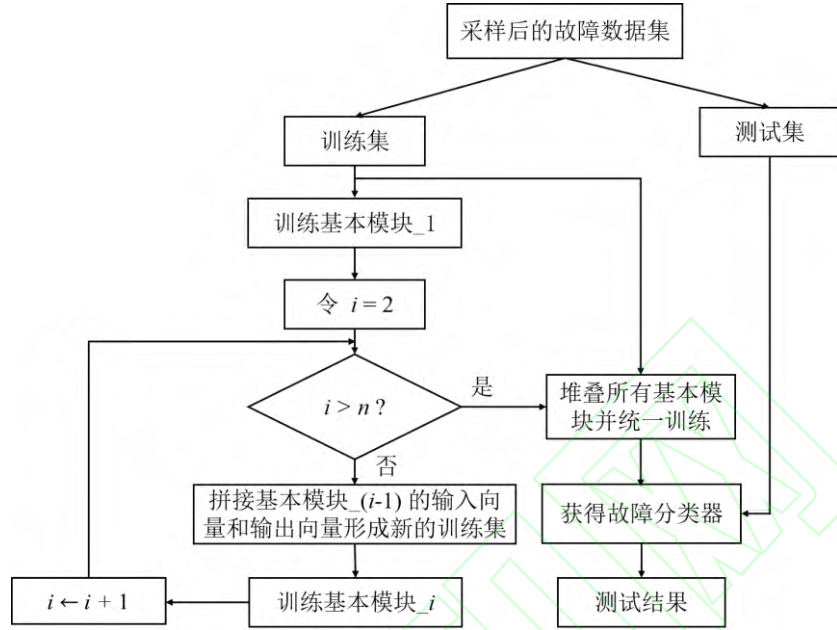


图 5 基于 DSN 的故障分类器训练与测试过程

4 实验结果与分析

为了验证基于主动生成式过采样和 DSN 的故障诊断方法，本文选择凯斯西储大学轴承数据中心的故障数据进行仿真实验。通过与六个基准算法的实验结果比较，说明了本文方法的优越性。

4.1 实验配置

4.1.1 数据集及其预处理

凯斯西储大学^[33]的原始实验数据包含外圈 12 点钟方向故障(Opposite outer race fault, OOF)、滚珠故障(Ball fault, BF)、内圈故障(Inner race fault, IF)和外圈 6 点钟方向故障(Centered outer race fault, COF)共四个故障类别，如图 6 所示。通过移动滑窗和信号交叠方法，选取样本数据并增加样本数量。如图 7 所示，当滑窗位移步长为 130 时，具有 67048 个采样时刻点的振动信号，可以提供 501 个长度为 2048 的故障样本。为了减少个别过大或过小噪声数据对实验结果的影响，本文对所有输入的训练数据特征进行高斯归一化操作，将样本的各个属性值映射到 $[-1, 1]$ 。其次，对故障的类别标签进行 One-Hot 编码，将类别标签转换为由 0 或 1 组成的向量，即类别编号对应的位置被标记为 1，其它位置标记为 0。

为了验证本文过采样方法的有效性，通过对原始数据集进行不放回式随机抽样，对不同样本选取不同的个体数目，从而构造了两个不平衡数据集：数据集 1 和数据集 2。其中，数据集 1 只有一个少数类（200 个样本），而其它 3 个多数类故障各 1000 个样本。数据集 2 包括 2 个少数类（分别有 50 个和 100 个样本）以及两个多数类（各 1000 个样本）。

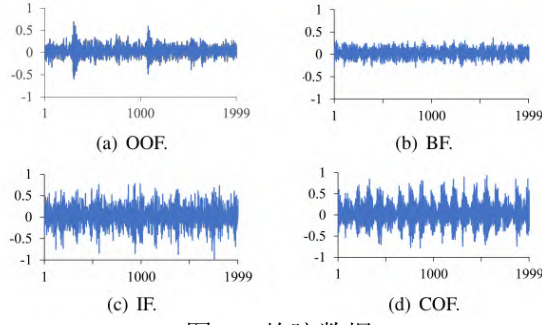


图 6 故障数据

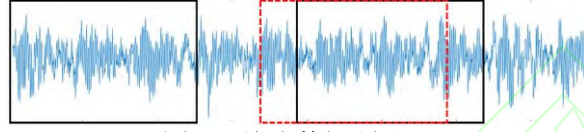


图 7 故障数据预处理

4.1.2 评价指标

为了综合比较本文所提出方法与基准方法，除了常用分类器在正确识别样本类别方面的评价指标，即准确率外，还采用精确率、召回率和 $F1$ score 来评估本文方法识别少数类样本的性能。假设少数类样本为阳性、多数类样本为阴性，用 TP 表示样本数据集中被正确预测的少数类样本数量， FP 表示被预测为少数类样本但实际为多数类样本的数量， FN 表示预测为多数类样本但实际为少数类样本个数， TN 表示被正确分类的多数类样本个数。

