

DOI: 10.19783/j.cnki.pspc.240523

基于改进生成对抗网络的海上风电机组故障数据增强及诊断

魏书荣¹, 殷世杰¹, 闫梦飞², 周海林¹

(1. 教育部海上风电技术工程研究中心(上海电力大学), 上海 200090;

2. 国网上海市电力公司市南供电公司, 上海 201100)

摘要: 海洋复杂运行环境下, 风电机组故障多样, 故障有效样本数据明显不足, 严重影响了故障诊断效果。为解决海上风电运行数据及故障样本积累不足的问题, 提出了一种基于 GRA-rACGAN 生成对抗网络的数据增强方法, 可有效扩充海上风机异常工况数据, 并通过实际运行数据进行诊断验证。首先, 对 SCADA 系统采集的数据进行灰色关联分析(grey relation analysis, GRA), 筛选出与海上风电机组运行状态高度相关的状态变量, 对数据进行归一化处理, 将特征的最小最大范围添加为每个样本的两个附加属性, 避免异常数据干扰, 提高数据生成能力。然后, 将筛选出的状态变量数据集输入至改进型辅助分类器, 采用生成对抗网络进行学习, 扩充故障数据。最后, 以海上风机实际运行数据的增强结果作为样本进行故障诊断, 检验故障数据增强方法的可靠性。通过对海上风电场的实际运行数据实测结果表明, 本模型相比于传统数据增强技术可以有效地生成故障样本, 提高故障诊断的准确率与稳定性, 为海上风机故障的准确预警提供技术支持。

关键词: 海上风机; 数据增强; 灰色关联分析; 辅助分类器生成对抗网络; 故障诊断

Offshore wind turbine fault data enhancement and diagnosis based on an improved generative adversarial network

WEI Shurong¹, YIN Shijie¹, YAN Mengfei², ZHOU Hailin¹

(1. Engineering Research Center of Offshore Wind Technology, Ministry of Education, Shanghai University of Electric Power, Shanghai 200090, China; 2. Shinan Power Supply Company, SMEPC, Shanghai 201100, China)

Abstract: Given the complex operational environment of the ocean, wind turbine faults are diverse, and the effective sample data of faults are obviously insufficient. This seriously affects fault diagnosis results. To solve the problem of insufficient accumulation of offshore wind turbine operational data and fault samples, a data enhancement method based on a GRA-rACGAN generative adversarial network is proposed. This can effectively expand offshore wind turbine abnormal working condition data and carry out diagnosis validation through actual operational data. First, grey relation analysis (GRA) is performed on the data collected by the SCADA system to screen out the state variables that are highly correlated with the operating state of the wind turbines, normalize the data, and add the minimum and maximum ranges of the features as two additional attributes for each sample to avoid the interference of abnormal data and to improve the ability of data generation. Then, the filtered state-variable dataset is fed into an improved auxiliary classifier that employs a generative adversarial network for learning and expanding the fault data. Finally, the reliability of the fault data enhancement method is tested using the enhancement results of actual offshore wind turbine operation data as a sample for fault diagnosis. The measured results of the actual operation data of offshore wind farms show that this model can more effectively generate fault samples and improve the accuracy and stability of fault diagnosis than traditional data enhancement techniques, providing technical support for the accurate early warning of offshore wind turbine faults.

This work is supported by the National Natural Science Foundation of China (No. 52377063).

Key words: offshore wind turbine; data enhancement; grey relation analysis; ACGAN; fault diagnosis

基金项目: 国家自然科学基金项目资助(52377063); 上海市教委自然科学基金重大项目资助(2021-01-07-00-07-E00122); 上海科技创新行动计划资助(22dz1206100); 上海高校特聘教授(东方学者)(TP2020066)

0 引言

海上风电开发是实现国家“双碳”目标的重要举措。海上风电储量巨大,可开发海上风能约 30 亿 kW,且靠近沿海负荷中心,具有消纳方便、效率高、运行小时数长等优势,成为我国发展新能源的一个必然选择,“双碳”目标下海上风电高质量发展具有重大的意义^[1-2]。

海上风电机组电气故障预警与预判意义重大,需求迫切。海上风电机组作为海上风电场中最关键的电能转换装置,其安全可靠运行是海上风电场全生命周期可持续运行亟需解决的“卡脖子”难题,是海上风电场赖以生存的关键。对海上风电机组电气故障进行预警和趋势预判,是保障机组经济安全可靠运行的重要基础。但是,受恶劣海洋环境影响,海上风电机组可及性差、故障率高、维护困难及停运损失大。近期上海启动了全球最大的 580 万 kW 深远海风电示范工程,最远离岸距离达到 90 km,更大的装机容量和更远的离岸距离使得海上风电机组特有的可及性差、故障率高、维护困难的问题愈发加剧,海上风电机组安全运行与故障识别面临巨大挑战。

我国海上风电开发以来,大批量海上风电机组的数据采集与监控(supervisory control and data acquisition, SCADA)系统能够获取海上风电场海量的运行数据并进行远程或本地监控^[3-5],这些海量的时序数据记录了海上风电机组实时运行工况^[6],风电机组的故障特征隐含在能表征其运行状态的数据变量中^[7-9]。截至 2023 年底,我国海上风电累计装机约为陆上风电的 9.34%,海上风电运行数据及故障样本积累明显不如陆上,也无法把陆上风电基于数据驱动的预警方法照搬至海上。由于海上机组扩充未发生的故障样本信息困难,且多年的海上风电运行经验表明,机组故障并非独立、隔绝的,而是呈现互相关联、连锁显现的态势,仍依靠传统陆上风电的数据驱动经验模型方法局限性凸显。

考虑 SCADA 系统的数据采样频率较低,大部分参数每隔 10 min 采集一次,且风机长时间处于正常工作状态,运行数据冗余但故障样本稀缺^[10-11],严重制约了故障诊断模型的准确性,特别是发电机故障在各种风电机组故障中占比较低,导致其正常样本和故障样本存在严重的数据不平衡问题。在数据不平衡的情况下,模型难以通过不平衡的数据进行训练,实现精确的故障诊断任务^[12]。

为了解决故障数据不足的问题,文献[13]提出了生成对抗网络(generative adversarial network, GAN)

方法,该方法目前已普遍应用在了计算机视觉和文本分析等领域。在传统的深度学习中,生成模型的优化目标一般是将生成样本与真实样本之间的绝对误差达到最小,如变分自编码器等。这种优化目标侧重于生成样本的真实性,但却忽略了样本的多样性。与此不同,生成对抗网络通过学习原始数据的分布,并在判断器的约束下,建立从高斯分布到真实数据分布的映射关系,以产生新的样本,显著提高了样本的多样性。

