

# 基于改进 CGAN 的电力系统暂态稳定评估样本增强方法

谭本东<sup>1</sup>, 杨 军<sup>1</sup>, 赖秋频<sup>1</sup>, 谢培元<sup>2</sup>, 李 军<sup>2</sup>, 徐 箭<sup>1</sup>

(1. 武汉大学电气工程学院, 湖北省武汉市 430072; 2. 国网湖南省电力有限公司, 湖南省长沙市 410000)

**摘要:** 基于数据驱动的暂态稳定评估方法已成为电网安全领域研究的重点, 由于实际电力系统中暂态失稳情况极少, 给通过数据挖掘方法判断失稳情况带来了极大困难。针对这个问题, 提出了一种用于暂态稳定评估中失稳样本合成的数据增强方法, 对条件生成对抗神经网络(CGAN)训练方法的适应性进行改进以提高其学习稳定性, 在离线训练时利用改进 CGAN 交替训练生成器和判别器, 学习电力系统暂态数据的分布特性, 然后采用极限学习机(ELM)分类器筛选出改进 CGAN 所生成的多组样本中 G-mean 值最高的生成样本, 将其中失稳样本对原始失稳样本进行增强, 最后用增强后的原始样本训练分类器, 实现在线暂态稳定评估。仿真结果表明, 所提出的样本数据增强方法通过改进 CGAN 实现对原始数据分布特征的有效学习, 进而提升暂态稳定评估的正确率, 具有抗噪声干扰性强、对高维数据鲁棒性好的优点, 能够有效平衡电力系统失稳数据。

**关键词:** 电力系统; 暂态稳定评估; 数据增强; 条件生成对抗神经网络; G-mean 值

## 0 引言

暂态稳定是电力系统发生大干扰后能够保持同步稳定的一种能力<sup>[1]</sup>。随着大电网之间的互联、高渗透率新能源接入, 以及电力市场建设的推进, 电力系统的动态特性越来越复杂, 发生暂态失稳的风险也越来越大<sup>[2]</sup>, 因此快速、准确地对电力系统的暂态稳定状况进行评估是电力系统安全运行的保障。

时域仿真方法<sup>[3]</sup>计算精度随着模型复杂度的增加, 同时也带来了计算速度的问题。直接法<sup>[4-5]</sup>对模型进行了一定简化, 导致系统的稳定估计结果较为保守, 无法满足大电网中应用的需求。这些方法在实际在线计算时无法同时达到精度、速度的要求, 因此学者们提出了基于数据驱动的暂态稳定评估方法。

目前基于数据驱动的暂态稳定评估相关研究成果可以分为两种思路。第一种思路是利用机器学习算法构建广域测量系统(WAMS)数据<sup>[6-7]</sup>和暂态稳定性之间的映射关系以及相应的稳定规则, 一旦发生故障就能根据这个映射关系找到实时数据对应的稳定性<sup>[8-11]</sup>。这方面早期的工作均集中于研究各种分类算法的性能和应用场景, 例如决策树方法<sup>[12]</sup>、人工神经网络<sup>[13]</sup>、支持向量机<sup>[14-16]</sup>等, 而自适应评

估方法<sup>[17-18]</sup>利用故障后轨迹和稳定性之间的映射关系进行连续预测。另外一种思路是利用 WAMS 提供的系统动态过程的时间序列响应, 直接应用某种时间序列预测方法或机器学习方法预测系统未来的受扰轨迹, 并判断系统的稳定性<sup>[19]</sup>。

但是基于数据驱动的暂态稳定评估方法存在着严重的类别不平衡问题。由于现代电力系统网架的坚强性, 实际电网在遭受大干扰之后一般都能恢复到正常运行状态, 因此失稳的样本数量极少。在这种情况下, 如果直接使用 WAMS 数据训练机器学习算法模型将很容易造成误判。一个极端的例子是在 100 个样本中只有一个失稳样本, 则算法只需要全部评估成稳定就能达到 99% 的正确率, 这显然不合理。目前的解决思路主要从算法和数据层面上进行改善; 算法的改进主要包括模型集成和代价敏感方法, 如 Easy ensemble<sup>[20]</sup> 和 Adacost<sup>[21]</sup> 等, 其泛化性能不理想; 数据上的改进可以分为欠采样和过采样方法<sup>[22]</sup> 两类。欠采样方法由于会损失多数类的数据信息, 因此应用范围有限。过采样方法中随机采样(random oversampling, ROS)<sup>[23]</sup> 容易造成过拟合问题, 对模型训练不利; 基于 SMOTE 和 ADASYN<sup>[24]</sup> 算法的数据合成方法则属于线性插值方法, 其合成的数据和实际电网数据的物理特性和运行状态存在相关性不强的问题<sup>[22]</sup>。而现有的非线性数据合成方法<sup>[22]</sup> 利用小波变换等方式非线性化, 再进一步线性插值以合成数据, 因此只适用于多