1) 准确率 (*Accuracy*)：反映分类器在正确识别故障样本上的性能，可按式(20)计算：

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (20)$$

2) 召回率 (*Recall*)：表示被正确分类的少数类样本占少数类样本总数的比例，可按式(21)计算：

$$Recall = \frac{TP}{TP + FN} \quad (21)$$

3) 精确率 (*Precision*)：反映了被正确预测为少数类的样本在所有被分类为少数类的样本中所占的比例，可按式(22)计算：

$$Precision = \frac{TP}{TP + FP} \quad (22)$$

4) $F1$ ：它综合考虑了 *Recall* 和 *precision*，其计算公式(23)：

$$F1 = \left(\frac{Recall^{-1} + Precision^{-1}}{2} \right)^{-1} \quad (23)$$

4.1.3 对比算法及其参数设置

为了评价本文提出的主动生成式过采样方法的有效性，本文选择三种基于 GAN 变体的过采样方法，如 WGAN、info GAN^[34]和 ACGAN，以及经典的过采样算法 SMOTE 进行对比实验。此外，为了验证本文改进的 ACGAN，即 W_ACGAN 的有效性，选择 ACGAN+QBC 和 ACGAN+AWU 为

对比方法，其中 ACGAN+QBC 表示仅使用 QBC 算法对 ACGAN 进行训练引导，ACGAN+AWU 表示仅使用本文提出的自适应训练方法对 ACGAN 模型参数进行更新。

需要说明，本文的 WGAN、infoGAN 和 ACGAN 均采用相同的模型结构，即生成器输入层包含 2000 个神经元节点，生成器和判别器的隐层包含 1024 个节点，判别器的输出层包含 4 个节点，生成器的输出层和判别器的输入层所包含的节点数与样本的属性维数相同。另外，生成器使用 Relu、Matmul 和 Sigmoid 激活函数，判别器仅使用 Relu 和 Matmul，且训练使用的优化算法为 RMSProp。

在本文的 QBC 算法中，将委员会个数 K 设为 5、熵值较高样本个数 M 设为 10、基于 Diversity 指标挑选的样本数 \hat{M} 设为 5；所设计的故障分类器基本模块数 $n=2$ ，且第一、第二个模块的网络结构分别为 2048-2048-1024-4 和 2052-2048-1024-4，其学习率均设为 0.005，批处理样本大小为 400，迭代次数为 3000，且输出层采用 Softmax 激活函数。

4.2 实验结果分析

为了全面评价本文基于主动生成式过采样和 DSN 的故障诊断方法 AGO (Active Generative Over-sampling) 的性能，我们在两个数据集上开展对比实验，并采用四个评价指标 Accuracy、Recall、Precision 和 F1 score 对实验结果进行分析与评价。

在数据集 1 上的实验，旨在验证不同过采样方法对模型诊断性能的影响。该数据集中只包含一个少数类，即每个过采样方法仅需生成一种类别的样本。由于实际故障分类问题往往包含多个少数类，因此数据集 2 包含两个少数类，以测试不同过采样方法的普适性，即对不同少数类样本的生成情况。

(1) 准确率 (Accuracy) 评价

图 8 展示了不同方法的准确率比较。在数据集 1 上，本文方法 AGO 获得了最高的准确率 94.9%，基于 GAN 的方法，其准确率均超过了 90%，只有 SMOTE 的准确率大约 86%（低于 90%）。由此可知，基于 GAN 的方法总体上能生成相对较好的样本，说明了对抗学习过程的有效性。虽然 SMOTE 能在一定程度上改善准确率，可是其样本生成是通过在现有样本中插入新样本，缺乏学习样本的过程，因此可能使生成样本与原始样本差异太大，甚至引入噪声，从而限制了其准确率的提升。在数