针对生成对抗网络扩充数据的真实性的研究,已受到国内外学者的关注,如文献[14]采用用于插补的门控循环单元神经网络细胞来处理不规则时滞关系,从而提高 GAN 输入数据的稳定性和可靠性。文献[15]通过在训练稳定性优越的边界平衡生成对抗网络基础上,融合条件信息和辅助局放脉冲分类任务,实现了对于均衡脉冲样本的修复与多源放电诊断。文献[16]通过在辅助分类器生成对抗网络(auxiliary classifier generative adversarial network, ACGAN)后引入自然梯度提升的主轴承状态监测模型,使主轴承温度异常残差可以在控制线范围内进行筛选,提高数据置信度。

以上方法对于输入 GAN 的状态变量均来自经验选取,而未考虑状态变量间的相关性^[17],且未关注数据边界条件的约束。仅依据经验选取状态变量的准确性受限,在此情况下 GAN 训练过程中极易出现被极端数据干扰学习的情况,严重制约了生成数据的准确性。

本文提出一种基于 GRA-rACGAN 生成对抗网络的数据增强与诊断方法。首先构建基于 GRA-rACGAN 生成对抗网络的数据增强模型,有效弥补海上风电场故障数据样本不足的难题。对 SCADA 运行数据中的状态变量进行相关性分析,提取与海上风电机组故障高度相关的状态变量作为输入数据。为了稳定训练过程并减少海上突风带来的影响,引入边界寻优损失,对 ACGAN 模型中数据分类的正确概率损失公式进行改进,避免了海上突风状态下数据异常导致的数据分类不平衡。将特征的最小最大范围添加为每个样本的两个附加属性,避免在 GAN 中极端数据引起的模式崩溃问题,提高数据训练的稳定性。最后以国内某海上风电场 SCADA 实际运行数据为例进行分析,证明本文提出的 GRA-rACGAN 数据增强方法与故障诊断结果的准确性。

1 改进型辅助分类器生成对抗网络算法

1.1 ACGAN

GAN 是一种基于纳什均衡理论的博弈数据生

成算法。该算法由生成器 G 和判断器 D 组成^[18], 生成器不断生成接近真实的数据, 与判断器进行博弈, 从而产生更为逼真的数据样本。该优化过程可以表述为如式(1)所示的目标函数。

$$\min_G \max_D V(D, G) = E_{x \sim P_{\text{data}}(x)} [\log D(x)] + E_{z \sim P_z(z)} [\log(1 - D(G(z)))] \quad (1)$$

式中: E 为期望的分布; $G(z)$ 为生成器生成数据; $D(x)$ 为判断器网络输出。

该最大最小优化过程意味着, 当生成器和判别器的博弈达到纳什均衡时, 生成器能够学习到符合真实数据分布的规律。然而传统的 GAN 存在着网络不收敛、不稳定模式极易崩溃的问题, 对此条件生成对抗网络(conditional generative adversarial network, CGAN)、半监督生成对抗网络(semi-supervised generative adversarial network, SGAN)和信息最大化生成对抗网络(information maximizing generative adversarial network, infoGAN)在一定程度上予以解决, 但都存在各自的局限性。

CGAN 在生成器与判断器中加入了额外信息 y , 使得 GAN 模型扩展为条件模型, 其中 y 可以是任何形式的条件变量, 例如类标签等。CGAN 的局限在于不能达到完全无监督学习的效果, 所有的数据类别必须手动输入。

SGAN 中生成器的输出分为两个部分, 分别是分类向量与伪选项, 如果类的个数为 N , 输出向量的长度则为 $N + 1$, 判断器的输入是真实数据和没有类信息的生成数据。虽然通过这种输出使 SGAN 达到了识别数据的能力, 但是生成器并不能按类别生成数据。

InfoGAN 使用互信息理论, 试图防止互信息 c 在生成器中丢失。相比原始 GAN, 其多了一项 $\lambda I(c; G(z, c))$, 代表 c 与生成器输出的互信息。这一项越大, 表示 c 与生成器输出的相关性越大。但 infoGAN 的局限性在于潜变量不是实验者可以控制的类别。

针对 CGAN、SGAN 和 infoGAN 各自的不足, ACGAN 分别进行了改进^[19-20], 结合了 CGAN、SGAN 和 infoGAN 的优势, 通过赋予判断器判断数据真实程度和分类的能力, 使生成器能够同时最大限度地利用噪声 Z 和标签 C 的信息。ACGAN 的损失函数有两部分^[21], 如式(2)和式(3)所示。

$$L_S = E[\log P(S = \text{real} | X_{\text{real}})] + E[\log P(S = \text{fake} | X_{\text{fake}})] \quad (2)$$

$$L_C = E[\log P(C = c | X_{\text{real}})] + E[\log P(C = c | X_{\text{fake}})] \quad (3)$$

式中: S 为样本的真假标签; X_{real} 为真实样本, X_{fake}

为生成样本; P 为概率; L_S 为记录数据为真的概率; L_C 为记录数据分类正确的概率。

对于判断器 D 的要求是 L_S 和 L_C 的和达到最大, 而生成器 G 则要求 L_C 和 L_S 之差达到最大, 当二者达到纳什均衡时, 便可认为输出的数据可信。

1.2 基于边界条件的改进型 rACGAN

虽然 ACGAN 融入了标签信息并具备辅助分类能力, 可以控制生成样本模式并显著提高生成样本质量, 但 ACGAN 易存在训练不稳定、分类效果不佳等问题。同时海洋复杂运行环境下, 风电机组故障多样, 但多样性故障样本不足。与陆上风机不同, 海洋特有的风、涌、浪、流引起的风速、风向、湿度、湍流等海洋运行环境对海上风机运行状态影响巨大。如图 1 所示, 海上常见的突风使得海上机组及控制系统启停频繁, 且海上湍流强度大, 风轮和叶片将不平衡载荷的频繁冲击传递给电机, 使其承受机械应力、电应力等各种振动。此外, 如图 2 所示, 冬季海上风电机组故障频发, 机组运行受海洋气象因素影响显著, 使得海上风电机组运行状态与陆上差异显著。根据 SCADA 系统运行记录显示, 原始数据集中含有多类运行场景及历史相关数据等多类特征, 不同场景之间的数据差异较大。

海上风电由于复杂的运行环境和噪声影响, SCADA 系统记录的特征之间相关性关系变化较陆上风电更为明显。传统生成对抗网络在复杂海洋环境下的应用存在局限性, 难以生成具有多样性的样本, 制约了诊断模型的精准性。因此, 本文在 ACGAN 的基础上进一步提出改进型 rACGAN: 首先, 将输入数据进行归一化处理, 并将特征的最小和最大范围添加为每个样本的附加属性; 其次, 基于数据的两种极端特征重新定义损失函数, 并对输入噪声信号