收稿日期: 2018-05-22; 修回日期: 2018-10-11。

上网日期: 2018-11-20。

国家重点研发计划资助项目(2017YFB0902900)。

<http://www.aeps-info.com> 149

维时间序列数据,无法应用于暂态稳定这样单个时刻的特征数据合成。

针对电力系统暂态稳定失稳样本不足造成的数据不平衡问题,本文提出了一种基于改进条件生成对抗神经网络(conditional generative adversarial network,CGAN)的非线性数据增强方法。该方法对 CGAN 进行改进以提高其学习稳定性,针对 CGAN 训练困难以及暂态稳定问题存在临界稳定而难以区分的情况,增加预训练判别器步骤以提高判别器对暂态数据的区分性能,从而提升生成器学习能力使得实际数据分布与生成分布的 Wasserstein 距离逼近到 0;建立了暂态稳定评估数据增强框架,用改进的 CGAN 交替训练判别器和生成器以学习原始数据的分布特性,提取训练得到的 G-mean 值最大的合成数据,并反归一化用于样本数据增强以实现在线暂态稳定评估。

## 1 改进 CGAN 算法

### 1.1 生成对抗神经网络

如附录 A 图 A1 所示,生成对抗神经网络(generative adversarial network,GAN)是一种近几年兴起的生成模型,它通过对抗训练来克服基于极大似然估计相关生成模型中存在概率估计计算困难的问题,由生成器 G 和判别器 D 组成。相比于线性数据合成方法以及目前只适用于时间序列特征的非线性合成方法,GAN 的优点在于能够学习到原始数据的分布,因此可以被视为一种能够用于数据合成的非线性插值方法,适合解决电力系统暂态单个时刻不平衡数据的机器学习问题。在本文中生成器相关的参数均用下标“g”表示,判别器相关参数均用下标“d”表示。

在暂态稳定评估过程中,每个样本输入都是一个特征向量,因此对抗框架可以直接使用多层神经网络。定义数据集中第  $i$  个样本的符号为  $x_i$ ,则整个数据集可以表示成  $X = \{x_1, x_2, \dots, x_N\}$ ,其中  $N$  为样本总数目。为在数据集  $X$  的实际分布  $P_r$  中学习生成分布  $P_g$ ,生成器 G 构建了一个从先验分布  $P_z$ (也被称为隐空间)到生成数据空间的映射关系  $G(z; \theta_g)$ ,其中  $\theta_g$  为生成器的参数, $z$  为先验分布的随机变量。先验分布一般选择高斯分布,它表征的是在高维空间中实际分布  $P_r$  和生成分布  $P_g$  的支撑集。而判别器 D 则为判别输入是实际数据还是生成数据的一种函数  $D(I; \theta_d)$ ,其中  $I$  代表判别器 D 的输入, $\theta_d$  为判别器 D 的参数。

生成器 G 由多层神经网络构成,输入为先验分布  $P_z$  对应的随机变量  $z$ ,输出为和  $x_i \in X$  维度一

致的向量  $O_g$ :

$$O_g = G(z; \theta_g) \quad (1)$$

生成器的目标是生成尽量逼真的数据,从而达到迷惑判别器的目的,因此它的损失函数可以定义为  $E_{z \sim P_z}[-D(G(z))]$ 。当生成器 G 生成的数据不够逼真时, $E_{z \sim P_z}[D(G(z))]$  的值越小;相反,生成器 G 生成的数据足够逼真时,判别器输出值则越大,因此生成器的目标可以视为最大化  $E_{z \sim P_z}[D(G(z))]$ ,即可定义生成器 G 的目标函数为:

