

去噪扩散概率模型在电池数据增强中的应用

常 伟, 胡志超, 潘多昭, 师继文

(南通乐创新能源有限公司, 江苏 南通 226000)

摘要: 解决电池运行过程中关键数据收集困难以及电化学阻抗谱(EIS)数据量有限的问题,能够优化电池性能评估和健康状态监测,以及电池使用和充电策略。研究方法涉及使用数据增强技术,在保障数据质量的同时,增加数据样本量。去噪扩散概率模型(DDPM)作为一种新兴的生成模型,被应用于电池数据的增强;对于低维度的电池数据(如电流、电压、温度和容量),直接应用 DDPM 模型进行数据增强;对于高维度的 EIS 数据,首先采用自编码(AE)模型进行降维,随后在低维空间进行数据增强,并将增强后的数据还原至原始空间。研究结果证实,所提出的数据增强方法在 NASA (美国国家航空和宇宙航行局)和 EIS(电化学阻抗谱)公开数据集上能够生成高质量的数据,并有效降低了计算复杂度。研究成果可为电池性能评估和健康管理提供有效的数据增强策略,具有一定的参考和应用价值。

关键词: 去噪扩散概率模型; 自编码模型; 阻抗谱; 温度; 电流; 电压; 容量; 数据增强

中图分类号: TM912.9 **文献标志码:** A **文章编号:** 1671-1807(2025)09-0092-07

在电池运行过程中,可以收集一些运行数据,这些数据包括但不限于以下几项:① 电池容量。电池的总能量储存能力,单位为毫安时(mA·h)或安时(A·h)。② 电压。电池的电势差,常用单位为伏特(V)。③ 电流。通过电池的电荷流动速率,一般以安培(A)计量。④ 温度。电池的工作温度,过高或过低的温度可能影响电池性能。

电池运行数据可以用于评估电池性能和健康状况,以及优化电池使用和充电策略。通过监测电池容量、电压和充电状态等数据,可以确保电池在正常工作范围内,并及时了解电池是否需要充电或更换。此外,电池的温度数据可以用来预防过热或过冷造成的安全问题。通过分析电池运行数据,还可以优化电池的使用寿命和效率,提高设备的续航时间和可靠性。然而,在某些情况下,电池运行数据的获取较为困难,如极端工况条件下,电池运行数据就很少。此时,可以考虑使用数据增强的方法,在保证质量的前提下,对电池运行数据的数量进行扩充。

电化学阻抗谱(electrochemical impedance spectroscopy, EIS)是一种用于研究电化学系统的分析技术。它通过在系统中施加交流电信号并测

量响应来分析电化学过程中的电阻和电容。这种技术可以被广泛应用于许多领域,包括电池、腐蚀研究、传感器、涂层评估以及生物电化学等。

EIS 基于对电化学系统中的交流电信号响应进行频率扫描来获取信息。在实验中,通过在一定频率范围内施加小振幅的交流电信号,然后测量系统的响应(通常是电流和电压),可以获得电阻、电容和电化学反应动力学等信息。

总的来说,电化学阻抗谱是一种强大的工具,可以为电化学体系提供深入的理解和评估。然而,由于获取 EIS 数据依赖于昂贵的专业设备,导致可用的数据量通常有限,这可能会影响使用 EIS 模型的泛化能力和鲁棒性,并增加过拟合的风险。为了解决这一问题,可以采用数据增强策略来扩充 EIS 数据集,从而为使用 EIS 的模型训练提供更丰富的信息,提高模型在各种情况下的表现稳定性。

数据增强技术是通过变换原始数据或生成新的数据本来增加数据集的多样性和规模。这些技术在深度学习领域,特别是在图像识别、自然语言处理和语音识别等领域中,已成为提高模型性能的关键。数据增强技术是一个快速发展的领域,主要用于解决数据隐私、数据稀缺性和数据偏差等

收稿日期: 2024-09-13

作者简介: 常伟(1976—),男,湖南长沙人,硕士,研究方向为新能源电池算法;通信作者胡志超(1984—),男,湖北黄梅人,硕士,研究方向为电池热失控、EIS、SOH 估计;潘多昭(1991—),男,壮族,新疆阿勒泰人,MBA,研究方向为储能与能源工程、AI 算法、智慧能源;师继文(1990—),男,陕西宝鸡人,硕士,研究方向为新能源材料。