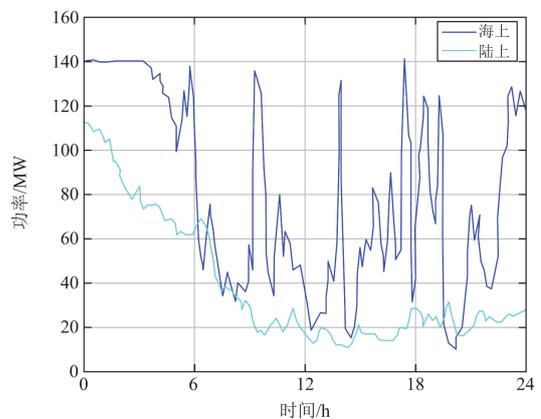


图 1 海上和陆上风电场功率变化图

Fig. 1 Graphs of power changes on-shore and off-shore wind farm

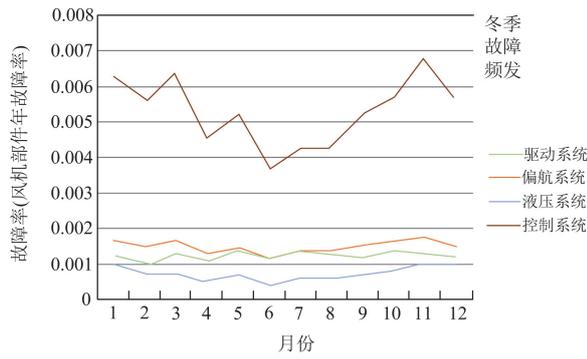


图2 受天气影响的海上风电机组故障率分布

Fig. 2 Distribution of weather-affected offshore wind turbine failure rates

进行预处理；最后，在判断器中加入小批量鉴别，并引入正则化互信息以增强模型在噪声标签数据场景下的鲁棒性，实现强鲁棒分类。rACGAN 在处理数据特征之间的相关性关系信息上表现更为灵活和先进。

同时，为了稳定训练过程并减少突变工况带来的影响，本文引入了边界寻优损失。为了提高训练的稳定性，首先对离散数据提出了边界寻求损失。这一策略也可以推广到连续数据。边界 L_B 损失公式为

$$L_B = E \left[\frac{1}{2} (\log D(G(z)) - \log(1 - D(G(z))))^2 \right] \quad (4)$$

另外，考虑到数据分类的不平衡，对数据分类的正确概率损失公式进行如式(5)所示的修改。

$$L_C = E \left[\log \sum_j P(j|X_{\text{real}}) C(j|i) \right] + E \left[\log \sum_j P(j|X_{\text{fake}}) C(j|i) \right] \quad (5)$$

式中， $C(j|i)$ 表示将一个样本从其真实类别 i 分类到预测类别 j 的误分类代价。因此，生成器试图最小化 $L_C + L_B$ ，判断器最小化 $L_C - L_B$ 。

rACGAN 对发生器进行优化，克服了故障特征的不稳定性和实际应用场景中提取特征的困难，生成无噪声故障样本，具有以下优点：

1) 解决带噪声标签及有限数据驱动状态下，海上风电机组实际运行工况中故障数据生成精度不足的问题；

2) 提高故障数据广义特征提取的效果；

3) 提高故障分类的效果。

2 数据增强模型框架及实验数据集的构建

2.1 数据增强模型构架

本文通过改进型 rACGAN 框架扩充风电机组发电机的故障样本，以弥补故障数据不足的问题，对

风电机组发电机故障诊断提供充足的故障样本。本文的改进型 rACGAN 框架如图 3 所示。

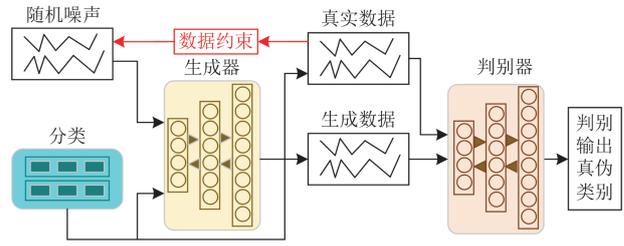


图3 rACGAN 框架

Fig. 3 Framework of rACGAN

在模型的网络结构中，为了提高模型的稳定性，在各隐含层后均添加了批量标准化层，从而使每层的输出均匀一致。对于深度学习网络来说，优化算法与相关参数选择对其性能有着很大影响，所以在搭建时需要进行反复调试与对比，形成生成器与判别器的最终生成参数。

2.2 灰色关联分析

在输入参数的选取阶段，本文使用灰色关联分析(grey relation analysis, GRA)对 SCADA 状态变量进行相关性分析。GRA 可以很好地计算出海上风电机组运行产生的各状态参数之间的相关关系，同时准确提取各变量间的主要特征^[22-24]。当海上风电机组的运行状态参数繁杂时，GRA 可以有效删除无关变量，提取出有效特征，避免了因为输入状态变量过多而影响故障诊断的精度与速度。GRA 的计算步骤如下所述。

1) 将数据划分成反映系统行为特征的参考数列与影响系统行为的比较数列两类。即表述风力发电机状态的数据为参考数列 Y ，而发电机待选参数则组成了数列 X ，如式(6)和式(7)所示。

$$Y = Y(k) | k = 1, 2, \dots, n \quad (6)$$

$$X_i = X_i(k) | k = 1, 2, \dots, n; i = 1, 2, \dots, m \quad (7)$$

式中： X_i 表示比较数列； k 表示数列中的位置； i 表示比较数列中的一个特征； n 表示数列的数量。

2) 考虑到不同数列之间的量纲不同，不便于对比，因此需要对数列进行无量纲化处理，如式(8)所示。

$$X_i(k) = \frac{X_i(k)}{X_i(1)} | k = 1, 2, \dots, n; i = 0, 1, 2, \dots, m \quad (8)$$

3) 计算关联系数 $\xi_i(k)$ 的公式如式(9)所示。

$$\xi_i(k) = \frac{\left[\min_i \min_k |X_0(k) - X_i(k)| + \rho \min_i \min_k |X_0(k) - X_i(k)| \right]}{\left[|X_0(k) - X_i(k)| + \rho \max_i \max_k |X_0(k) - X_i(k)| \right]} \quad (9)$$

式中, ρ 为分辨系数, 其值越小代表分辨能力越大, 当 $\rho < 0.5463$ 小时分辨力最好, 通常 $\rho = 0.5$ 。

4) 考虑到对比每个比较数列时每个点的关联系数各不相同, 数值过于分散不利于数列之间的互相比较, 因此 GRA 算法中将每组数列的关联系数求平均值, 作为比较数列与参考数列之间的关联度 r , 如式(10)所示。

$$r = \frac{1}{n} \sum_{k=1}^n \xi_i(k) \quad (10)$$

5) 将不同比较数列的关联度按从小到大进行排序, 即可得到相关性更大的参数。

2.3 算法流程

基于本文提出的 GRA-rACGAN 模型, 可以根据需求生成高质量故障数据, 通过模型中的分类模块, 可实现海上风电机组多种故障类型的生成与识别。基于 GRA-rACGAN 的海上风机故障数据增强与识别框架如图 4 所示, 故障数据生成的主要步骤如下。

1) 数据预处理: 利用灰色关联分析 GRA 分析影响海上风电机组运行状态的主要因素, 不仅考虑发电机自身因素, 还考虑复合因素以及海洋环境因素的影响, 通过灰色关联度分析, 将高度相关的海上风电机组运行状态变量作为改进型 ACGAN 的输入。

