

基于改进CVAE-GAN的电力系统暂态稳定 评估样本增强方法

马彬喻^{1,2}, 杨军^{1,2}, 彭晓涛^{1,2}, 李蕊^{1,2}, 申锦鹏^{1,2}, 江克证³, 柳丹³, 曹侃³

(1. 交直流智能配电网湖北省工程中心, 湖北 武汉 430072;

2. 武汉大学 电气与自动化学院, 湖北 武汉 430072;

3. 国网湖北省电力有限公司电力科学研究院, 湖北 武汉 430077)

摘要:实际电力系统的暂态失稳样本占比少,不平衡数据降低了数据驱动的暂态稳定评估的失稳样本识别率和可靠性。对此,提出了基于改进条件变分生成对抗网络(CVAE-GAN)的电力系统暂态稳定评估样本增强方法。通过改进输入样本组成比例提高模型对失稳样本分布的学习能力,改进模型网络结构以适应电力系统量测数据特点,采用预训练方式为模型提供良好的初始状态促进训练的收敛。利用训练完成的改进CVAE-GAN模型合成高质量失稳样本,添加到原始样本中实现样本增强。重新训练分类器,实现在线暂态稳定评估。改进的IEEE 39节点系统和改进的南卡罗莱纳州500节点电网测试结果表明,所提方法能够有效学习原始数据分布特性,实现样本增强,从而提升暂态稳定评估精度和失稳样本的识别率。

关键词:数据增强;数据不平衡;条件变分生成对抗网络;暂态稳定评估;电力系统

中图分类号:TM712

文献标志码:A

DOI:10.16081/j.epae.202504023

0 引言

由于电力系统中新能源和电力电子设备的大量接入、电网规模的不断扩大以及运行工况的日趋复杂,电力系统的暂态稳定评估迎来重大挑战^[1]。传统的电力系统暂态稳定评估方法,例如时域仿真法^[2]和直接法^[3],在面对复杂电网时已难以兼顾评估的高精度与高效的速度需求。随着数据驱动技术的发展以及智能电网的数据监测设备同步相量测量单元(phasor measurement unit, PMU)的应用^[4],近年来基于数据驱动的电力系统暂态稳定评估方法成为了学者们青睐的研究方向。例如支持向量机^[5]、随机森林^[6]等浅层学习模型,以及深度学习领域的卷积神经网络^[7]、长短时记忆网络^[8]、堆叠自动编码器^[9]、深度置信网络^[10]、图神经网络^[11]等主流神经网络,可利用大量历史数据训练评估模型构建系统故障后的特征与系统稳定状态之间的映射关系,进而获得预测结果。

在实际电网的运行过程中,其在绝大部分时间呈现稳定状态,这导致从电网获取的大量历史数据中,暂态失稳的样本比例极小,若模型经过训练后将所有的失稳样本预测为稳定样本,则能够获得较高的准确率,但是对失稳样本的预测正确率会严重下

降,而在实际应用中,对电网失稳情况的判断反而更加重要,因此解决数据不平衡问题对于提升暂态稳定评估的失稳样本识别率以及可靠性具有重要意义。目前已有的研究成果主要从算法和数据两方面进行改进。算法层面主要采取代价敏感方法,文献[12]通过焦点损失函数修正轻梯度提升机(light gradient boosting machine, LightGBM)改变样本重叠区域的样本权重以提高失稳样本识别率;文献[13]通过改进焦点损失函数提高失稳样本的拟合度,增加了错分样本的权重来改善总体准确率。然而代价敏感方法改变了失稳样本权重,将造成模型对稳定样本的判别效果下降。在数据层面,随机过采样(random oversampling, ROS)^[14]会造成模型过拟合;合成少数类过采样技术(synthetic minority oversampling technique, SMOTE)^[15]是线性插值法,并不能很好地拟合电力系统真实数据,且通常只用于经典的机器学习模型。文献[16]提出基于改进条件生成对抗网络(conditional generative adversarial network, CGAN)的样本增强方法,通过生成器和判别器的交替训练,生成有效的失稳样本,然而生成对抗网络(generative adversarial network, GAN)容易出现梯度消失造成训练的不稳定,样本分布规律学习能力较弱。

针对以上方法的不足,本文提出了一种基于改进条件变分生成对抗网络(conditional variational autoencoder-generative adversarial network, CVAE-GAN)的电力系统暂态稳定评估样本增强方法。该方法针对条件变分自编码器(conditional variational

收稿日期:2024-07-19;修回日期:2025-02-16

在线出版日期:2025-05-08

基金项目:国家电网有限公司科技项目资助(合同号:5100-202399365A-2-2-ZB)

Project supported by the Science and Technology Project of State Grid Corporation of China(5100-202399365A-2-2-ZB)

autoencoder, CVAE)部分实施预训练,以学习良好的潜在空间表示和重构能力,促进训练的收敛,同时调整输入数据类别结构比例和网络结构,以提高模型的稳定性和生成质量;使用平均特征匹配和成对特征匹配促使模型学习真实数据的分布特性,合成逼真样本。该方法属于非线性数据合成方法,不仅能够避免改变失稳样本权重,且能通过平均特征匹配提高GAN训练的稳定性,在生成样本分布规律方面具备优势,实现更有效的样本增强效果,从而获得更好的暂态稳定评估结果。在改进IEEE 39节点系统和改进美国南卡罗莱纳州500节点电网仿真上验证了该方法的有效性。

1 CVAE-GAN原理

1.1 CVAE-GAN

CVAE-GAN是一种CVAE和GAN相结合的深度学习框架^[17]。传统的GAN结构如附录A图A1所示,其优化目标为:

$$\min_G \max_D \left\{ E_{x \sim P_r} (\ln D(x)) - E_{z \sim P_z} (\ln (1 - D(G(z_{\text{noise}})))) \right\} \quad (1)$$

式中: $E(\cdot)$ 为期望值函数; P_r 和 P_z 分别为真实样本分布和噪声分布; x 为真实样本; z_{noise} 为随机噪声; G 表示生成器; D 表示鉴别器。