问题。

尽管数据增强技术在提高模型精确度方面发挥了重要作用,但它在数据质量和法律法规方面仍面临挑战。数据增强的使用有助于根据模型的需求生成正确的数据,从而提高模型的可靠性。

常用的数据增强方法有^[1-9]:①生成对抗网络(generative adversarial networks, GANs)。GANs由生成器和判别器组成,通过对抗训练生成逼真的新数据样本。适用于生成复杂的电池运行数据,如电压、温度和电流数据。②变分自编码器(variational autoencoders, VAEs)。VAEs通过学习数据的分布生成新样本,生成器通过编码器-解码器结构实现,适用于生成高质量的电池数据,保留原始数据的统计特性。③数据增强方法集成(data augmentation ensemble)。集成多种数据增强方法,如结合 GANs、VAEs 模型生成多样化的数据集,提高数据增强的多样性和覆盖面。这些方法通过生成新的数据样本或扩展现有数据集,提高了模型训练的多样性和数据量,增强了模型的泛化能力和鲁棒性,能够更好地适应各种工况下的电池性能预测和优化任务。

去噪扩散概率模型(denoising diffusion probabilistic models, DDPM)^[10]是一种基于概率扩散过程的生成模型。它通过逐步添加噪声来破坏数据,然后通过学习一个去噪过程来逐步还原数据。这个模型具有生成高质量样本的能力,并且在某些方面优于其他生成模型如生成对抗网络(GAN)和变分自编码器(VAE)。

DDPM 模型的优点有:①稳定性。相比 GANs,DDPM 的训练过程更加稳定,因为它不涉及对抗训练。②样本质量。相比 VAEs,DDPM 能够生成高质量的样本,尤其在图像生成任务中表现出色。③理论基础。DDPM 具有坚实的理论基础,通过马尔可夫链和概率扩散过程进行建模。

DDPM 模型可以应用于图像生成、图像修复、图像去噪等任务,也可以扩展到其他领域如音频生成和文本生成。通过其高质量的生成能力,DDPM 为各类生成任务提供了一种新的有效方法。

最近几年,以生成式大模型为代表的技术的研究和应用取得显著进展。在图像、音频、视频合成和自然语言处理等领域有着广泛地应用。例如,VAE 具有很强的数据建模能力,其在需要理解数据潜在结构的领域中得到应用,如金融分析和生物信息学。GAN 模型在图像合成和风格迁移等

方面有广泛的应用,被许多科技公司和艺术创作者所采用,如 GAN 被用于生成高质量的合成图像,这些图像在医学影像分析等领域中发挥着重要作用。DDPM 模型作为一种新兴技术,其在图像和视频生成领域的应用正逐渐增加。另外,VAE 和 DDPM 模型在数据去噪和恢复方面表现出色。GAN 和 DDPM 模型在艺术创作、游戏设计、虚拟现实等领域有显著应用。

本文将 DDPM 模型引入到电池数据增强领域,将多维数据的增强方法拓展到一维数据的增强。

本文的创新点如下。

(1)对于电池运行数据进行增强。①高质量合成数据生成。DDPM 能生成高质量的电池运行数据,包括电压、电流和温度等参数,这些合成数据与真实数据在统计特性上高度一致,能够显著扩充数据集的规模,提高模型的泛化能力。②噪声处理与数据修复。DDPM 在处理噪声数据方面表现优异,可以有效去除电池运行数据中的噪声和异常值,生成更为平滑和准确的运行数据,提升数据质量。③逐步细化与多样化生成。采用逐步添加和去除噪声的方法,DDPM 能够生成多样性更高的数据。这意味着可以覆盖不同工况、不同老化状态和不同使用条件下的电池运行数据,丰富数据集的多样性。④处理多模态数据。DDPM 能够同时处理电池的多模态数据,如电压、温度和电流等,生成的多模态数据之间保持相关性和一致性,为综合分析和模型训练提供更加完整的数据支持。⑤增强数据的实用性。生成的高质量电池运行数据可以用于电池健康管理(BMS)系统的模型训练,提升剩余使用寿命预测(RUL)的准确性,提高电池管理系统的性能和可靠性。