2) 数据增强: 根据输入的样本库训练 rACGAN, 并针对样本不均衡情况, 设置 cfake 指导生成器合成少数类脉冲样本, 扩大训练集规模, 通过测试生成样本在分类模型中的表现, 评估数据增强的效果。

3) 多故障类型识别: 利用 rACGAN 的分类器进行预训练, 并结合数据增强样本对分类模型进行微调, 以稳定识别故障样本中每类信号的数据特征, 达到可以检测出多种海上风电机组故障的效果。

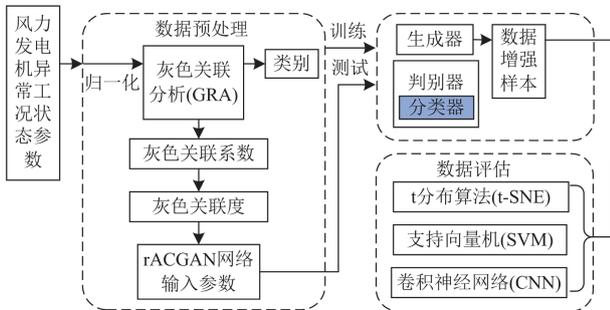


图 4 基于 GRA-rACGAN 的风机故障数据增强与识别框架
Fig. 4 A framework for wind turbine fault data enhancement and recognition based on GRA-rACGAN

3 实测算例分析

3.1 海上风电场实际运行数据描述与分析

本文使用国内已投运十余年的某海上风电场实际 SCADA 数据进行分析, 该风电场共装备 34 台 3 MW 海上双馈风力发电机, 每台机组均配有 SCADA 系统, 该系统每 10 min 储存一次风机状态监测数据。本文选取 2012—2018 年的 SCADA 数据, 并结合风电场台账及维修记录进行研究。其中, 台账记录了确切的停机时刻, 如图 5 所示。该风机于 2018 年 8 月 21 日 11 时 4 分 27 秒停机。

照片 追踪	
塔筒偏差驱动轴方向 [米/秒²]	0.00
塔筒偏差非驱动轴方向 [米/秒²]	0.00
偏航状态	(20) 自动
偏航刹车液压状态	1
机舱位置故障指令 [度]	6.53
偏航变频器功率 [千瓦]	0
偏航变频器故障	0
偏航刹车压力 [巴]	161.49
主刹车储能压力 [巴]	0.79
主轴承温度 [摄氏度]	52.22
齿轮箱轴承温度001 [摄氏度]	53.07
齿轮箱轴承温度002 [摄氏度]	57.67
齿轮箱轴承温度003 [摄氏度]	55.97
齿轮油进油温度 [摄氏度]	53.32
齿轮油电气泵压力 [巴]	1.16
齿轮油机械泵压力 [巴]	1.99
齿轮油进油压力 [巴]	2.28
发电机前端轴承温度 [摄氏度]	42.32
发电机后端轴承温度 [摄氏度]	43.87
发电机U相绕组温度 [摄氏度]	56.57
发电机V相绕组温度 [摄氏度]	51.67
发电机W相绕组温度 [摄氏度]	57.17

图 5 国内某海上风电场台账记录

Fig. 5 A domestic offshore wind farm ledger records

根据风电场台账和维修记录, 发现风力发电机的滑环碳刷磨损、转速异常或者绕组温度过高等异常工况频发。本文选择 3 个出现频次最高的典型异常工况进行数据扩充, 经过对实际运行数据进行统计, 本文选择发电机滑环碳刷磨损、发电机三相绕组温度过高和发电机水冷循环故障。这 3 类发电机异常工况均发生在 22 号机组, 具体信息如表 1 所示。

表 1 22 号机组发电机异常工况信息描述

Table 1 Description of unit 22 generator abnormal operating condition information

编号	异常工况描述	开始时间	结束时间
工况 1	发电机滑环碳刷	2012-10-01	2012-10-04
	磨损	09:25	09:55
工况 2	发电机三相绕组	2012-10-08	2012-10-10
	温度过高	15:50	17:50
工况 3	发电机水冷循环故障	2016-03-16	2016-03-17

3.2 GRA 状态量提取

风电机组 SCADA 系统监测点^[25-26]较多, 图 6 展示了部分监测点布置位置。为了建立有效的异常工况检测模型, 需要从 SCADA 系统监测的 100 多个运行参数中选取与发电机运行状态相关的变量作

为检测模型的输入, 使用 GRA 选择与风力发电机故障相关性较高的参数。

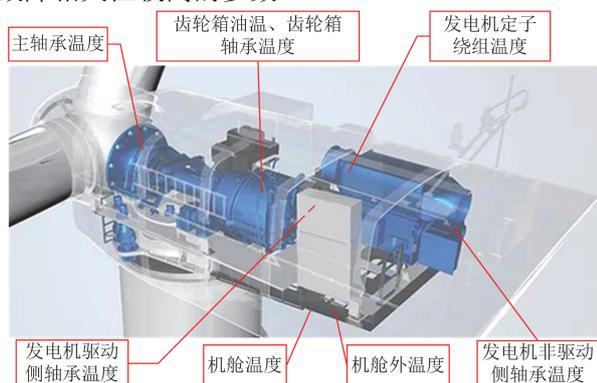


图 6 温度传感器布置示意图

Fig. 6 Temperature sensor arrangement diagram

在风力发电机组的温度监测中, 风电场主要有 9 个点的温度被监测: (1) 机舱温度; (2) 机舱外温度; (3) 齿轮箱油温; (4) 齿轮箱轴承温度; (5) 发电机轴承温度; (6) 定子绕组温度; (7) 主轴承温度; (8) 发电机冷却水入水口温度; (9) 发电机冷却水出水口温度。通过监测各主要部位的温度了解风力发电机组系统的实时运行状况。由于温度自身的热惯性特点, 温度变量具有很好的抗扰动能力, 可以有效反映海上风机部件劣化程度的演变, 且海上风电机组温度变量的测点较多, 综合考虑选择温度变量作为重点分析的状态变量。通过监测各主要部位的温度了解风力发电机组系统的实时运行状况。根据 SCADA 数据中温度状态变量的变化趋势, 判断部件的劣化程度以及是否已经发生故障。

GRA 状态量提取过程可以简述如下:

1) 根据专家经验, 将 SCADA 系统中收集的与风力发电机相关的 15 个状态变量设为比较序列, 进行归一化处理, 从而消除量纲不同的影响;

2) 选择将风力发电机的各测点温度设为参考序列, 设为 Y , 分别计算 Y 与其他状态变量的灰色关联系数;

3) 使用熵权理论优化各权重, 计算各个状态变量与发电机温度之间的灰色关联度;