CVAE结构如附录A图A2所示,其优化目标为:

$$\min \left\{ -F_{\text{KL}}(q(z|x, y) \| p(z|x)) + E_{q(z|x, y)} (p(y|x, z)) \right\} \quad (2)$$

式中: z 为潜在空间向量; y 为样本类别; $F_{\text{KL}}(\cdot \| \cdot)$ 表示KL散度; $q(z|x, y)$ 为 z 关于输入样本的后验分布; $p(z|x)$ 为 z 的先验分布; $p(y|x, z)$ 为根据潜在空间向量 z 生成的样本。

构造CVAE-GAN模型,结构如图1所示。该模型由编码器网络 E_n 、生成器网络 G 、鉴别器网络 D 、分类器网络 C 组成。

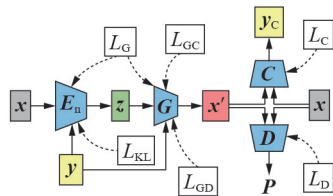


图1 CVAE-GAN结构

Fig.1 Structure of CVAE-GAN

CVAE-GAN模型可表示为:

$$\begin{cases} z = E_n(x, y; \theta_E) \\ x' = G(z, y; \theta_G) \\ P = D(I; \theta_D) \\ y_c = C(x; \theta_C) \end{cases} \quad (3)$$

式中: x 为电网量测真实数据; y 为样本暂态稳定状态的类别标签; x' 为生成数据; I 为 x 或 x' ; P 为属于真实样本的概率; y_c 为分类器输出的类别概率; $E_n(\cdot)$ 、 $G(\cdot)$ 、 $D(\cdot)$ 和 $C(\cdot)$ 分别为编码器、生成器、鉴别器、分类器网络的映射函数; θ_E 、 θ_G 、 θ_D 和 θ_C 分别为编码器、生成器、鉴别器、分类器网络的参数。

1.2 损失函数与网络更新

分类器网络 C 采用交叉熵损失,其损失函数 L_C 可表示为:

$$L_C = -E_{x \sim P_r} (\ln P(y|x)) \quad (4)$$

式中: $P(y|x)$ 为Sigmoid函数转换的电力系统样本分类后验概率。

鉴别器网络 D 试图区分真实训练数据与合成数据,其也采用交叉熵损失,损失函数 L_D 可表示为:

$$L_D = -E_{x \sim P_r} (\ln D(x)) - E_{z \sim P_z} (\ln (1 - D(G(z)))) \quad (5)$$

传统的GAN训练易出现生成器梯度消失或训练不稳定的情况,故采用平均特征匹配目标,即生成样本的特征中心与真实样本的特征中心相匹配,分类器与生成器间亦如此,损失函数可表示为:

$$L_{GD} = \frac{1}{2} \| E_{x \sim P_r} (f_D(x)) - E_{z \sim P_z} (f_D(G(z))) \|^2 \quad (6)$$

$$L_{GC} = \frac{1}{2} \| E_{x \sim P_r} (f_C(x)) - E_{z \sim P_z} (f_C(G(z, y))) \|^2 \quad (7)$$

式中: L_{GD} 、 L_{GC} 分别为鉴别器与生成器间、分类器与生成器间的平均特征匹配目标损失函数; f_D 和 f_C 分别为鉴别器与分类器中间层特征。

编码器网络 E_n 使用KL损失减少先验分布 $P(z)$ 与所提议分布 $P_z \sim N(0, 1)$ 间的差距,KL损失函数 L_{KL} 可表示为:

$$L_{KL} = \frac{1}{2} (\mu^T \mu + F_{\text{sum}}(\exp(\varepsilon) - \varepsilon - 1)) \quad (8)$$

式中: μ 和 ε 分别为潜在空间向量 z 的均值和协方差; F_{sum} 为求和函数。

为提高生成的样本和真实样本分布的相似性,在真实样本 x 和合成样本 x' 之间引入重建损失和成对特征匹配损失 L_G :

$$L_G = \frac{1}{2} (\|x - x'\|_2^2 + \|f_D(x) - f_D(x')\|_2^2 + \|f_C(x) - f_C(x')\|_2^2) \quad (9)$$

CVAE-GAN的总体损失函数 L 可表示为:

$$L = L_D + L_C + \lambda_1 L_{KL} + \lambda_2 L_G + \lambda_3 L_{GD} + \lambda_4 L_{GC} \quad (10)$$

式中: λ_1 、 λ_2 、 λ_3 、 λ_4 分别为 L_{KL} 、 L_G 、 L_{GD} 、 L_{GC} 的权重系数。

CVAE-GAN的四部分网络采用反向传播求梯度来更新网络参数,更新网络参数过程为:

$$\begin{cases} \theta_c \leftarrow \theta_c - \eta \frac{dL_c}{d\theta_c} \\ \theta_d \leftarrow \theta_d - \eta \frac{dL_d}{d\theta_d} \\ \theta_G \leftarrow \theta_G - \eta \frac{d(\lambda_2 L_G + \lambda_3 L_{GD} + \lambda_4 L_{GC})}{d\theta_G} \\ \theta_E \leftarrow \theta_E - \eta \frac{d(\lambda_1 L_{KL} + \lambda_2 L_G)}{d\theta_E} \end{cases} \quad (11)$$

式中: η 为网络的学习率。经过迭代更新,最终计算求得CVAE-GAN的网络参数。

CVAE-GAN能对特征空间维度高且复杂的电力数据进行有效地压缩和保留其重要特征,基于平均特征匹配特性和成对特征匹配充分学习失稳样本的分布特性,从而确保了生成的失稳样本具备多样性且更符合真实暂态失稳样本的全局分布规律,故将该模型用于电力系统的暂态稳定评估样本增强。

2 基于改进CVAE-GAN的暂态稳定评估样本增强模型

2.1 评估模型改进

在构建暂态稳定评估样本增强模型时,对CVAE-GAN模型进行了改进。原始CVAE-GAN模型通常用于图像处理,但直接应用在电力系统暂态稳定评估的数据不平衡问题场景时存在一定的局限性。因此,针对电力系统暂态稳定评估的应用场景,从输入样本类别构成比例、模型神经网络结构、模型训练方式3个方面对原始CVAE-GAN模型进行改进,形成基于改进CVAE-GAN的暂态稳定评估样本增强模型。