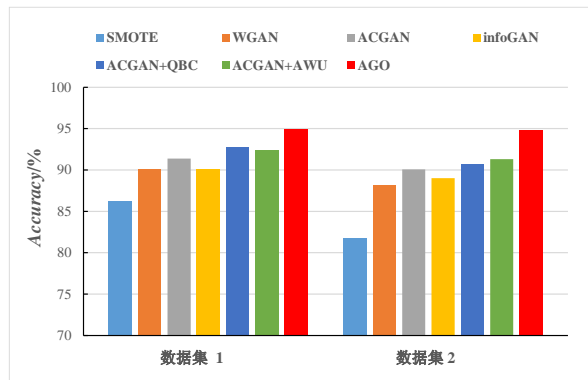


图 8 不同方法的 Accuracy

据集 2 上，本文方法 AGO 也取得了最高的准确率，与其在数据集 1 上的性能表现很接近，说明我们的方法能够很好地适应不同的数据集，且对不同的少数类，也能生成质量较好的少数类样本。

此外，在数据集 1 上，基于 GAN 的过采样方法中，ACGAN 的表现最好，说明引入辅助分类器对提升生成样本的质量有一定作用；WGAN 的准确率最低，说明结构改进和信息引入对于分类效果的提升更加明显。SMOTE 在数据集 1 和数据集 2 上的准确率差别最大，说明 SMOTE 随机插入样本的普适性较差，也从侧面反映出通过学习原始样本分布来生成少数类样本的优越性。

（2）精确率（Precision）评价

图 9 展示了两个数据集上的精确率对比。由于精确率是指模型正确预测的少数类样本，因此可以反映出在生成的少数类样本中引入噪声的情况，引入噪声越少，生成样本与原始样本越接近，即模型学习到的信息越准确，少数类的识别结果也越准确。由图 9 可知，在数据集 1 上，本文 AGO 方法获得了最高的精确率，表明 AGO 在生成与原始少数类样本相似样本的同时，也增加了与多数类样本的差异，更加容易被识别，即引入的噪声很少。此外，ACGAN+QBC 的精确率超过了 90%，说明 QBC 算法的使用可以抑制噪声的引入，生成对分类有利的样本；可是 SMOTE 相对容易带来噪声，从而使模型学到不属于少数类样本或者噪声的特征，最终将部分多数类样本识别成了少数类，所以精确率较低，只有 83%。

相比于数据集 1，SMOTE 在数据集 2 上的精确率下降很明显，这反映了在少数类类别数较多的情况下，SMOTE 更容易引入噪声，其少数类识别的精确率也更低。此外，ACGAN+AWU 在两个数据集上的精确率差别较小，说明了本文的自适应训练策略在数据集中少数类类别数目增多时，能保持较好的性能。在数据集 2 上，基于各类 GAN 变体的过采样方法效果相差不大，而 ACGAN+QBC 和 ACGAN+AWU 的效果相较原始 ACGAN 出现了较大变化，这反映了引入其他训练策略的有效性。

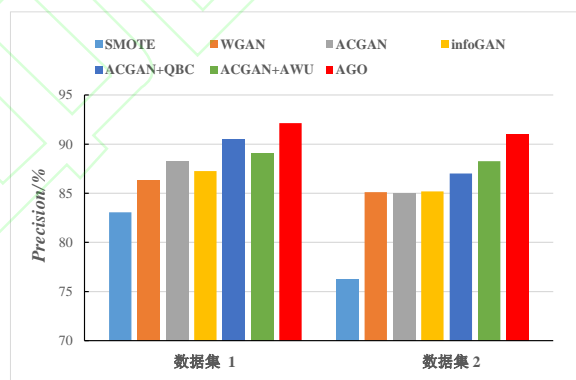


图 9 不同方法的 Precision

（3）召回率（Recall）评价

与精确率对应的另外一个指标为召回率，即少数类样本被识别出来的百分比。该评价指标可以从侧面反映生成的少数类样本的多样性，即当少数类样本足够多样时，训练集中的少数类样本能较好地表征所有少数类样本的模式，模型也能学到足够多样的特征，所以识别出更多的样本。由图 10 可知，在两个数据集上，我们的方法都取得了最高的召回率，充分说明了 AGO 生成样本的多样性。