4) 将灰色关联度排序, 选择高相关度的状态变量作为 rACGAN 的输入。

经过上述步骤, GRA 计算结果按灰色关联度的大小进行排序, 排序结果如表 2 所示。

根据灰色关联分析的原理, 灰色关联度越大, 则说明比较序列越对参考序列有影响, 因此通过表 2 可以看出发电机绕组温度与其他状态变量的灰色关联度最大, 而齿轮箱轴承温度的灰色关联度最小,

一般认为, 灰色关联度在 0~0.35 之间时, 属于低关联度; 灰色关联度在 0.35~0.65 之间时, 属于中等强度关联度; 当灰色关联度大于 0.65 时, 即为强关联度。因此, 可以忽略灰色关联度较小的状态变量, 只选取关联度大于 0.65 的状态变量作为改进型辅助分类器生成对抗网络模型的输入。

表 2 检测模型的输入参数

Table 2 Input parameters for the detection model

编号	状态变量	关联度	编号	状态变量	关联度
1	发电机定子绕组温度	0.8986	10	发电机电网侧电压	0.6120
2	风速	0.8354	11	齿轮箱油温	0.598 55
3	发电机定子绕组线电流	0.8057	12	发电机转速	0.5685
4	发电机出水口温度	0.7865	13	主轴承温度	0.5386
5	功率	0.7752	14	发电机电压	0.5267
6	环境温度	0.7351	15	齿轮箱轴承温度	0.4955
7	湍流	0.7098	16	发电机驱动侧轴承温度	0.4752
8	发电机入水口温度	0.7006	17	发电机非驱动侧轴承温度	0.4257
9	机舱温度	0.6641	—	—	—

3.3 数据集准备与模型训练

将 SCADA 中的数据按照 GRA 筛选出来的状态变量作为实验样本, 并将实验样本划分为训练集与测试集两部分, 其中训练集用于对生成对抗网络进行训练, 测试集用于验证训练的效果。经过筛选, 得到正常运行数据有 16 052 组, 3 种工况对应的数据分别为 4350 组、3060 组和 1780 组。因此将故障工况数据作为数据集 1, 并将 3 类异常工况各取 1500 个样本作为测试集。同时考虑到本模型需要应用在故障数量稀少情况下以达到数据增强的效果, 因此在数据集 1 中随机选取各种异常工况的数据 200~300 个作为数据集 2, 将数据集 2 作为训练集。

通过该模型对风电发电机故障的 3 种典型异常工况样本进行数据扩充。首先将训练集分别输入 ACGAN 模型与改进型 rACGAN 模型中进行训练, 每种异常工况样本设置为 1000 个生成量。其中 ACGAN 的结构参数与改进型 rACGAN 设置相同, 进行实验对比, 比较两种模型的生成器与判断器的损失情况, 结果如图 7 所示。在 ACGAN 模型中, 生成损失代表着模型生成数据与真实数据的损失, 分类损失代表着模型分辨样本是真实或生成的损耗, 两者共同衡量了生成对抗网络在生成数据质量以及判断数据可靠性方面的能力。由图 7 可见, 在

相同的迭代次数下,改进型 rACGAN 的生成损失与判断损失以及收敛速度均优于 ACGAN 模型。

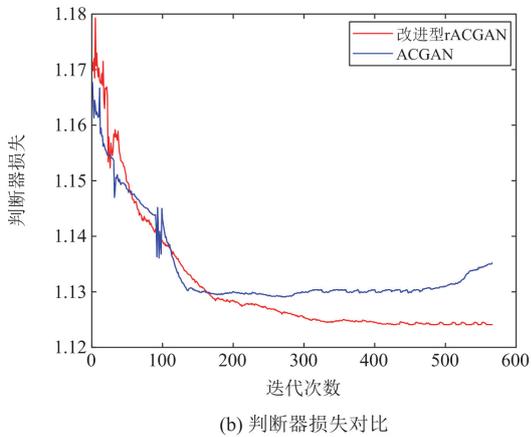
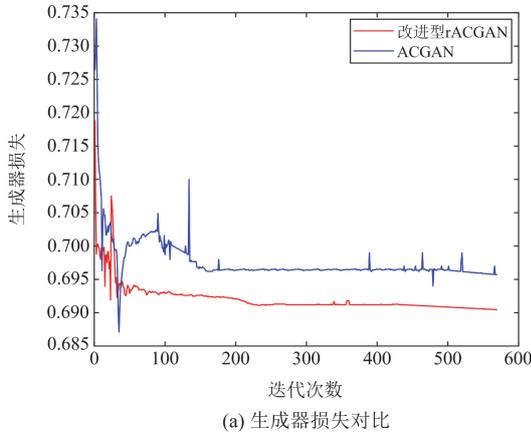


图 7 改进型 rACGAN 与 ACGAN 的性能对比

Fig. 7 Performance comparison between improved rACGAN and ACGAN

对比可得:改进型 rACGAN 模型的生成损失、分类损失两个指标均超过 ACGAN 模型。

3.4 模型生成数据评估与应用

3.4.1 数据增强方法的分类效果对比

分别将 ACGAN^[27]与改进型 rACGAN 生成的异常工况数据划分为生成集 1 与生成集 2,生成集中风力发电机每种异常工况均有 1000 个样本。同时分别将原 ACGAN 生成的生成集 1 与改进型 rACGAN 生成的生成集 2 与原始故障数据集 1 中的部分数据合并,分别命名为增强集 1 与增强集 2。经过数据增强后,增强集 1,2 中风力发电机异常工况每种均能达到 1500 以上。为了直观表示,数据集划分如表 3 所示。

为了验证本文对于风力发电机组异常工况数据增强的优越性,可以使用 t-分布随机邻域嵌入算法(t-distributed stochastic neighbor embedding, t-SNE)将本文改进型 rACGAN 的扩充样本进行降维分析,

并通过可视化导出,将数据集 2 与生成集 2 输入 t-SNE 进行降维处理后,本文改进型 rACGAN 生成数据在分布上与真实样本接近,而且可以在一定程度上生成在原始数据中未出现过的高仿真样本,扩充了风力发电机异常工况数据的多样性,如图 8 所示。

表 3 实验数据集划分

Table 3 Experimental dataset partitioning

工况	数据集 1	测试集	数据集 2	生成集 1	生成集 2	增强集 1	增强集 2
1	4350	1500	260	1000	1000	1650	1650
2	3060	1500	250	1000	1000	1520	1520
3	1780	1500	280	1000	1000	1580	1580

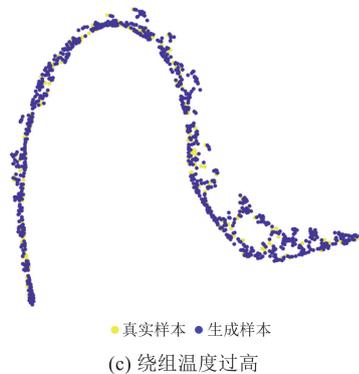
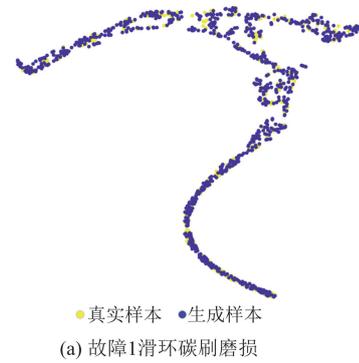