2.1.1 样本特征构造

对于在某一工况下能够稳定运行的特定电力系统,其稳态运行参数应是确定的,当系统发生故障时,利用神经网络建立系统暂态稳定状态与系统故障后动态特性信息的映射关系,若电力系统在某一故障后的暂态过程中不再次发生故障,则该故障切除后系统的运行状态将完全由系统内各元件的物理特性决定,而采用故障切除前的数据无法反映故障期间的变化,因此选择故障切除瞬间与暂态稳定性强相关的电气量作为神经网络模型的输入^[18]。本文选择故障切除瞬间的发电机有功功率和无功功率、线路电压和相角、线路有功功率和无功功率、负荷有功功率和无功功率作为输入特征。

在电力系统暂态稳定评估的应用场景中,能够获取到的暂态失稳样本数量仅占总体样本的少数,但失稳样本比稳定样本更为关键,直接采用不平衡样本训练样本增强模型,会使样本增强模型学习到的训练样本分布情况以数量占大多数的暂态稳定样本为主,而对占比较少的暂态失稳样本分布学习不

强,甚至会将其忽略。因此,为了增强模型对暂态失稳样本的学习能力,对输入样本类别构成比例进行了改进。具体而言,在将训练样本输入模型前,随机剔除部分暂态稳定样本,使2类样本输入模型的数量一致,从而改善暂态失稳样本的生成效果,提高学习暂态失稳样本的特征和分布的能力。

2.1.2 网络结构

原始CVAE-GAN模型多用于图像处理,其卷积层结构适合捕捉图像数据中的局部模式,因为图像的相邻像素间具备较强的关系,但是在本文的电力系统暂态稳定评估场景下,所采用的电气特征数据并不具备图像的特点,每个训练样本由多电气特征组成,相邻的电气特征针对系统暂态稳定并不一定强相关,使用卷积层的滑动窗口操作,考虑局部的窗口内特征,导致生成的暂态失稳样本的分布仅与真实样本的局部分布特性相近。因此,对原始CVAE-GAN的结构进行改进,以Dense层构造的隐藏层替代卷积层,Dense层以固定长度输入,考虑训练样本的多电气特征的共同作用及全局的数据关系,更好地学习高维电气特征之间的复杂非线性关系,从而建立电力系统多个电气特征与暂态稳定状态的映射关系。网络结构的改进方面,增强了模型对输入样本电气特征与系统暂态稳定状态对应关系的学习能力,因而能使模型生成的暂态失稳样本更符合真实暂态失稳样本的全局分布关系。改进后的CVAE-GAN结构如附录B表B1所示,各部分神经网络由多层隐藏层组成。

表B1中:编码器和生成器的嵌入层和乘法层的作用是将类别标签与对应样本向量或潜在空间向量融合后一同输入神经网络,为训练提供标签信息;隐藏层(m)表示该隐藏层的神经元数量为 m ;激活函数均为LeakyReLU函数。在每个隐藏层后加入了批规范化(batch normalization)层,其目的是加快梯度下降的收敛速度、加速模型训练、减少过拟合。编码器的全连接层激活函数为Linear函数;生成器的全连接层输出维度始终保持与输入模型的真实样本的特征维度一致,其激活函数为Tanh函数;鉴别器和分类器的全连接层均为1层,激活函数均为Sigmoid函数。各网络选用Adam优化器对网络参数进行计算更新,其具备自适应学习率的特点,使网络能更好地收敛到最优解。

2.1.3 CVAE-GAN预训练

电力系统暂态稳定评估模型在处理高维、复杂的电力系统数据时,训练过程中可能难以建立高维特征与暂态稳定状态间的关系,导致模型陷入局部最优,模型的学习效率和效果不理想。为了进一步提升样本增强模型训练的收敛效果以及模型的性能,本文对CVAE-GAN的训练方式进行了改进,新增

了CVAE-GAN模型的预训练步骤,其主要针对的是模型前段的CVAE预训练。编码器以真实的电力系统样本和对应的暂态稳定条件信息作为输入,提前捕捉输入数据的重要特征,生成器则通过最小化重构损失来学习真实样本分布,改善生成样本的质量。预训练的目标函数的显式表达为:

$$\min \frac{1}{N_s} \sum_{i=1}^{N_s} \left[q(z_i|y_i) \ln \frac{q(z_i|y_i)}{p(z)} + \frac{1}{2} \|x_i - x'_i\|_2^2 \right] \quad (12)$$

式中: z_i 和 y_i 分别为第 i 个样本的潜在空间向量和类别标签; $q(z_i|y_i)$ 为编码器网络 E_n 的实际输出分布; $p(z)$ 为目标分布,采用单位高斯分布; N_s 为样本数量。

由于暂态稳定评估样本增强模型输入的高维电气特征较为复杂,直接训练模型以建立特征与暂态稳定间的关系难度较大,而通过预训练CVAE这一改进,为整体的CVAE-GAN模型提供了良好的初始状态,降低了模型的训练难度,并且能够提前捕捉重要电气特征以及学习原始样本分布特征,因此模型在后续的整体训练中更容易收敛,避免局部陷入最优,从而建立更准确的高维电气特征与暂态稳定状态间的映射关系,提升生成样本的质量。

2.2 数据预处理与模型评价指标

神经网络在训练前通常需要对输入的数据进行归一化处理,如附录B式(B1)所示。

为衡量样本不平衡情形下的模型性能,引入混淆矩阵,如附录B表B2所示,其将所有实例划分为 N_{TS} 、 N_{FS} 、 N_{FU} 、 N_{TU} 这4组,并通过总体准确率 φ_{Acc} 、稳定样本被正确分类的比例 φ_{SR} 、失稳样本被正确分类的比例 φ_{UR} 以及 φ_{SR} 和 φ_{UR} 的几何平均值 φ_{G-mean} 这4个指标反映模型评估的性能,具体表达式见附录B式(B2)~(B5)。