$$\min E_{z \sim P_z}[-D(G(z))] \quad (2)$$

判别器也是一个神经网络,只不过它的输入是和  $x_i \in X$  维度一致的向量  $I$ ,因此  $I$  既可以是实际数据也可以是生成数据。判别器的主要任务是区分两种数据,所以它的输出是一个在 0 到 1 之间的标量,也就是属于实际数据或者生成数据的概率。判别器的损失函数定义为  $E_{x \sim P_r}[D(x)] + E_{z \sim P_z}[-D(G(z))]$ ,由于判别器的目的是尽可能区分实际数据和生成数据,因此它会努力将  $E_{x \sim P_r}[D(x)]$  最大化而将  $E_{z \sim P_z}[D(G(z))]$  最小化。判别器的目标函数为:

$$\max E_{x \sim P_r}[D(x)] + E_{z \sim P_z}[-D(G(z))] \quad (3)$$

为了构建对抗神经网络整体训练的目标函数,本文引入 Wasserstein 距离  $W(P_r, P_g)$ <sup>[25]</sup> 来描述数据分布  $P_r$  和  $P_g$  的期望距离:

$$W(P_r, P_g) = \inf_{\gamma \in f(x, O_g)} E_{(x,s) \sim \gamma}[\|x - O_g\|] \quad (4)$$

式中: $f(x, O_g)$  代表  $P_r$  和  $P_g$  的联合分布  $\gamma(x, O_g)$  的集合,那么  $\gamma(x, s)$  的边缘分布就分别为  $P_r$  和  $P_g$ ;  $\inf$  表示下确界。

$W(P_r, P_g)$  越小,表明实际分布和生成分布的相似度越高,相比于其他常用的描述数据分布间的距离,即使两个分布的支撑集没有重叠或者重叠非常少,Wasserstein 距离仍能反映两个分布的远近。根据 Kantorovich-Rubinstein duality 定律<sup>[26]</sup>,可以将  $W(P_r, P_g)$  表示为:

$$\max_{\|D(I)\| \leq 1} E_{x \sim P_r}[D(x)] + E_{z \sim P_z}[-D(G(z))] \quad (5)$$

$\|D(I)\| \leq 1$  约束是为了在数据发生轻微变化的时候判别器 D 的输出变化不至于太剧烈,可以等价表示为:

$$\max E_{x \sim P_r}[D(x)] + E_{z \sim P_z}[-D(G(z))] + \lambda E_{\hat{x} \sim P_{\hat{x}}}[\|\nabla \hat{x} D(\hat{x})\|_p - 1]^2 \quad (6)$$

式中: $\|\cdot\|_p$  表示  $p$  范数; $\lambda$  为正则项系数; $\hat{x} = \epsilon x + (1 - \epsilon)G(z)$ ,其中  $\epsilon \sim U[0, 1]$ , $U$  为均匀分布。

则 GAN 整体训练的目标函数为:

$$\min_G \max_D E_{x \sim P_r} [D(x)] + E_{z \sim P_z} [-D(G(z))] + \lambda E_{\hat{x} \sim P_{\hat{x}}} [\|\nabla \hat{x} D(\hat{x})\|_p - 1]^2 \quad (7)$$

一般采用判别器 D 和生成器 G 交替训练的模式进行学习来获得有效的 GAN 模型。也就是说,在一轮训练中判别器 D 先训练  $k$  次,紧接着训练生成器 G,因此存在一个较好的  $k$  值使得整个 GAN 达到良好的水平。GAN 训练过程如附录 A 图 A2 所示。在刚开始训练的时候,由于生成器 G 首先进行随机初始化,因此其对应的数据分布距离实际分布很远,判别器 D 能够很容易将生成的数据区分开来。但是随着训练程度的深入,生成器 G 生成的数据越来越接近实际数据分布,这时判别器 D 辨别也越来越困难;当判别器 D 和生成器 G 之间的博弈达到纳什均衡时,表明模型训练完成。理论上该模型训练过程存在最优解<sup>[27]</sup>,此时  $D(I) = 0.5$ 。根据 Wasserstein 距离的定义,Wasserstein 距离在 0 附近波动时,表明生成分布非常接近实际分布,因此本文将其作为评估 GAN 训练程度的指标。