图 8 生成数据可视化对比

Fig. 8 Visualization comparison of the generated data

同时针对本文改进型 rACGAN 对于 3 种异常工况的生成能力, 通过 t-SNE 降维方式可视化表达如图 9 所示, 从图中可以看出, 改进型 rACGAN 生成的 3 种异常工况各自聚集成群, 界限明显, 证明了此生成对抗网络生成数据的准确性。

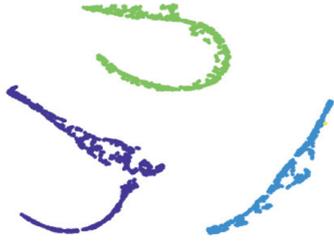
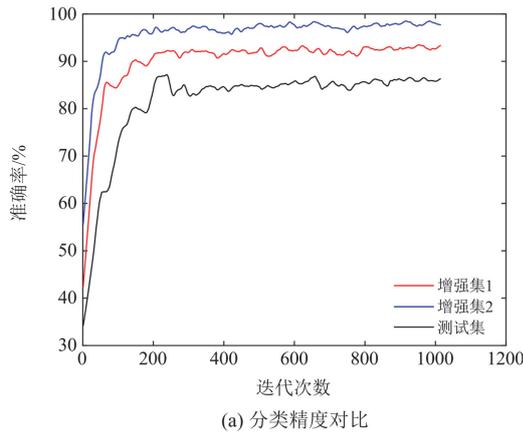


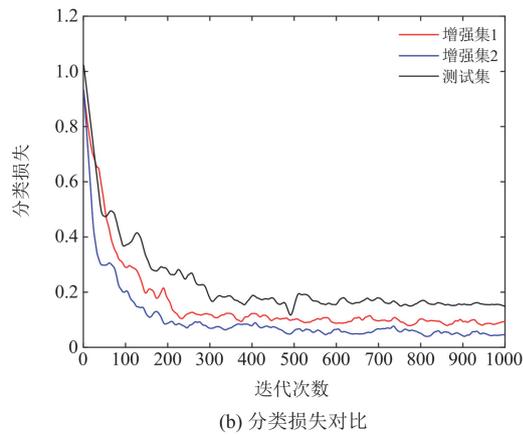
图 9 生成数据可视化对比

Fig. 9 Visualization comparison of the generated data

同时, 采用 CNN^[28] 算法用于对生成数据分类性能测试, 将测试集、增强集 1 和增强集 2 分别作为训练集输入 CNN 网络进行分类, 在迭代 1000 次后, 其分类损失与分类精度对比如图 10 所示, 为了方便



(a) 分类精度对比



(b) 分类损失对比

图 10 不同数据集分类性能对比

Fig. 10 Comparison of classification performance on different datasets

对比, 对于曲线进行了平滑(smooth)处理, 可以直观对比几种方式分类效果的优劣。从图中可以看出, 由于测试集数据量少且分布不平衡, 测试集的效果最差, 一直到迭代 300 次才达到收敛, 分类精度上升与分类损失下降最慢, 而本文改进型 rACGAN 的分类效果最高, 生成集 2 与增强集 2 的分类精度均最高, 而且收敛速度更快, 可以看出数据增强效果明显优于原 ACGAN。

同时, 为了验证数据增强效果的准确性, 本文使用支持向量机(support vector machine, SVM)^[29]、极限学习机(extreme learning machine, ELM)^[30]这两种目前普遍使用的分类算法作为对比。将每种异常工况取 100 个样本作为测试集, 各分类模型准确率如表 4 所示。

表 4 各种分类算法的准确率对比

Table 4 Comparison of accuracy of various classification algorithms

算法	准确率		
	测试集	增强集 1	增强集 2
CNN	0.961	0.980	0.991
SVM	0.949	0.970	0.983
ELM	0.953	0.976	0.985

由表 4 可以看出, 经过本文改进型 rACGAN 数据增强的生成集 2 与增强集 2 在不同分类诊断算法中均有出色的表现, 证明本文改进型辅助分类器生成对抗网络相对于原框架训练过程更稳定, 在海上复杂环境及突风工况下, 其生成数据的质量更接近于原始数据, 能够有效平衡复杂海洋环境下的故障数据, 进一步提升了海上风电机组故障诊断的正确率。

3.4.2 数据增强方法的故障诊断能力对比

将本文提出的数据增强方法应用到海上风电机组的故障诊断中, 使用 Stacking 融合的故障诊断算法^[31]对故障增强的效果进行了测试。基于 Stacking 融合算法可以充分融合各种不同优势的算法, 从多个角度对海上风电机组 SCADA 数据进行深入挖掘, 使各种算法优势互补, 获得更优的诊断结果, 有效提高了海上风力发电机早期故障预警的诊断精度。

为了评估故障诊断模型的性能, 采用正确率 R_T 、漏警率 R_F 和 F_1 -score 三个指标评估故障诊断的有效性。各指标计算过程为

$$R_T = \frac{TP}{TP + FP} \quad (11)$$

$$R_F = \frac{FN}{TP + FN} \quad (12)$$

$$F_1\text{-score} = \frac{2R_T(1 - R_F)}{R_T + (1 - R_F)} \quad (13)$$

式中： TP 表示正确诊断出特定故障的个数； FP 表示错将其他状态诊断为特定故障的个数； FN 表示将特定故障诊断为其他状态的个数； F_1 -score 表示分类器综合性能，是 R_T 和 R_F 的调和平均值。

诊断模型评价指标中，故障诊断效果与 R_T 和 F_1 -score 呈正相关，而与 R_F 呈负相关。

表 5 Stacking 融合算法诊断结果
Table 5 Diagnosis results of Stacking fusion algorithm

诊断数据	故障类型及评价指标								
	工况 1			工况 2			工况 3		
	R_T	R_F	F_1 -score	R_T	R_F	F_1 -score	R_T	R_F	F_1 -score
测试集	94.47	5.89	95.55	94.38	5.62	94.18	98.90	1.10	98.28
增强集 1	95.64	4.36	96.98	95.15	4.85	95.45	98.88	1.00	98.60
增强集 2	97.37	2.63	98.89	97.00	3.00	97.65	99.50	0.50	99.50

由表 5 可知，对于 3 种异常工况，使用本文改进型 rACGAN 样本增强效果的增强集 2 的 R_T 值与 F_1 -score 值最高， R_F 值最低。可以证明本文改进型 rACGAN 样本增强效果在风电机组故障样本增强与诊断上的可靠性。

同时，本文对比分析了 RF、SVM、KNN、GBDT、XGBoost 算法的诊断结果，如表 6 所示。

表 6 诊断模型性能比较
Table 6 Performance comparison of diagnostic models

诊断模型	诊断精度		
	测试集	增强集 1	增强集 2
RF	94.19	96.79	96.88
SVM	74.76	76.21	79.87
KNN	87.20	88.78	91.89
GBDT	95.51	95.81	96.37
XGBoost	97.39	98.99	98.12