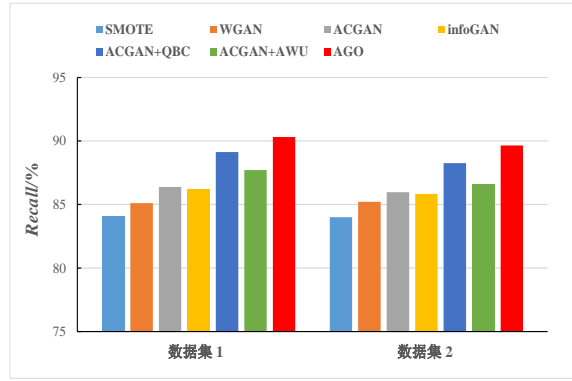


图 10 不同方法的 Recall

此外，ACGAN+QBC 也取得了较好的召回率，说明了 QBC 算法引导的有效性。SMOTE 与基于 GAN 的过采样方法的召回率相差不大，这是由于 SMOTE 的随机插值导致生成样本的模式较多，从某种程度上可以提高模型的泛化能力。所有方法在数据集 2 与数据集 1 上的效果均相差较小，说明了从数据层面解决不平衡分类问题，对于提高少数类样本识别召回率的有效性。生成样本有助于模型学习到较多的少数类样本特征，从而识别出更多的少数类样本。

(4) $F1$ 评价

$F1$ score 是一个综合评价指标， $F1$ score 的值越大，表示该方法不仅可以准确地识别少数类，还可以识别出较多的少数类样本。由图 11 可知，无论是在数据集 1 还是数据集 2 上，我们的方法 AGO 都获得了最高的 $F1$ score，其值约为 91。这就表明 AGO 生成的少数类样本不仅接近于原始样本、引入噪声较少，而且生成样本的多样性也好，可以覆盖较多的少数类故障模式。在数据集 2 上，基于 GAN 变体的过采样方法的 $F1$ score 约为 85，比在数据集 1 上约少 1 分；而 SMOTE 在数据集 2 上的 $F1$ score 大约 80，比在数据集 1 上（约为 84）低 3 分左右，反映出虽然 SMOTE 可以在一定程度上提高少数类的识别效果，但其适应性较差。

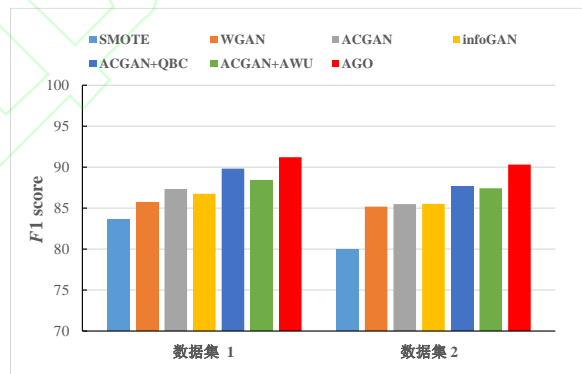


图 11 不同方法的 $F1$ score

5 结束语

面向故障诊断领域存在的类别数据不平衡问题，本文提出一种基于主动生成式过采样和深度堆叠网络的故障诊断方法。首先，针对 GAN 生成的样本质量差且多样性不足问题，提出了一种基于

QBC 和 ACGAN 的主动生成式过采样方法。借助于主动学习算法 QBC 和多样性评价指标 Diversity, 引导 W_ACGAN 的模型训练过程, 并生成多样的、信息量丰富的少数类故障样本, 为后续故障分类效果的提升做好数据准备。同时, 为了提高 ACGAN 模型训练的稳定性以及所生成样本的质量, 采用 Wasserstein 距离代替原有的 JS 散度衡量真实样本分布与生成样本分布的距离, 提出了一种基于损失值的自适应模型参数更新策略, 进一步改善生成样本的质量。最后, 采用过采样后的数据集训练基于 DSN 的故障分类器。两个不平衡故障数据集上的一系列实验结果表明, 本文提出的故障诊断方法, 在少数类故障识别率以及数据集适应性方面, 均优于其基准方法。

本文主要从数据处理层面解决因类别不平衡带来的少数类故障难以识别问题, 在未来的研究中, 我们将进一步从算法/分类器层面入手, 通过设计不同的特征提取或模型训练策略, 以增加模型对少数类样本的关注度, 提升少数类故障识别性能。此外, 我们将尝试更多不同的数据集, 研究并探索故障种类数目较多且边界条件不够清晰时的复杂系统故障诊断问题, 以进一步提高我们设计的故障分类模型的泛化能力。

参考文献:

- [1] LEI Y, YANG B, JIANG X, et al. Applications of machine learning to machine fault diagnosis: A review and roadmap[J]. Mechanical Systems and Signal Processing, 2020, 138:106587.
- [2] CHEN Xiaolu, WANG Ruixuan, WANG Jing, et al. Industrial process monitoring and fault diagnosis based on hybrid discriminant analysis[J]. Acta Automatica Sinica, 2020, 46(8): 1600-1614 (in Chinese). [陈晓露, 王瑞璇, 王晶, 等. 基于混合型判别分析的工业过程监控及故障诊断[J]. 自动化学报, 2020, 46(8): 1600-1614.]
- [3] LI H, HU G, LI J, et al. Intelligent fault diagnosis for large-scale rotating machines using binarized deep neural networks and random forests[J]. IEEE Transactions on Automation Science and Engineering, doi: 10.1109/TASE.2020.3048056, Jan. 18, 2021.
- [4] HU G, LI H, XIA Y, et al. A deep Boltzmann machine and multi-grained scanning forest ensemble collaborative method and its application to industrial fault diagnosis[J]. Computers in Industry, 2018, 100: 287-296.
- [5] KAUR H, PANNU H S, MALHI A K. A systematic review on imbalanced data challenges in machine learning: Applications and solutions[J]. ACM Computing Surveys (CSUR), 2019, 52(4): 1-36.
- [6] MAO W, LIU Y, DING L, et al. Imbalanced fault diagnosis of rolling bearing based on generative adversarial network: A comparative study[J]. IEEE Access, 2019, 7: 9515-9530.
- [7] JOHNSON J M, KHOSHGOFTAAR T M. Survey on deep learning with class imbalance[J]. Journal of Big Data, 2019, 6(1): 1-54.
- [8] FANG Hao, LI Yun. Random undersampling and POSS method for software defect prediction[J]. Journal of Shandong University (Engineering Science), 2017, 47(1):15-21 (in Chinese). [方昊, 李云. 基于多次随机欠采样和 POSS 方法的软件缺陷检测[J]. 山东大学学报(工学版), 2017, 47(1):15-21.]
- [9] MANI I, ZHANG I. KNN approach to unbalanced data distributions: a case study involving information extraction[C] //Proceedings of workshop on learning from imbalanced datasets. United States: ICML, 2003, 126.
- [10] BARANDELA R, VALDOVINOS R M, SÁNCHEZ J S, et al. The imbalanced training sample problem: Under or over sampling?[C] // Proceedings of Joint IAPR international workshops on statistical techniques in pattern recognition (SPR) and structural and syntactic pattern recognition (SSPR). Springer, Berlin, Heidelberg, 2004: 806-814.