## 1.2 改进 CGAN

GAN 只能学习到实际数据的整体分布,但是对于暂态稳定评估来说,其关注的是具体的稳定状况。为了加入有关稳定状况的额外信息,本文提出了一种改进 CGAN<sup>[28]</sup>,其网络结构如附录 A 图 A3 所示。这里的条件是指类别条件,可以同时同时对判别器 D 和生成器 G 的输入融入类别 one-hot 向量,从而实现 CGAN 做出类别信息上的指导。one-hot 向量利用二进制数来表示类别信息,而暂态稳定评估可归类为二分类问题,那么暂态稳定可以表示为  $[0, 1]$ ,暂态失稳可以表示为  $[1, 0]$ 。

定义类别 one-hot 向量为  $y$ ,在先验分布  $P_z$  中结合向量  $y$  一起作为 CGAN 生成器 G 的隐空间表达输入。对于 GAN 的判别器 D,采用和生成器 G 同样的处理方式,将数据  $I$  和类别信息  $y$  一起融合作为判别器 D 的输入。在一轮训练中,判别器 D 和生成器 G 所用的类别信息是一致的,因此 CGAN 整体训练的目标函数为:

$$\min_G \max_D E_{x \sim P_r} [D(x|y)] + E_{z \sim P_z} [-D(G(z|y))] + \lambda E_{\hat{x} \sim P_{\hat{x}}} [\|\nabla \hat{x} D(\hat{x}|y)\|_p - 1]^2 \quad (8)$$

其显式表达式为:

$$\min_G \max_D \frac{1}{N} \sum_{i=1}^N D([\mathbf{x}_i, \mathbf{y}_i]) - D([\mathbf{G}([\mathbf{z}_i, \mathbf{y}_i]), \mathbf{y}_i]) + \lambda \frac{1}{N} \sum_{i=1}^N [\|\nabla D([\hat{\mathbf{x}}_i, \mathbf{y}_i])\|_p - 1]^2 \quad (9)$$

式中:  $[\cdot, \cdot]$  表示对两个向量进行拼接;下标  $i$  表示第  $i$  个样本。

CGAN 同样采用的是多层神经网络结构和交替对抗训练方式,隐含层激活函数采用 Relu 函数<sup>[29]</sup>,输出层激活函数为 sigmoid 函数,因此在具体应用时,需要将数据归一化至区间  $[0, 1]$  内。在优化 CGAN 时采用梯度下降的策略,为了保证稳定训练过程,本文不采用基于动量的梯度下降策略,而选择 RMSProp 算法<sup>[30]</sup>进行训练。针对 CGAN 训练困难以及暂态稳定问题存在临界稳定而造成难以区分的情况,进一步增加预训练判别器步骤实现对 CGAN 的改进,以提高判别器对暂态数据区分的性能,从而促进生成器的学习能力,使得实际数据分布与生成分布的 Wasserstein 距离减小至 0 左右。

从基于数据驱动的暂态稳定评估角度来看,CGAN 生成器针对失稳条件下的生成分布可以看成一种非线性插值的过程。实际数据集通过模拟多种运行方式及故障来产生大量故障样本,但是这些数据并不能代表整个数据空间,而数据生成方法则相当于对这些离散的运行点进行插值,合成一些不曾模拟的运行点。因此应用在暂态稳定评估中,相当于增加了失稳样本的数目,能够有效帮助发现失稳模式。

## 2 基于数据增强方法的暂态稳定评估模型

### 2.1 基于极限学习机的基础分类器

极限学习机 (extreme learning machine, ELM)<sup>[31]</sup> 是一种简单易用、有效的单隐层前馈神经网络,目前在电力系统暂态稳定领域取得了广泛应用<sup>[17,32]</sup>。对于一个样本数目为  $N$  的数据集  $(x_i, y_i)$ ,其中  $x_i \in \mathbf{R}^n, y_i \in \mathbf{R}^2, y_i$  为第  $i$  个样本的类别标签,具有  $M$  个隐藏层单元的 ELM 输出为:

$$f(x_j) = \sum_{i=1}^M \beta_i h(w_i x_j + b_i) = y_j \quad (10)$$

式中:  $j=1, 2, \dots, N; h$  为激活函数;  $w_i \in \mathbf{R}^n$ , 为隐藏层第  $i$  个单元的权重向量;  $\beta_i \in \mathbf{R}^2$ , 为隐藏层第  $i$  个单元的乘子;  $b_i \in \mathbf{R}^1$ , 为隐藏层第  $i$  个单元的偏置。