2.3 基于样本增强的暂态稳定评估流程

为了找出能够对原始真实数据增强效果最佳的合成数据,本文提出了基于改进CVAE-GAN的电力系统暂态稳定样本增强框架,如附录B图B1所示,其由数据预处理、改进CVAE-GAN模型训练、基于平衡数据集暂态稳定评估模型构建、在线评估四部分构成,具体流程如下。

1)数据预处理:根据历史运行数据和时域仿真数据构建原始样本数据集,获取故障切除时刻数据;本文仅针对电力系统暂态功角稳定进行研究,对样本类别进行划分,依据电力系统安全稳定导则判断系统暂态功角稳定状态,故障发生后电力系统各机组之间的功角相对增大,若经过第1个或第2个振荡周期后不失步,则视为暂态稳定,若任意2台发电机间的最大功角差摆开,则视为暂态失稳^[19],对样本稳

定状态进行标注,然后进行数据归一化处理,将数据集分为训练集和测试集。

2)改进CVAE-GAN模型训练:首先对改进CVAE-GAN模型中CVAE部分的网络进行预训练,使改进CVAE-GAN模型的参数具备良好的初始状态,预训练完成后迭代训练整体的改进CVAE-GAN模型,计算迭代过程中各部分网络损失函数的值,对分类器网络 C 、鉴别器网络 D 、生成器网络 G 、编码器网络 E_n 四部分依次采用反向传播求梯度来更新网络参数,使生成器网络 G 不断学习真实样本的特征空间分布特性,为合成样本构建平衡数据集提供可用模型。

3)基于平衡数据集的暂态稳定评估模型构建:生成多组随机噪声向量 z_p ,其服从标准正态分布,利用步骤2)中已训练完毕的改进CVAE-GAN模型中的生成器网络 G 合成多组暂态失稳样本,利用分类器网络 C 对每组暂态失稳样本进行评估,筛选出总体准确率最高的1组失稳样本,将其添加到原始样本中,保证最终样本中失稳样本数量与稳定样本数量一致,从而获得平衡数据集,实现样本增强;为了使暂态稳定评估模型更全面地学习平衡后数据集的分布特性,提高失稳样本的识别效果,使用平衡数据集重新单独训练分类器网络 C ,使评估模型更全面地学习真实数据集的原始概率分布,获得基于数据增强的电力系统暂态稳定评估分类器模型。

4)在线评估:电力系统运行过程中,利用广域测量系统采集系统电气量,获取在线数据后,输入上述已训练完成的电力系统暂态稳定评估模型中,快速得到在线评估预测结果。

3 仿真与结果分析

为了验证本文所提暂态稳定评估样本增强方法的有效性,基于编程语言Python的深度学习框架Tensorflow2.1和Keras搭建改进CVAE-GAN模型,仿真计算机配置为Intel® Core (TM) i5-8265U CPU、16.0 GB RAM。

3.1 数据集生成

选取的测试系统为改进IEEE 39节点系统。考虑到可再生能源的影响,改进IEEE 39节点系统整合了2座风电场,对母线34、36上的同步发电机进行了等容量替换,风机类型为双馈风机,风电的渗透率约为20.7%,系统拓扑结构见附录C图C1。采用可实现机电暂态仿真的PSS/E软件进行时域仿真,系统中所有负荷均采用ZIP综合负荷模型,选取同步发电机各组件模型,发电机模型为GENTPJU1,励磁系统模型为IEEEX1,调速器模型为IEESGO;风机由4个组件组成,其中发电机模型为WT3G,风力涡轮机模型为WT3T,变桨控制模型为WT3P,变流控制

模型为WT3E,采用恒功率因数控制。

为使离线仿真能够覆盖实际应用中的不同运行水平的系统状态,分别设置70%、75%、…、120%共11种负荷水平,同步发电机出力进行相应的调整使系统功率平衡;故障方式设置为最严重的三相接地故障;对不同负荷水平下的电力系统进行潮流计算,若满足潮流收敛,则设置故障地点分别位于每条交流线路总长的2%、20%、40%、60%、80%处,仅考虑 $N-1$ 故障;故障的切除时间设置为0.1、0.3、0.5 s,仿真时长设置为20 s。仿真一共得到5 610个不同运行条件下的有效样本,确保样本包含改进IEEE 39节点系统在各典型工况下的 $N-1$ 故障情况,然后随机剔除部分失稳样本以模拟实际电力系统失稳样本稀少的情况,使得最终总样本数量和失稳样本数量的比例为20:1^[20]。随机选取其中70%的样本作为训练集,剩余30%的样本作为测试集。

3.2 改进CVAE-GAN的训练效果

由于改进CVAE-GAN模型的各网络结构采用多层神经网络进行搭建,因此使用小批量训练的方式能使训练趋于收敛,达到平坦最小化,增强模型的泛化性能。模型的部分重要参数设置如下:模型小批量训练的批大小(batch size)设置为91,训练的迭代次数设置为1 000次,CVAE预训练的迭代次数设置为100次;潜在空间向量 z 的维度设置为100;损失函数权重系数设置为 $\lambda_1=3$ 、 $\lambda_2=\lambda_3=\lambda_4=1$;设置Adam优化器的初始学习率 $\eta=0.005$ 。模型的各神经网络层数如表B1所示。

由式(9)可知,生成器网络 G 的损失函数采用2范数构建真实样本和生成样本的重建损失,反映二者分布的相似程度。改进前、后CVAE-GAN的生成器网络损失值变化如图2所示。图中:改进前CVAE-GAN的生成器网络损失在训练前期的波动较大,且训练稳定后,其损失值始终大于改进后CVAE-GAN的生成器网络损失值,这是由于改进后CVAE-GAN的预训练环节能够提供给模型更好的初始状态,使模型训练更加稳定,同时说明改进后的CVAE-GAN的生成样本更接近真实样本分布。

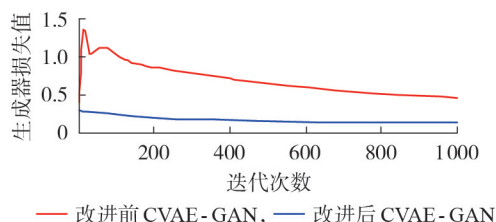


图2 改进前、后CVAE-GAN的生成器网络损失值变化

Fig.2 Loss value changes of generator network of CVAE-GAN before and after improvement