- [11] CHAWLA N V, BOWYER K W, HALL L O, et al. SMOTE: synthetic minority over-sampling technique[J]. Journal of artificial intelligence research, 2002, 16: 321-357.
- [12] NGUYEN H B, HUYNH V N. On sampling techniques for corporate credit scoring[J]. Journal of Advanced Computational Intelligence and Intelligent Informatics, 2020, 24(1): 48-57.
- [13] ZHU T, LIN Y, LIU Y. Synthetic minority oversampling technique for multiclass imbalance problems[J]. Pattern Recognition, 2017, 72: 327-340.
- [14] HAN H, WANG W Y, MAO B H. Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning[C] // Proceedings of International conference on intelligent computing. Springer, Berlin, Heidelberg, 2005: 878-887.
- [15] BUNKHUMPORNPAT C, SINAPIROMSARAN K, LURSINSAP C. Safe-level-smote: Safe-level-synthetic minority over-sampling technique for handling the class imbalanced problem[C] // Proceedings of Pacific-Asia conference on knowledge discovery and data mining. Springer, Berlin, Heidelberg, 2009: 475-482.
- [16] HAN Zhiyan, WANG Jian. Fault Diagnosis Method Based on Weighted Synthetic Minority Oversampling Technique[J]. Computer Technology and Development, 2016, 26(9): 43-46 (in Chinese). [韩志艳,王健.基于加权合成少数类过采样技术的故障诊断[J]. 计算机技术与发展, 2016, 26(9):43-46.]
- [17] ANDO S, HUANG C Y. Deep over-sampling framework for classifying imbalanced data[C] // Proceedings of Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, Cham, 2017: 770-785.
- [18] ZHANG K, SU Y, GUO X, et al. MU-GAN: Facial attribute editing based on multi-attention mechanism[J]. IEEE/CAA Journal of Automatica Sinica, 2020.
- [19] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[J]. Advances in neural information processing systems, 2014, 27.
- [20] LEE Y O, JO J, HWANG J. Application of deep neural network and generative adversarial network to industrial maintenance: A case study of induction motor fault detection[C] // Proceedings of 2017 IEEE international conference on big data (big data). IEEE, 2017: 3248-3253.
- [21] DOUZAS G, BACAO F. Effective data generation for imbalanced learning using conditional generative adversarial networks[J]. Expert Systems with applications, 2018, 91: 464-471.
- [22] WU J, ZHAO Z, SUN C, et al. Ss-InfoGAN for class-imbalance classification of bearing faults[J]. Procedia Manufacturing, 2020, 49: 99-104.
- [23] WANG Z, WANG J, WANG Y. 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.
- [24] HAN H, HAO L, CHENG D, et al. GAN-SAE based fault diagnosis method for electrically driven feed pumps[J]. Plos one, 2020, 15(10): e0239070.
- [25] HAN T, LIU C, YANG W, et al. A novel adversarial learning framework in deep convolutional neural network for intelligent diagnosis of mechanical faults[J]. Knowledge-based systems, 2019, 165: 474-487.
- [26] XIE Y, ZHANG T. Imbalanced learning for fault diagnosis problem of rotating machinery based on generative adversarial networks[C] // Proceedings of 37th Chinese Control Conference (2018 CCC). IEEE, 2018: 6017-6022.
- [27] BAO Ping, LIU Yunjie. Research on fault identification based on improved deep model in combination of generative adversarial networks under unbalanced data sets[J]. Journal of Electronic Measurement and Instrumentation, 2019,33(3):176-183 (in Chinese). [包萍, 刘运节. 不平衡数据集下基于生成对抗网络的改进深度模型故障识别研究[J]. 电子测量与仪器学报, 2019, 33(3): 176-183.]
- [28] REN P, XIAO Y, CHANG X, et al. A survey of deep active learning[J]. arXiv preprint arXiv:2009.00236, 2020.
- [29] ODENA A, OLAH C, SHLENS J. Conditional image synthesis with auxiliary classifier GANs[C] // Proceedings of International conference on machine learning. PMLR, 2017: 2642-2651.

- [30] SUN C, MA M, ZHAO Z, et al. Sparse deep stacking network for fault diagnosis of motor[J]. IEEE Transactions on Industrial Informatics, 2018, 14(7): 3261-3270.
- [31] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein generative adversarial networks[C] // Proceedings of International conference on machine learning. PMLR, 2017: 214-223.
- [32] HU Haiyang, ZHUANG Zaixiong, YU Jiacheng, et al. Blurred workpiece angle detection method based on generative adversarial networks[J]. Computer Integrated Manufacturing Systems, 2019, 25(8): 1936-1945 (in Chinese). [胡海洋, 庄载雄, 俞佳成, 等. 基于生成对抗网络的模糊工件角度检测[J]. 计算机集成制造系统, 2019, 25(8): 1936-1945.]
- [33] SUN Chen, WEN Long, LI Xinyu, et al. New automated machine learning based imbalanced learning method for fault diagnosis[J]. Computer Integrated Manufacturing Systems, 2021, 27(10): 2837-2847 (in Chinese). [孙晨, 文龙, 李新宇, 等. 基于自动机器学习的不平衡故障诊断方法[J]. 计算机集成制造系统, 2021, 27(10): 2837-2847.]
- [34] Chen X, Duan Y, Houthoof R, et al. InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets[C] // Proceedings of the 30th International Conference on Neural Information Processing Systems. 2016: 2180-2188.

作者简介:

李慧芳 (1965-), 女, 陕西人, 副教授, 博士, 研究方向: Petri 网、工作流、云计算、任务调度、故障诊断以及深度学习的应用, E-mail: huifang@bit.edu.cn;

徐光浩 (1999-), 男, 河南卫辉人, 硕士研究生, 研究方向: 工作流、云计算、任务调度、故障诊断等, E-mail: 3220200812@bit.edu.cn;

黄双喜 (1972-), 男, 河南人, 副教授, 博士, 研究方向: 企业信息化工程、复杂系统建模与分析以及协同工程等, E-mail: huangsx@tsinghua.edu.cn。