(2)对于电池 EIS 数据进行增强。①扩充有限数据集。由于 EIS 数据获取成本高且数据量通常有限,DDPM 可以生成大量高质量的 EIS 数据,有效扩充数据集,减少过拟合风险,提高模型的泛化能力和鲁棒性。②生成具有复杂特征的 EIS 数据。DDPM 可以生成具有复杂频域特征的 EIS 数据,这对于分析和建模电池的电化学特性至关重要。生成的 EIS 数据在特征空间上与真实数据高度一致,提供了丰富的信息支持。③去噪与增强数据质量。对于存在噪声的 EIS 数据,DDPM 能够进行有效的去噪处理,生成更为干净和准确的 EIS 数据,提升数据的可靠性。④多模态数据融合。DDPM 可以将 EIS 数据与其他电池运行数据(如电压、温度、电

流等)进行多模态融合,生成包含多种信息的综合数据,提升对电池性能和健康状态的全面分析能力。⑤提升模型训练效果。通过生成丰富多样的 EIS 数据和其他电池运行数据,可以显著提升基于这些数据的机器学习模型的训练效果,提高模型在不同条件下的预测精度和鲁棒性。⑥创新的建模方法。DDPM 在电池数据增强中的应用,展示了一种新的数据生成与处理方法,拓展了电池数据增强的技术手段,为电池管理和预测模型提供了更为丰富和高质量的数据支持。

这些创新点表明,DDPM 在电池运行数据和 EIS 数据增强中的应用,不仅提高了数据集的质量和多样性,还显著提升了电池管理系统和预测模型的性能,具有重要的研究和应用价值。本文的实用性在于,提供了一种电池数据增强的可行方法。

1 电池数据增强流程

1.1 电池运行数据增强流程

电池运行数据包括电流、电压、温度、容量等数据,由于维度较低,DDPM 的计算复杂度不高,因此可以直接将 DDPM 模型应用于原始数据。

基于 DDPM 模型的电池运行数据增强流程如图 1 所示。

第 1 步,采集电池电压、温度、电流和容量等数据。在真实工况下,电池管理系统(BMS)都会记录电池的电压、温度、电流和容量等数据,按照一定的时间间隔,如 30 s,采集相应的数据并保存到数据库中。

第 2 步,训练 DDPM 模型,使用 Keras 平台,采用 U-Net 框架来训练 DDPM 模型。

第 3 步,对输入数据进行随机采样,并使用 DDPM 模型进行预测,得到新的生成数据,即得到增强数据。

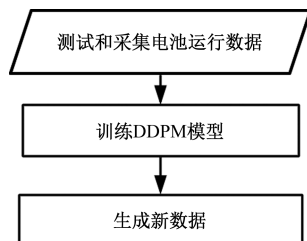


图 1 基于 DDPM 模型的电池运行数据增强流程

1.2 电池 EIS 数据增强流程

电池 EIS 数据覆盖一定的频率范围,由实部数据和虚部数据组成^[11],维度往往较高,如果直接训

练 DDPM 模型,将会非常消耗资源,因此,可以先对 EIS 数据进行降维处理。

基于 DDPM 模型的电池 EIS 数据增强流程如图 2 所示。

第 1 步,测试并采集电池 EIS 数据。使用电化学工作站配置频率扫描测试,在不同频率点测试电池的阻抗值和相位,记录频率响应数据,主要是测试阻抗的实部数据和虚部数据。

第 2 步,训练 AE 模型。基于 Keras 平台进行 AE 模型搭建和训练。

第 3 步,计算 EIS 的潜在空间向量数据。将 EIS 数据作为输入数据,使用 AE 模型的编码器 Encoder 生成潜在空间向量。

第 4 步,训练 DDPM 模型。将 EIS 的潜在空间向量作为 DDPM 模型的输入数据,训练 DDPM 模型。使用 Keras 平台,采用 U-Net 框架来训练 DDPM 模型。