为了分析改进后CVAE-GAN在完成训练后生成样本的效果,使用t分布随机近邻嵌入(t-distributed

stochastic neighbor embedding, t-SNE)工具将模型生成的失稳样本非线性降维到2维平面^[20-21],生成的失稳样本数量与原始训练数据中的失稳样本数量一致,对比原始数据、改进前CVAE-GAN生成的失稳样本以及改进后CVAE-GAN生成的失稳样本在2维空间的分布情况如附录C图C2所示。

改进前CVAE-GAN模型生成的失稳样本(浅蓝色样本)仅集中在图C2中椭圆内的区域,是原始训练数据的失稳样本(深蓝色样本点)较为集中的区域,而改进后CVAE-GAN生成的失稳样本(红色样本点)更接近原始数据的失稳样本,改进后CVAE-GAN仍保持原本平均特征匹配和成对特征匹配这2个特性,使生成样本的整体分布与真实样本相似。改进前CVAE-GAN对电力系统数据的学习能力有限,限制了其在本文中包含多个电气特征的数据上的表达能力,不能全面学习到所有失稳样本的分布特性,而改进后的CVAE-GAN不仅能够学习到集中区域的失稳样本分布,还能够学习到远离集中区域的一些失稳样本点分布,说明改进后预训练的加入为生成器提供了更好的初始状态,从而增强了其样本生成和重构能力;网络结构的改变也增强了模型对包含多个电气特征的电力系统数据的全局关系学习能力,因此改进后CVAE-GAN保持了原本学习整体分布的特性,并提高了样本分布学习能力,其生成样本的分布更接近真实样本。

3.3 数据增强方法的性能测试

在采用改进CVAE-GAN生成数据阶段,利用生成器网络 G 生成200组失稳样本,利用分类器网络 C 进行评估,各组失稳样本的评估准确率如附录C图C3所示,并筛选出准确率最高的1组失稳样本与原始数据融合,使得到的数据集中失稳样本数量与稳定样本数量一致,即获得平衡数据集。

为了体现本文所提方法的优越性,利用无数据增强方法时的不平衡数据集和有数据增强方法时获得的平衡数据集分别训练暂态稳定分类器,对比这2种分类器的分类性能,其中,数据增强方法包括ROS、SMOTE这些传统的数据增强方法以及文献[16]所提改进CGAN方法和本文所提方法;暂态稳定评估分类器均采用表B1中的分类器网络 C 。各方法对应的分类器在同一数据测试集中的评估结果如表1所示,具体分析如下。

1)由于原始训练数据中失稳样本数量极少,因此利用未经数据增强方法处理的原始训练数据训练的暂态稳定评估模型有明显将每个样本判定为稳定样本的趋势,从而获得了较高的稳定样本被正确分类的比例 φ_{SR} 和总体准确率 φ_{Acc} ,造成失稳样本被正确分类的比例 φ_{UR} 很低,仅有84.42%。

2)采用数据增强方法后,用平衡数据集训练后

表1 各种数据增强方法性能对比

Table 1 Performance comparison among various data augment methods

数据增强方法	$\varphi_{SR} / \%$	$\varphi_{UR} / \%$	$\varphi_{G-mean} / \%$	$\varphi_{Acc} / \%$
无数据增强	99.80	84.42	91.78	99.03
ROS	99.93	92.21	95.99	99.55
SMOTE	99.25	94.81	97.00	99.03
改进CGAN	99.32	96.10	97.70	99.16
改进CVAE-GAN	99.80	98.70	99.25	99.74

的分类模型的评估效果都得到了不同程度的提升。具体而言,传统的数据增强方法ROS以及SMOTE表现较差,改进CGAN方法的 φ_{UR} 和 φ_{G-mean} 得到了进一步的提升,本文方法的 φ_{UR} 、 φ_{G-mean} 和 φ_{Acc} 均最高;与采用原始训练数据的结果相比,本文方法的 φ_{UR} 提升了14.28%, φ_{G-mean} 提升了7.47%,这说明本文方法促使评估模型学习到更多失稳样本的信息,使评估模型对2类样本在特征空间上的分类边界更为明显,部分临界的失稳样本能远离边界,提高评估模型对失稳样本的识别率,因此本文方法对失稳样本识别的可靠性最好,暂态稳定评估的准确性最高。

为了对比各种数据增强方法合成数据的质量,观察t-SNE降维后的数据分布情况,如附录C图C4所示。由于各方法均向原始训练样本中添加了各自生成的失稳样本,故各方法获得的平衡数据集的整体概率分布各不相同,而t-SNE降维使降维前的高维空间中点对之间的概率分布与降维后低维空间中点对之间的概率分布保持一致,因此图C4中各方法的稳定和失稳数据分布并不相同。但基于t-SNE降维使相似的样本尽量靠近,不相似的样本尽量远离的特性,易于观察各方法生成的失稳样本与真实失稳样本的相似情况。

已有的数据增强方法存在一定不足:图C4(a)中,ROS只是通过简单复制已有失稳样本来增强数据,样本大量重合容易造成过拟合问题。图C4(b)中,采用的SMOTE是一种线性插值方法,合成的失稳样本总是在失稳样本的线性连接上,其拟合效果与真实电力系统的数据分布会有较大差别;图C4(c)中,改进CGAN方法合成的失稳样本分布能够区别于真实的稳定样本,但对于失稳样本分布的学习能力仍较弱。本文的改进CVAE-GAN属于非线性数据合成方法,为体现其生成样本的真实有效性,从样本分布和评估性能两方面说明。

1)从生成样本的分布来看,在图C4(d)中的椭圆区域内,增强数据符合真实失稳样本(蓝色样本点)的分布,同时在原本有限的失稳场景下,增加了更多符合失稳分布规律的样本(红色样本点),以此不断逼近原始训练数据集中没有的真实失稳数据,丰富了原始数据中包含的失稳信息。