从表 6 可知，增强集 2 在各算法中的诊断精度均有显著提升，这表明通过本文方法实现的数据增强可以显著提高数据的诊断能力，再次验证了本文提出的 GRA-rACGAN 算法在海上风力发电机故障数据增强中的有效性。

4 结论

针对目前海上风电机组故障诊断中面临的故障样本少且不同类别的数据样本不平衡的问题，提出了一种基于边界损失约束条件的改进 GRA-rACGAN 数据增强模型，该模型具有以下优点：

1) 通过灰色关联分析与提取与风机发电机故障呈高度相关的状态变量，避免了基于经验选择的

将 3.1 节所述的 3 种海上风力发电机故障的数据输入到 Stacking 融合模型中，诊断输出的四种发电机运行状态，即正常、工况 1、工况 2 和工况 3。使用该诊断模型分别对测试集、增强集 1 与增强集 2 的 4 种发电机状态样本进行诊断，诊断结果如表 5 所示。

偶然性；

2) 将基于边界损失约束条件的改进型 rACGAN 模型应用到海上风电机组故障诊断领域，从有限的非平衡故障样本中提取出数据特征，在一定程度上降低了复杂海洋环境及海上频发的突风等影响，生成得到质量更高的故障样本，实现了故障数据增强，显著提升了海上风电机组基于 SCADA 运行数据的故障诊断精度。

本文基于国内某海上风电场的实际运行数据进行数据增强与故障诊断，结果验证了改进型 rACGAN 模型的准确性，表明了本文所提方法能有效地应用于实际海上风电机组异常工况检测，为海上风电特别是深远海风电开发中风电机组的精准故障预警提供了技术支持。

本文通过改进生成对抗网络寻找数据发展规律、扩充完善故障数据。目前，从数据驱动角度出发改进后的生成对抗模型考虑了海洋运行环境对数据增强的影响，暂未考虑海上风电机组故障机理对于数据增强的作用及影响，亦未考虑故障特性中不同状态变量之间的耦合关系，这部分内容将在后续进一步开展研究。

参考文献

[1] FRIEDLINGSTEIN P, O'SULLIVAN M, JONES M W, et al. Global carbon budget 2022[J]. Earth System Science Data, 2022, 14(11): 4811-4900.
 [2] 中国能源大数据报告(2022)[R]. 北京: 中能传媒研究院, 2022.
 China energy big data report (2022)[R]. Beijing: China Energy Media Research Institute, 2022
 [3] 刘军, 安柏任, 张维博, 等. 大型风力发电机组健康状态

- 评价综述[J]. 电力系统保护与控制, 2023, 51(1): 176-187.
LIU Jun, AN Bairen, ZHANG Weibo, et al. A review of health condition evaluation of large wind turbines[J]. Power System Protection and Control, 2023, 51(1): 176-187.
- [4] 李特, 王荣喜, 高建民. 基于改进 DBSCAN 的风电机组 SCADA 异常数据识别方法[J]. 西安交通大学学报, 2024, 58(3): 1-10.
LI Te, WANG Rongxi, GAO Jianmin. Improved DBSCAN-based SCADA anomaly data identification method for wind turbines[J]. Journal of Xi'an Jiaotong University, 2024, 58(3): 1-10.
- [5] 胡爱军, 邢磊, 张鑫, 等. 基于时空特征深度融合的双馈式风力发电机故障早期预测[J]. 中国工程机械学报, 2023, 21(4): 353-357.
HU Aijun, XING Lei, ZHANG Xin, et al. Early prediction of doubly-fed wind turbine faults based on deep fusion of spatio-temporal features[J]. China Journal of Construction Machinery, 2023, 21(4): 353-357.
- [6] 赵洪山, 闫西慧, 王桂兰, 等. 应用深度自编码网络和 XGBoost 的风电机组发电机故障诊断[J]. 电力系统自动化, 2019, 43(1): 81-90.
ZHAO Hongshan, YAN Xihui, WANG Guilin, et al. Fault diagnosis of wind turbine generator applying deep self-coding network and XGBoost[J]. Automation of Electric Power Systems, 2019, 43(1): 81-90.
- [7] 郑毅, 王承民, 刘保良, 等. 基于多层次时空图神经网络的风电机组在线异常检测[J]. 电力系统自动化, 2024, 48(5): 107-119.
ZHENG Yi, WANG Chengmin, LIU Baoliang, et al. Online anomaly detection of wind turbines based on multilevel spatio-temporal graph neural network[J]. Automation of Electric Power Systems, 2024, 48(5): 107-119.
- [8] 苏向敬, 山衍浩, 周汶鑫, 等. 基于 GRU 和注意力机制的海上风机齿轮箱状态监测[J]. 电力系统保护与控制, 2021, 49(24): 141-149.
SU Xiangjing, SHAN Yanhao, ZHOU Wenxin, et al. Condition monitoring of offshore wind turbine gearbox based on GRU and attention mechanism[J]. Power System Protection and Control, 2021, 49(24): 141-149.
- [9] 李斌, 宋威, 赵凯, 等. 基于 mRMR-GRU 的风电机组运行状态评估[J]. 可再生能源, 2023, 41(7): 906-911.
LI Bin, SONG Wei, ZHAO Kai, et al. Operational condition assessment of wind turbines based on mRMR-GRU[J]. Renewable Energy Resources, 2023, 41(7): 906-911.
- [10] ZHU L, ZHANG X. Time series data-driven online prognosis of wind turbine faults in presence of SCADA data loss[J]. IEEE Transactions on Sustainable Energy, 2021, 12(2): 1289-1300.
- [11] 张玉彦, 张永奇, 孙春亚, 等. 不平衡样本下基于生成式对抗网络的风机叶片开裂状态识别[J]. 计算机集成制造系统, 2023, 29(2): 532-543.
ZHANG Yuyan, ZHANG Yongqi, SUN Chunya, et al. Cracked wind turbine blade state identification based on generative adversarial network under unbalanced samples[J]. Computer Integrated Manufacturing Systems, 2023, 29(2): 532-543.
- [12] 范敏, 杨青, 郭祥富, 等. 面向不平衡数据的配电网故障停电预测方法[J]. 电力系统保护与控制, 2023, 51(8): 96-106.
FAN Min, YANG Qing, GUO Xiangfu, et al. A fault outage prediction method for unbalanced data in distribution networks[J]. Power System Protection and Control, 2023, 51(8): 96-106.
- [13] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[J]. Advances in Neural Information Processing Systems, 2014: 2672-2680.
- [14] 赵厚翔, 沈晓东, 吕林, 等. 基于 GAN 的负荷数据修复及其在 EV 短期负荷预测中的应用[J]. 电力系统自动化, 2021, 45(16): 143-151.
ZHAO Houxiang, SHEN Xiaodong, LÜ Lin, et al. GAN-based load data restoration and its application in EV short-term load forecasting[J]. Automation of Electric Power Systems, 2021, 45(16): 143-151.
- [15] 朱永利, 张翼, 蔡炜豪, 等. 基于辅助分类-边界平衡生成式对抗网络的局部放电数据增强与多源放电识别[J]. 中国电机工程学报, 2021, 41(14): 5044-5053.
ZHU Yongli, ZHANG Yi, CAI Weihao, et al. Local discharge data enhancement and multi-source discharge identification based on auxiliary classification-boundary balance generative adversarial network[J]. Proceedings of the CSEE, 2021, 41(14): 5044-5053.
- [16] 尹诗, 侯国莲, 胡晓东, 等. 基于 AC-GAN 数据重构的风电机组主轴承温度监测方法[J]. 智能系统学报, 2021, 16(6): 1106-1116.
YIN Shi, HOU Guolian, HU Xiaodong, et al. Wind turbine main bearing temperature monitoring method based on AC-GAN data reconstruction[J]. CAAI Transactions on Intelligent Systems, 2021, 16(6): 1106-1116.
- [17] WILKIE D, GALASSO C. Impact of climate-change scenarios on offshore wind turbine structural performance[J]. Renewable and Sustainable Energy Reviews, 2020, 134: 110323.
- [18] SUN Q, PENG F, YU X, et al. Data augmentation strategy for power inverter fault diagnosis based on Wasserstein distance and auxiliary classification generative adversarial network[J]. Reliability Engineering & System Safety, 2023, 237: 109360.