第 5 步,在 EIS 潜在空间生成新数据。将 EIS 潜在空间向量中的数据作为输入数据,使用 DDPM 模型进行预测,生成新的 EIS 潜在向量数据。

第 6 步,还原 EIS 数据。使用 Decoder 解码器将 EIS 潜在空间生成的新数据还原至正常的 EIS 数据,即生成新的 EIS 数据。

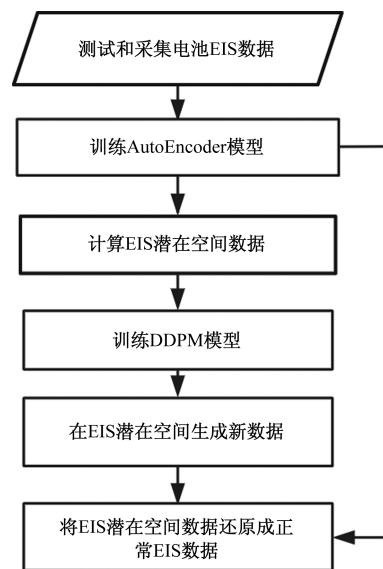


图 2 基于 DDPM 模型的电池 EIS 数据增强流程

2 电池数据增强模型

2.1 电池运行数据增强模型

模型总体结构如图 3 所示。

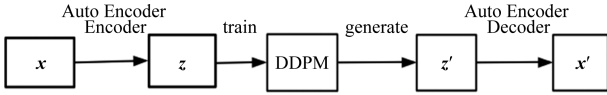
2.2 电池 EIS 数据增强模型

模型总体结构如图 4 所示。



x 为输入数据,有电流、电压、温度和容量等;train 为基于输入数据训练 DDPM 模型;generate 为 DDPM 模型生成新的数据 x'

图 3 DDPM 模型总体结构



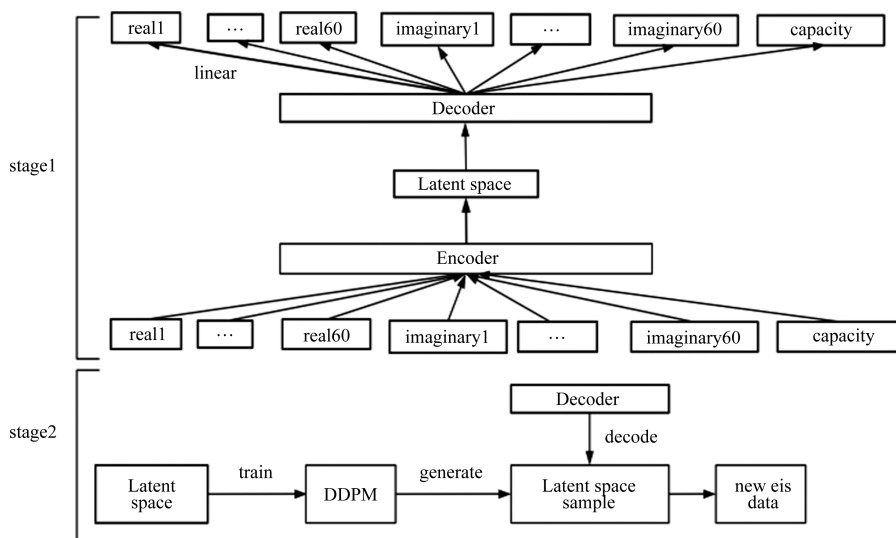
x 为输入数据;Encoder 为编码器; z 为潜在空间向量;train 为基于潜在空间向量训练 DDPM 模型;generate 为 DDPM 模型生成新的潜在空间向量 z' ;Decoder 为解码器,将 z' 还原成新的数据 x'

图 4 AE-DDPM 模型总体结构

数据增强模型结构如图 5 所示,分为 2 个阶段,第 1 阶段为 AE 模型,第 2 阶段为 DDPM 模型和 EIS 数据还原。

第 1 阶段(stage 1):输入部分 real1 至 real60 表示从高频至低频对应的阻抗数据的实部数据,imaginary1 至 imaginary60 表示从高频至低频对应的阻抗数据的虚部数据,capacity 表示电池容量;接下来输入数据连接到一个编码器(Encoder),编码器将输入数据映射到潜在分布 Latent space 中;然后,潜在分布连接到一个解码器(Decoder),用于还原输入数据;之后,待 AE 模型训练完成之后,保存潜在空间的数据和解码器 Decoder,待后续使用。

第 2 阶段(stage 2):将潜在空间的数据作为训练数据集,训练 DDPM 模型。之后,使用 DDPM 模型在潜在空间中生成新的数据;最后,使用 Decoder 将潜在空间的新数据还原成 EIS 数据,即新