2)从暂态稳定分类器的评估性能来看,对比表1中数据增强前和采用改进CVAE-GAN进行数据增强后的结果,分类器对测试集失稳样本的识别率 φ_{UR} 从84.42%提高到了98.70%,说明分类器从新生成的失稳样本中学习到了更多工况下的失稳信息,进一步降低了失稳样本的漏判。

因此,从样本分布和评估性能分析来看,改进CVAE-GAN都能适用于电力系统暂态稳定评估的样本增强。由于改进CVAE-GAN具备平均特征匹配和成对特征匹配的特性,提高了生成样本与真实样本的相似程度,而改进网络结构提升了模型对多维电气特征的学习能力,加入的预训练也使得模型在整体训练前捕捉到重要电气特征,学习原始样本的分布特征,从而提高了改进CVAE-GAN的模型性能,使其更全面地学习到真实失稳样本的失稳信息,生成的失稳样本可进一步提高暂态稳定评估模型的失稳样本识别率。

3.4 模型对噪声的鲁棒性

同步PMU在采集数据时可能存在测量误差,使含噪声干扰的数据对暂态稳定评估结果产生影响。对于符合IEEE C37.118标准的同步PMU,其测量总矢量误差应小于1%^[22],表达式如附录C式(C1)所示。

根据信噪比(signal to noise ratio, SNR)的定义,20 dB的高斯白噪声意味着存在1%的测量误差,故本节通过加入20 dB的高斯白噪声来模拟同步PMU测得的最大误差下的电网数据,以检验所提模型对噪声干扰的鲁棒性。噪声影响下,各数据增强方法的测试效果如表2所示。

表2 广域噪声下各种数据增强方法的性能对比

Table 2 Performance comparison among various data augment methods under wide-area noise

数据增强方法	$\varphi_{SR} / \%$	$\varphi_{UR} / \%$	$\varphi_{G-mean} / \%$	$\varphi_{Acc} / \%$
无数据增强	99.86	89.61	94.60	99.36
ROS	99.12	94.81	96.94	98.90
SMOTE	98.30	94.81	96.54	98.13
改进CGAN	98.78	93.51	96.11	98.52
改进CVAE-GAN	99.86	96.10	97.97	99.68

对表2的分析如下。

1)未采用无数据增强方法时,表2中的准确率高于表1中无噪声干扰情况下的准确率,这可能是由于20 dB噪声是随机添加的,使原始训练样本中部分类别临界的样本被广域噪声干扰从而远离了分类边界,因此评估模型更准确。

2)然而,噪声干扰也可能使原本远离分类边界的样本靠近分类边界,增加模型的判别难度,导致4种数据增强方法在噪声干扰下的总体准确率 φ_{Acc} 均有所下降。其中,ROS法大量重复生成失稳样本,

分类器网络不会过分关注重复样本,因此其对噪声的敏感程度较低, φ_{UR} 反而有所提升;SMOTE这一线性插值方法容易在特征空间中的受噪声干扰样本和正常样本间合成与真实样本不相符的异常样本,存在过拟合问题,因此其稳定样本被正确分类的比例 φ_{SR} 也有所下降;改进CGAN的 φ_{UR} 低于SMOTE,其对噪声的抗扰能力仍较弱;本文所提改进CVAE-GAN的 φ_{SR} 、 φ_{UR} 、 φ_{G-mean} 和 φ_{Acc} 的值在所有数据增强方法中均最高,表明改进CVAE-GAN合成的数据受噪声干扰较小,对原始数据特征分布特性的学习能力最强,即使存在20 dB的噪声,仍能发掘到原始数据的特性,合成高质量的电力系统失稳数据,不至于发生过拟合,影响稳定样本的识别率,因此其对噪声的鲁棒性较强,有效提高了数据不平衡情形下暂态稳定评估的可靠性。

3.5 模型在N-2场景下的测试效果

为验证模型在某些N-2场景下的泛化性能,设置母线30上的同步发电机处于检修状态暂时脱网以及连接母线17、18的交流线路检修暂时脱网情况下线路发生三相接地短路故障(即N-2故障),负荷水平、发电机出力、故障地点及故障时间与3.1节中的仿真设置相同,从生成的仿真样本中随机各取500个样本组成一个新的测试集以测试评估模型的效果,测试效果如表3所示。

表3 模型在N-2场景下的测试效果

Table 3 Test performance of model in N-2 scenarios

数据增强方法	$\varphi_{SR} / \%$	$\varphi_{UR} / \%$	$\varphi_{G-mean} / \%$	$\varphi_{Acc} / \%$
无数据增强	99.55	79.25	88.82	97.40
改进CVAE-GAN	98.55	96.23	97.38	98.30

在表3中,评估模型的总体准确率 φ_{Acc} 相比N-1测试场景下有所下降,因为系统脱离了原来的运行场景,但依然能够有效提升失稳样本被正确分类的比例 φ_{UR} ,说明本文所提方法能够学习到未知故障场景下的失稳情形,生成符合真实失稳样本规律的样本,不仅适用于固定拓扑,对于拓扑变化不大的N-1、N-k等故障场景也具有较好的泛化性能。

3.6 模型的泛化性

为证明模型的泛化性能,测试其在大电网中的数据增强能力,利用改进的美国南卡罗来纳州500节点电网对本文所提改进CVAE-GAN进行性能测试。该电网原本包含56台同步发电机以及500条母线,正常运行时共有466条交流线路,其单线图如附录C图C5所示。改进的美国南卡罗来纳州500节点电网将其中的7台同步发电机等容量替换为双馈风机,风电渗透率约为20.3%。

针对该系统的考虑70%、80%、90%、100%、110%这5种负荷水平;分别在每条交流线路总长的

2%、20%、40%、60%、80%处设置三相接地故障,仅考虑N-1故障;故障的切除时间设置为0.1、0.3、0.5 s。最终共得到34 950个不同运行状态下的样本,其中暂态稳定样本数量为34 055个,失稳样本数量为895个,随机选择70%的样本作为训练集,其他样本作为测试集。由此可见,在稳定且坚固的大电网情景下,暂态失稳样本占总体样本的比例仅为2.56%左右,样本不平衡的问题更加突出。同时,模型的输入特征同样选择发电机有功功率和无功功率、线路电压和相角、线路有功功率和无功功率、负荷有功功率和无功功率,共7 856维特征。改进CVAE-GAN的模型参数与3.2节中的设置参数一致,在高维数据且2类样本数量极不平衡的大规模电网下,本文所提改进CVAE-GAN与其他数据增强方法的测试结果对比如表4所示。