- [19] SHAO S, WANG P, YAN R. Generative adversarial networks for data augmentation in machine fault diagnosis[J]. *Computers in Industry*, 2019, 106: 85-93.
- [20] 邱吉尔, 陶洪峰, 程龙, 等. 基于自注意力机制辅助分类生成对抗网络的轴承故障诊断[J]. *信息与控制*, 2022, 51(6): 753-762.
QIU Jier, TAO Hongfeng, CHENG Long, et al. Bearing fault diagnosis based on self-attention mechanism assisted classification generative adversarial network[J]. *Information and Control*, 2022, 51(6): 753-762.
- [21] HUANG N, CHEN Q, CAI G, et al. Fault diagnosis of bearing in wind turbine gearbox under actual operating conditions driven by limited data with noise labels[J]. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70: 1-10.
- [22] 杨路春, 杨晨俊, 李学斌. 基于多目标进化算法和决策技术的螺旋桨优化设计研究[J]. *中国造船*, 2019, 60(3): 55-66.
YANG Luchun, YANG Chenjun, LI Xuebin. Research on optimization design of propeller based on multi-objective evolutionary algorithm and decision-making technology[J]. *Shipbuilding of China*, 2019, 60(3): 55-66.
- [23] 钱玉村, 杨博, 郑如意, 等. 基于经验模态分解-灰色关联度分析-蒲公英优化器改进 Elman 网络的锂离子电池健康状态估计[J/OL]. *电网技术*: 1-10[2024-05-03]. DOI: 10.13335/j.1000-3673.pst.2024.0409.
QIAN Yucun, YANG Bo, ZHENG Ruyi, et al. Lithium-ion battery health state estimation based on empirical modal decomposition-gray correlation analysis-dandelion optimizer improved Elman network [J/OL]. *Power System Technology*: 1-10[2024-05-03]. DOI: 10.13335/j.1000-3673.pst.2024.0409.
- [24] 程军照, 徐敏, 冯磊, 等. 基于 GRA-K-means++算法的配电网供电可靠性水平评价[J]. *南方电网技术*, 2023, 17(8): 113-122.
CHENG Junzhao, XU Min, FENG Lei, et al. Evaluation of power supply reliability level of distribution network based on GRA-K-means++ algorithm[J]. *Southern Power System Technology*, 2023, 17(8): 113-122.
- [25] 张杰. 基于 SCADA 数据统计的风电机组异常状态识别[D]. 沈阳: 沈阳工业大学, 2023.
ZHANG Jie. Identification of abnormal state of wind turbine based on SCADA data statistics[D]. Shenyang: Shenyang University of Technology, 2023.
- [26] 万安平, 龚志鹏, 张银龙, 等. 基于 XGBoost-KDE 的风机主轴承温度预测与故障预警方法[J]. *热力发电*, 2022, 51(12): 164-171.
WAN Anping, GONG Zhipeng, ZHANG Yinlong, et al. A wind turbine main bearing temperature prediction and fault warning method based on XGBoost-KDE[J]. *Thermal Power Generation*, 2022, 51(12): 164-171.
- [27] 李俊卿, 胡晓东, 王罗, 等. 考虑数据不足和基于合作博弈模型融合的风电机组轴承故障诊断方法[J]. *太阳能学报*, 2024, 45(1): 234-241.
LI Junqing, HU Xiaodong, WANG Luo, et al. Wind turbine bearing fault diagnosis method considering insufficient data and fusion based on cooperative game model[J]. *Acta Energetica Solaris Sinica*, 2024, 45(1): 234-241.
- [28] 常淼, 沈艳霞. 基于改进卷积神经网络的风电轴承故障诊断策略[J]. *电力系统保护与控制*, 2021, 49(6): 131-137.
CHANG Miao, SHEN Yanxia. Wind power bearing fault diagnosis strategy based on improved convolutional neural network[J]. *Power System Protection and Control*, 2021, 49(6): 131-137.
- [29] 李俊卿, 李忠徽, 仝宗义. 基于支持向量机和 D-S 证据理论的双馈风机定子匝间短路故障诊断[J]. *电机与控制应用*, 2018, 45(5): 99-103.
LI Junqing, LI Zhonghui, TONG Zongyi. Fault diagnosis of stator turn-to-turn short circuit in doubly-fed wind turbine based on support vector machine and D-S evidence theory[J]. *Electric Machines & Control Application*, 2018, 45(5): 99-103.
- [30] CHEN Z, GRYLLIAS K, LI W. Mechanical fault diagnosis using convolutional neural networks and extreme learning machine[J]. *Mechanical Systems and Signal Processing*, 2019, 133: 106272.
- [31] 魏书荣, 张鑫, 符杨, 等. 基于 GRA-LSTM-Stacking 模型的海上双馈风力发电机早期故障预警与诊断[J]. *中国电机工程学报*, 2021, 41(7): 2373-2383.
WEI Shurong, ZHANG Xin, FU Yang, et al. Early fault warning and diagnosis of offshore doubly-fed wind turbine based on GRA-LSTM-Stacking model[J]. *Proceedings of the CSEE*, 2021, 41(7): 2373-2383.

收稿日期: 2024-04-29; 修回日期: 2024-07-24

作者简介:

魏书荣(1980—), 女, 博士, 教授, 博士生导师, 主要从事海上风电优化规划、故障诊断与运行维护研究; E-mail: wsrmail@163.com

殷世杰(1996—), 男, 硕士研究生, 研究方向为海上风力发电系统故障诊断技术。E-mail: 109018300@qq.com

(编辑 姜新丽)