Latent space 为潜在空间;train 为训练模型

图 5 用于 EIS 数据增强的 AE-DDPM 模型结构

生成的 EIS 数据。

3 电池数据增强具体步骤

3.1 电池运行数据增强步骤

3.1.1 电池运行数据来源

本文使用公开数据集中的电池运行数据,包括电池的电压、温度、电流和容量等数据。

3.1.2 训练 DDPM 模型

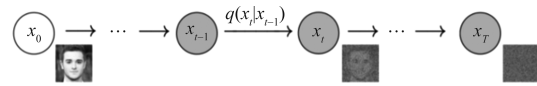
在 Keras 平台使用 U-Net 架构来训练 DDPM 模型。DDPM 模型的训练过程包括以下两个阶段。

(1)噪声化阶段。从数据集中获取真实样本,并在一定的时间步内逐步添加噪声,直到数据完全随机化。DDPM 模型加噪过程示意图如图 6 所示。

加噪过程公式为

$$q(x_t | x_{t-1}, 1) := N(x_t; \sqrt{1 - \beta_t} x_{t-1}, 1, \beta_t \mathbf{I}) \quad (1)$$

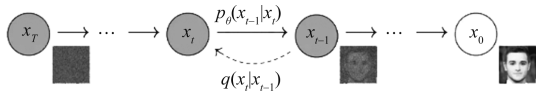
式中, $q(x_t | x_{t-1})$ 为在前向的每一步通过给 x_{t-1} 添加高斯噪声数据得到 x_t ; x_t 为在时间步 t 的数据状态; x_{t-1} 为在时间步 $t-1$ 的数据状态; β_t 为在时间步 t 的噪声比例,通常是一个小于 1 的正数; $N(\cdot | \mu, \sigma^2)$ 为均值为 μ 和方差为 σ^2 的正态分布; \mathbf{I} 为单位矩阵。在 T 步之后, x_T 将是一个接近于纯噪声的状态。



x_0 为真实样本; x_{t-1} , x_t 和 x_T 分别为时刻 $t-1$, t 和 T 时添加噪声之后的样本

图 6 DDPM 模型加噪过程示意图

(2)去噪阶段:训练去噪模型,使其能够预测在任意时间步的噪声,并从噪声数据中恢复出原始数据。DDPM 去噪过程示意图如图 7 所示。



x_T 为完成加噪之后的样本; x_t, x_{t-1} 分别为时刻 $t, t-1$ 时去噪之后的样本; x_0 为去噪完成之后还原的样本

图 7 DDPM 去噪过程示意图

去噪过程公式为

$$p_{\theta}(x_{t-1} | x_t) = N[x_{t-1}; u_{\theta}(x_t, t), \sum_{\theta}(x_t, t)] \quad (2)$$

式中: $p_{\theta}(x_{t-1} | x_t)$ 为在后向的每一步通过给 x_t 去掉高斯噪声数据得到 x_{t-1} ; $\mu_{\theta}(x_t, t)$ 为去噪模型预测的均值,它是关于 x_t 和时间步 t 的函数; $\sum_{\theta}(x_t, t)$ 为去噪模型预测的方差,它也是关于 x_t 和时间步 t 的函数; $N(\cdot | \mu, \sigma^2)$ 为均值为 μ 和方差为 σ^2 的正态分布。

在训练过程中,去噪模型通过最小化重构误差来学习预测准确的均值和方差,使得 x_{t-1} 尽可能接近于原始数据。在生成过程中,去噪模型从 x_T 开始,逐步预测 x_{t-1} ,直到恢复出 x_0 。

3.1.3 生成电池运行数据

在 DDPM 模型训练完毕之后,对原始的电池运行数据进行采样,并使用 DDPM 模型对采样数据进行预测,得到新的电池运行数据,即增强数据。

3.2 电池 EIS 数据增强步骤

3.2.1 电池运行数据来源

本文使用公开数据集中的电池 EIS 数据,包括电池的 EIS 和容量等数据。

3.2.2 构建和训练 AE 模型

AE 是一种无监督学习的神经网络模型,它的主要目的是通过编码器(Encoder)和解码器(Decoder)来学习输入数据的有效表示(编码)。自编码模型的结构通常是对称的,编码器负责将输入数据压缩成低维的编码,而解码器则负责将这个低维编码恢复成原始数据。