表4 各种数据增强方法在高维数据下的性能对比

Table 4 Performance comparison among various data augment methods under high dimensional data

数据增强方法	$\varphi_{SR} / \%$	$\varphi_{UR} / \%$	$\varphi_{G-mean} / \%$	$\varphi_{Acc} / \%$
无数据增强	99.98	73.93	85.97	99.28
ROS	99.76	76.79	87.52	99.15
SMOTE	99.82	87.86	93.65	99.50
改进CGAN	99.68	90.36	94.90	99.43
改进CVAE-GAN	99.79	96.43	98.10	99.70

对表4的分析如下。

1)大规模电网中极少的失稳样本使得分类器网络倾向于将所有样本归类为稳定样本,致使评估结果有较高的稳定样本召回率和总体准确率,但分类器网络的失稳样本召回率却很低。

2)ROS这一重复合成失稳样本方法对于高维大电网的样本增强效果很小,SMOTE和改进CGAN方法使失稳样本识别效果有了较大的提高,但仍低于本文所提改进CVAE-GAN方法。本文方法的 φ_{UR} 、 φ_{G-mean} 和 φ_{Acc} 值均高于其他数据合成方法,表明本文方法对高维数据的学习能力强,即使在失稳样本数量极少且特征数量多的情况下,其仍能准确学习到真实失稳样本的分布特性,大幅提高了分类器网络对稳定样本的判别效果和整体的评估性能,在大型的电力系统中仍具有较强的适应性和较好的泛化性。

4 结论

针对实际电力系统失稳样本数量稀少的问题,本文提出了一种基于改进CVAE-GAN的电力系统暂态稳定评估样本增强方法,该方法能够有效学习原始数据的分布特性合成高质量失稳样本,平衡原始数据集,提高暂态稳定评估的性能。在改进的含风机IEEE 39节点系统和美国南卡罗来纳州500节

点电网上进行了仿真测试,得出结论如下。

1)改进CVAE-GAN能更好地适用于电力系统多维电气特征的样本,其增强了特征与系统暂态稳定状态关系的学习能力,更好地学习到原始失稳样本的分布特性,合成更接近真实样本的高质量数据集。

2)改进CVAE-GAN得到的电力系统增强数据改善了暂态稳定评估失稳样本的识别率,其 φ_{SR} 、 φ_{UR} 、 φ_{G-mean} 和 φ_{Acc} 的值均优于ROS(简单合成)、SMOTE(线性插值合成)这些传统的数据增强方法和已有的改进CGAN(非线性合成)方法。

3)改进CVAE-GAN方法具有较好的抗噪声干扰能力和处理高维数据的泛化能力,因此其更适合用于实际大型电力系统的暂态稳定评估。

附录见本刊网络版(<http://www.epae.cn>)。

参考文献:

- [1] 胡力涛,王怀远,党然,等. 基于梯度范数的暂态稳定评估模型的不平衡修正方法[J]. 电力自动化设备, 2024, 44(4): 158-163, 177.
HU Litao, WANG Huaiyuan, DANG Ran, et al. Imbalance correction method of transient stability assessment model based on gradient norm[J]. Electric Power Automation Equipment, 2024, 44(4): 158-163, 177.
- [2] TANG C K, GRAHAM C E. Transient stability index from conventional time domain simulation[J]. IEEE Transactions on Power Systems, 1994, 9(3): 1524-1530.
- [3] 朱劭璇,王彤,王增平,等. 考虑主导不稳定平衡点变化的电力系统暂态稳定切机控制策略[J]. 电力系统保护与控制, 2021, 49(5): 20-28.
ZHU Shaoxuan, WANG Tong, WANG Zengping, et al. Generator tripping strategy in transient stability control of a power system considering the change of the controlling unstable equilibrium point[J]. Power System Protection and Control, 2021, 49(5): 20-28.
- [4] AHSAN F, DANA N H, SARKER S K, et al. Data-driven next-generation smart grid towards sustainable energy evolution: techniques and technology review[J]. Protection and Control of Modern Power Systems, 2023, 8(3): 1-42.
- [5] 姜涛,王长江,陈厚合,等. 基于正则化投影孪生支持向量机的电力系统暂态稳定评估[J]. 电力系统自动化, 2019, 43(1): 141-148.
JIANG Tao, WANG Changjiang, CHEN Houhe, et al. Transient stability assessment of power system based on projection twin support vector machine with regularization[J]. Automation of Electric Power Systems, 2019, 43(1): 141-148.
- [6] 武宇翔,韩肖清,牛哲文,等. 基于变权重随机森林的暂态稳定评估方法及其可解释性分析[J]. 电力系统自动化, 2023, 47(14): 93-104.
WU Yuxiang, HAN Xiaoqing, NIU Zhewen, et al. Transient stability assessment method based on variable weight random forest and its interpretability analysis[J]. Automation of Electric Power Systems, 2023, 47(14): 93-104.
- [7] 时纯,刘君,梁卓航,等. 基于GAN和多通道CNN的电力系统暂态稳定评估[J]. 电网技术, 2022, 46(8): 3191-3202.
SHI Chun, LIU Jun, LIANG Zhuohang, et al. Transient stability assessment of power system based on GAN and multi-channel CNN[J]. Power System Technology, 2022, 46(8): 3191-3202.
- [8] 李楠,朱娜,崔莹. 考虑代价敏感的AC-LSTM暂态稳定评估[J]. 电力系统保护与控制, 2022, 50(22): 160-169.
LI Nan, ZHU Yuan, CUI Ying. AC-LSTM transient stability assessment considering cost-sensitivity[J]. Power System Protection and Control, 2022, 50(22): 160-169.
- [9] 卫志农,李超凡,丁爱飞,等. 基于Tri-training-SSAE半监督学习算法的电力系统暂态稳定评估[J]. 电力自动化设备, 2023, 43(7): 110-116.
WEI Zhinong, LI Chaofan, DING Aifei, et al. Power system transient stability assessment based on Tri-training-SSAE semi supervised learning algorithm[J]. Electric Power Automation Equipment, 2023, 43(7): 110-116.
- [10] SU T, LIU Y, ZHAO J, et al. Deep belief network enabled surrogate modeling for fast preventive control of power system transient stability[J]. IEEE Transactions on Industrial Informatics, 2022, 18(1): 315-326.
- [11] CHEN L, GUAN L. Static information, K-neighbor, and self-attention aggregated scheme: a transient stability prediction model with enhanced interpretability[J]. Protection and Control of Modern Power Systems, 2023, 8: 1-16.
- [12] 李楠,李保罗,朱建华,等. 计及样本不平衡与重叠的暂态稳定评估方法[J]. 电力系统自动化, 2020, 44(21): 64-71.
LI Nan, LI Baoluo, ZHU Jianhua, et al. Transient stability assessment method considering sample imbalance and overlap[J]. Automation of Electric Power Systems, 2020, 44(21): 64-71.
- [13] 李嘉敏,杨红英,闫莉萍,等. 样本不平衡情况下的电力系统暂态稳定集成评估方法[J]. 电力系统自动化, 2021, 45(10): 34-41.
LI Jiamin, YANG Hongying, YAN Liping, et al. Integrated assessment method for transient stability of power system under sample imbalance[J]. Automation of Electric Power Systems, 2021, 45(10): 34-41.
- [14] BATISTA G E A P A, PRATI R C, MONARD M C. A study of the behavior of several methods for balancing machine learning training data[J]. ACM SIGKDD Explorations Newsletter, 2004, 6(1): 20-29.
- [15] 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.
- [16] 谭本东,杨军,赖秋频,等. 基于改进CGAN的电力系统暂态稳定评估样本增强方法[J]. 电力系统自动化, 2019, 43(1): 149-157.
TAN Bendong, YANG Jun, LAI Qiupin, et al. Data augment method for power system transient stability assessment based on improved conditional generative adversarial network[J]. Automation of Electric Power Systems, 2019, 43(1): 149-157.
- [17] BAO Jianmin, CHEN Dong, WEN Fang, et al. CVAE-GAN: fine-grained image generation through asymmetric training[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017: 2764-2773.
- [18] 周挺,杨军,周强明,等. 基于改进LightGBM的电力系统暂态稳定评估方法[J]. 电网技术, 2019, 43(6): 1931-1940.
ZHOU Ting, YANG Jun, ZHOU Qiangming, et al. Power system transient stability assessment method based on modified LightGBM[J]. Power System Technology, 2019, 43(6): 1931-1940.
- [19] 国家市场监督管理总局,国家标准化管理委员会. 电力系统安全稳定导则: GB 38755—2019[S]. 北京: 中国标准出版社, 2019.