(1)编码器(Encoder)。编码器的作用是将输入数据 x 映射到一个低维的潜在空间表示 z 。这个潜在空间通常具有比输入数据更低的维度,因此编码器可以看作是一个降维工具。编码器通常由一系列的线性层和非线性激活函数组成,非线性激活函数提供了模型的学习能力。

(2)潜在空间(Latent Space)。潜在空间 z 是输入数据在低维空间中的表示。在自编码模型中,潜在空间 z 尽量保留输入数据 x 的关键信息,同时去除不必要的噪声。通过这种方式,潜在空间可以用于数据压缩、特征提取或作为其他机器学习模型的输入。

(3)解码器(Decoder)。解码器的作用是将潜在空间 z 映射回原始输入数据的空间,尝试重构出与原始输入 x 尽可能接近的数据 x' 。解码器的结构通常与编码器相似,但是是编码器的镜像反转。

(4)损失函数。自编码模型的损失函数用于衡量重构数据 x' 与原始数据 x 之间的差异。由于 EIS 是连续数值,因此此处的损失函数是均方误差(MSE)。模型的目标是最小化这个损失函数,使得重构数据尽可能地接近原始数据。

在训练时,模型迭代次数为 500,样本批次为 10。

3.2.3 生成 EIS 潜在空间向量数据

将 EIS 数据作为 AE 模型编码器 Encoder 的输入数据,生成相应的潜在空间向量。

3.2.4 生成 EIS 潜在空间向量数据

基于步骤 3 中生成好的 EIS 潜在空间向量数据,训练 DDPM 模型。在 Keras 平台使用 U-Net 架构来训练 DDPM 模型。具体方法参见 3.1.2 节。

3.2.5 生成 EIS 潜在空间新数据

在 DDPM 模型训练完成之后,在 EIS 潜在空间中进行采样,然后使用 DDPM 模型对采样的数据进行预测,得到去噪之后的新生成的 EIS 潜在空间数据。

3.2.6 还原 EIS 数据

在步骤 5 中已经得到了新生成的 EIS 潜在空间数据,使用 AE 模型中的解码器 Decoder 对以上数据进行还原,得到正常的 EIS 数据。

4 预测效果

4.1 电池运行数据增强预测效果

在 NASA 电池数据集中的 B0018 电池的放电(discharge)阶段的数据上测试效果,使用其中的电池电流、电压、温度、电池容量和循环次数等数据来进行数据增强。测试数据特征的名称和含义见表 1。具体数据样例见表 2。基于 DDPM 模型的增强数据样例见表 3。

4.2 生成数据与真实数据相似度分析

通过计算新增数据与真实数据的最大 Cosine 相似度,可以发现大部分的相似度都在 0.98 以上,说明 DDPM 模型能够生成高质量的增强数据。

表 1 电池运行数据特征名称和含义

特征名称	含义	单位
Voltage_measured	测量的电压	V
Current_measured	测量的电流	A
Temperature_measured	测量的温度	°C
Current_charge	在负载下测量的电流	A
Voltage_charge	在负载下测量的电压	V
Capacity/(A·h)	放电至 2.7 V 的电池容量	A·h
id_cycle	循环次数	次

4.3 电池 EIS 数据增强预测效果展示

本文在公开数据集上测试效果,对其中的阻抗数据进行增强。输入数据共有 120 个特征,特征序号依次为 f1~f120,前 60 个特征是频率从高到低对应的阻抗数据的实部数据,后 60 个特征是频率从高到低对应的阻抗数据的虚部数据,Capacity 表示电池容量。

使用 AE 模型对 EIS 数据进行降维,映射到维

度为 20 的潜在空间,潜在空间的向量样例见表 4。

使用 AE 模型对 EIS 数据进行降维,映射到维度为 20 的潜在空间,潜在空间的向量样例见表 5,表中 d1~d20 分别表示潜在空间向量的第 1 维度至第 20 维度。

基于 DDPM 模型生成新的 EIS 潜在向量数据,样例见表 6。

将 EIS 潜在空间新数据还原至原始 EIS 数据,样例见表 7。

通过计算新增数据与原始数据的最大 Cosine 相似度,可以发现大部分的相似度都在 0.98 以上,说明 DDPM 模型能够生成高质量的增强数据。

4.3.1 NASA 数据相似度分析样例

某条 NASA(美国国家航空和宇宙航行局)原始数据见表 8,其生成的数据见表 9。

以上 NASA 真实数据与生成数据的 Cosine 相似度为 0.999 936 5,均方误差为 1.596 5。

表 2 电池运行数据样例

Voltage_measured	Current_measured	Temperature_measured	Current_charge	Voltage_charge	Capacity	id_cycle
4.188 108 651	0.000 130 667	23.819 520 25	0.000 6	0	1.855 004 521	2
4.188 195 943	0.001 459 081	23.828 807 16	0.000 6	4.203	1.855 004 521	2
...
3.469 907 316	0.001 515 875	34.785 943 38	0.000 6	0	1.341 051 441	318
3.477 276 739	-0.001 940 069	34.581 659 54	0.000 6	0	1.341 051 441	318

表 3 电池运行数据增强样例

Voltage_measured	Current_measured	Temperature_measured	Current_charge	Voltage_charge	Capacity	id_cycle
3.125 545 7	-2.023 314 5	38.369 830	1.867 825 9	2.498 738 0	1.752 299 5	39
3.716 371 8	-1.992 242 8	26.710 802	1.995 241 4	2.908 668 3	1.478 731 4	190
3.139 994 0	-2.023 314 5	38.186 714	1.877 815 6	2.532 322 6	1.739 751 5	45
3.336 758 4	-1.880 005 8	32.335 300	1.999 000 1	2.423 853 6	1.363 560 9	301
3.486 401 6	-1.957 776 4	30.444 775	1.999 000 1	2.660 656 2	1.384 201 9	253

表 4 电池 EIS 数据集样例(部分)

f1	f2	...	f119	f120	Capacity
0.384 70	0.391 56	...	0.290 26	0.327 95	37.202 71
0.388 86	0.393 43	...	0.280 64	0.323 00	36.223 03
0.390 38	0.39 60	...	0.287 73	0.329 55	35.589 30
0.391 94	0.396 43	...	0.297 68	0.336 73	35.108 08
0.392 25	0.397 60	...	0.283 96	0.325 62	34.767 28

表 5 电池 EIS 潜在空间数据样例(部分)

d1	d2	...	d19	d20
0.355 173 83	0.563 401 30	...	-0.128 173 65	-0.609 647 3
0.427 724 96	0.502 117 04	...	-0.093 782 78	-0.505 016 9
0.441 013 84	0.432 994 48	...	-0.172 946 16	-0.552 503 1
0.365 985 42	0.498 599 77	...	-0.131 373 17	-0.477 283 0
0.437 506 20	0.485 465 53	...	-0.131 951 51	-0.452 581 7

表 6 DDPM 模型生成的 EIS 潜在空间数据样例(部分)

d1	d2	...	d19	d20
0.355 173 83	0.563 401 3	...	-0.1281 736 5	-0.609 647 3
0.001 926 18	0	...	0	0
0.023 060 71	0	...	0	0
0.361 399 50	1	...	0	0
1	0.545 869 5	...	0	0

表 7 新生成 EIS 数据样例

f1	f2	...	f119	f120	Capacity
0.360 6	0.353 9	...	0.291 9	0.315 2	26.803 7
0.542 7	0.497 8	...	0.278 8	0.297 6	30.859 8
0.561 8	0.567 5	...	0.291 6	0.296 7	34.376 7
0.740 5	0.710 1	...	0.257 7	0.276 3	32.950 7
0.523 4	0.550 8	...	0.276 2	0.274 2	30.890 3

表 8 NASA 真实数据样例

Voltage_measured	Current_measured	Temperature_measured	
3.590 235 45	-2.009 522 69	30.040 395 44	
Current_charge	Voltage_charge	Capacity	id_cycle
1.998 8	2.684	1.780 938 61	41

表 9 NASA 生成数据样例

Voltage_measured	Current_measured	Temperature_measured	
3.465 477 5	-2.003 505 7	32.382 34	
Current_charge	Voltage_charge	Capacity	id_cycle
1.999 000 1	2.601 113 6	1.759 674 1	45

4.3.2 EIS 数据相似度分析样例

某条 EIS 原始数据见表 10, 其生成数据见表 11。

以上 EIS 真实数据与生成数据的 Cosine 相似度为 0.999 238 6, 均方误差为 0.049 02。以上样例都说明 DDPM 模型能够生成高质量的数据。