- [20] 李宝琴,吴俊勇,强子玥,等. 基于改进DCGAN的电力系统暂态稳定增强型自适应评估[J]. 电力系统自动化,2022,46(2): 73-82.
- LI Baoqin, WU Junyong, QIANG Ziyue, et al. Enhanced adaptive assessment on transient stability of power system based on improved deep convolutional generative adversarial network [J]. Automation of Electric Power Systems, 2022, 46(2): 73-82.
- [21] 周生存,罗毅,易焯承,等. 基于边界强化混合采样的两阶段电力系统暂态稳定评估[J]. 电力自动化设备,2024,44(4): 143-150.
- ZHOU Shengcun, LUO Yi, YI Xuancheng, et al. Two-stage transient stability assessment of power system based on boundary enhanced hybrid sampling [J]. Electric Power Automation Equipment, 2024, 44(4): 143-150.
- [22] HE Miao, VITTAL V, ZHANG Junshan. Online dynamic secu-

rity assessment with missing PMU measurements: a data mining approach [J]. IEEE Transactions on Power Systems, 2013, 28 (2): 1969-1977.

作者简介:

马彬喻(2001—),男,硕士研究生,主要研究方向为数据驱动在电力系统中的应用(**E-mail**: 2023202070054@whu.edu.cn);

杨 军(1977—),男,教授,博士研究生导师,主要研究方向为电力系统运行与控制、电动汽车(**E-mail**: JYang@whu.edu.cn);

李 蕊(1974—),女,副教授,通信作者,主要研究方向为电力系统运行与控制、电力市场(**E-mail**: lirui@whu.edu.cn)。

(编辑 任思思)

Sample augment method for power system transient stability assessment based on improved CVAE-GAN

MA Binyu^{1,2}, YANG Jun^{1,2}, PENG Xiaotao^{1,2}, LI Rui^{1,2}, SHEN Jinpeng^{1,2}, JIANG Kezheng³,
LIU Dan³, CAO Kan³

(1. Hubei Engineering and Technology Research Center for AC / DC Intelligent Distribution Network, Wuhan 430072, China;

2. School of Electrical Engineering and Automation, Wuhan University, Wuhan 430072, China;

3. State Grid Hubei Electric Power Research Institute, Wuhan 430077, China)

Abstract: The proportion of transient instability samples in actual power system is small, and imbalanced data reduces the recognition rate of instability samples and reliability in data-driven transient stability assessment (TSA). Aiming at this problem, a sample augment method for TSA of electric power systems based on improved conditional variational auto-encoder with generative adversarial network (CVAE-GAN) is proposed. The composition ratio of input samples is improved to enhance the model's ability to learn distribution of instability samples, the model network structure is improved to adapt to the characteristics of electric power systems measurement data, and the pre-training is adopted to provide a good initial state for the model to promote training convergence. The high-quality instability samples are synthesized by using the improved CVAE-GAN model trained, and added to the original samples for sample augment. Then the classifier is retrained to achieve online TSA. The test results of the modified IEEE 39-bus system and the modified 500-bus power grid in South Carolina show that the proposed method can effectively learn the distribution characteristics of the original data, achieve sample augment, and thus improve the accuracy of TSA and the recognition rate of unstable samples.

Key words: data augment; data imbalance; CVAE-GAN; transient stability assessment; electric power systems