表 10 EIS 真实数据样例

f1	f2	...	f119	f120
0.762 61	0.769 75	...	0.269 72	0.306 48

表 11 EIS 生成数据样例

f1	f2	...	f119	f120
0.759 07	0.860 05	...	0.304 93	0.330 75

5 结论

利用 DDPM 模型对电池运行数据进行增强。对于低维度的电池数据(如电流、电压、温度和容量), 直接应用 DDPM 模型进行增强。对于高维度的电池数据(如 EIS 数据), 首先使用 AE 对数据进行降维处理, 然后在低维空间进行数据增强, 并将增强数据还原到原始空间。通过在 NASA 和 EIS 公开数据集上的实验验证, 本文提出的方法能够生

成高质量的数据, 有效降低了计算复杂度, 具有重要的参考和应用价值。

参考文献

- [1] 贺伟, 马鸿雁, 张英达, 等. 基于改进门控循环单元神经网络的锂电池组荷电状态预测[J]. 科学技术与工程, 2023, 23(12): 5102-5109.
- [2] 李洁, 张震豪, 董亚冰, 等. 基于生成对抗网络的电动汽车电池数据增强和故障诊断[J]. 汽车技术, 2023(8): 1-6.
- [3] 郑永飞, 文怀兴, 韩昉, 等. 基于电池外特征的粒子群神经网络电池健康状态预测[J]. 科学技术与工程, 2019, 19(36): 184-189.
- [4] 郝可青, 吕志刚, 邸若海, 等. 基于鲸鱼算法优化长短时记忆神经网络的锂电池剩余寿命预测[J]. 科学技术与工程, 2022, 22(29): 12900-12908.
- [5] 常伟, 潘多昭, 胡志超. 一种基于 VAE 模型数据增强算法的电池 RUL 预测方法: 202311608920[P]. 2024-07-02.
- [6] 常伟, 潘多昭, 胡志超. 一种基于 VAE 模型数据增强算法预测电池 SOH 方法: 202311608918[P]. 2024-07-02.
- [7] 周雅夫, 万伟东, 董启超. 基于改进多频正弦激励的燃料电池电化学阻抗谱测量[J]. 科学技术与工程, 2022, 22(22): 9679-9685.
- [8] 李伟, 牛东晓. 基于灰色神经网络的短期电力负荷预测分析[J]. 科技和产业, 2008, 8(10): 57-60.
- [9] 范振杰, 罗娜. 基于改进 VAE 的时间序列数据增强方法[J]. 华东理工大学学报(自然科学版), 2024, 50(3): 400-410.
- [10] JONATHAN H, JAIN A, ABBEEL P. Denoising diffusion probabilistic models[J]. arxiv: 2006. 2020: 11239.
- [11] ZHANG Y, TANG Q, ZHANG Y. et al. Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning[J]. Nat Commun, 2020, 11: 1706.

Application of Denoising Diffusion Probability Model in Battery Data Augmentation

CHANG Wei, HU Zhichao, PAN Duozhao, SHI Jiwen

(Nantong Le Chuang New Energy Co., Ltd., Nantong 226000, Jiangsu, China)

Abstract: Addressing the difficulties in collecting key data during battery operation and the limited amount of electrochemical impedance spectroscopy (EIS) data, is able to optimize battery performance evaluation and health monitoring, as well as optimize battery usage and charging strategies. The research method involves using data augmentation techniques to increase the sample size while ensuring data quality. The denoising diffusion probability model (DDPM), as an emerging generative model, is applied to enhance battery data. For low dimensional battery data such as current, voltage, temperature, and capacity, the DDPM model is directly applied for data augmentation. For high-dimensional EIS data, the autoencoder (AE) model is first used for dimensionality reduction, followed by data augmentation in low dimensional space, and the enhanced data is restored to the original space. The research results confirm that the proposed data augmentation method can generate high-quality data on NASA (National Aeronautics and Space Administration) and EIS public datasets and effectively reduce computational complexity. The conclusion indicates that this study provides an effective data augmentation strategy for battery performance evaluation and health management, and has certain reference and application value.

Keywords: denoising diffusion probability mode; autoencoder; electrochemical impedance spectroscopy; temperature; electric current; voltage; capacity; data augmentation