在 ELM 中权值向量  $w_i$  和偏置  $b_i$  都是随机给定的,乘子  $\beta_i$  可以基于所有数据通过 Moore-Penrose 广义逆求取,因此不需要反向传播 (BP) 神经网络所必需的反向调节过程,所以 ELM 的计算速度要显著快于基于优化的分类算法。大量实践证明,ELM 的泛化性能非常优秀<sup>[33]</sup>,因此在本文中 ELM 作为基础分类器。

### 2.2 模型输入输出及评价指标

根据 GAN 理论<sup>[27]</sup>,目前它只能处理连续值;一

旦将数据离散化,那么将无法有效训练对抗神经网络。因此对电力系统在多个运行条件下的仿真,根据已有的文献<sup>[8,10-11]</sup>选择与暂态稳定性强相关的物理量,记录故障切除时刻的发电机功角、发电机有功及无功、线路电压、线路功率、负荷有功及无功等物理量原始数据,用于构建暂态稳定特征,在线应用时可以通过 WAMS 及时准确地获取。为了保证 GAN 稳定训练,对特征输入归一化处理得:

$$x_{ij}^* = \frac{x_{ij} - \min(x_{ij})}{\max(x_{ij}) - \min(x_{ij})} \quad (11)$$

式中:  $x_{ij}$  为第  $i$  个样本的第  $j$  个特征,经过归一化处理之后,所有输入数据数值范围都在  $[0,1]$  之间,在应用暂态稳定评估时再根据式(11)反归一化即可。

由于暂态稳定评估问题存在失稳样本极少的问题,如果只使用准确率作为模型的评价指标非常不合理。假设数据集中稳定样本占 99%,那么模型只需要全部评估为稳定就能达到 99% 的正确率。为了能够在类别不平衡情况下仍能对模型性能做出有效评估,定义如下指标来评估模型的性能:

$$\lambda_{TSR} = \frac{f_{TS}}{f_{TS} + f_{FU}} \times 100\% \quad (12)$$

$$\lambda_{TFR} = \frac{f_{TU}}{f_{FS} + f_{TU}} \times 100\% \quad (13)$$

$$\lambda_{G-mean} = \sqrt{\lambda_{TSR} \lambda_{TFR}} \times 100\% \quad (14)$$

$$\lambda_{Acc} = \frac{f_{TS} + f_{TU}}{f_{TS} + f_{FU} + f_{FS} + f_{TU}} \times 100\% \quad (15)$$

式中:  $f_{TS}$  为预测稳定正确(true stable)的样本数量;  $f_{FS}$  为预测稳定错误(false stable)的样本数量;  $f_{TU}$  为预测失稳正确(true unstable)的样本数量;  $f_{FU}$  为预测失稳错误(false unstable)的样本数量。

上述指标中,  $\lambda_{TSR}$  表示预测为稳定的正确结果在所有稳定样本中的占比;  $\lambda_{TFR}$  表示预测为失稳的正确结果在所有失稳样本中的占比,  $\lambda_{TFR}$  越大,漏警率(失稳情况判断成稳定)越低;  $\lambda_{G-mean}$  为  $\lambda_{TSR}$  和  $\lambda_{TFR}$  的几何平均,能够有效衡量不平衡数据的评估性能;而  $\lambda_{Acc}$  则代表了总体正确率。

### 2.3 基于改进 CGAN 的电力系统暂态稳定评估数据增强模型

为了能够有效增强电力系统数据,本文提出了一种基于改进 CGAN 的电力系统暂态稳定评估数据增强模型,如图 1 所示。为了能够有效评估改进 CGAN 合成数据的合理性,本文采用了“训练数据构建分类器,合成数据测试效果”的合成数据评估方法来选择  $\lambda_{G-mean}$  值最高的合成数据。

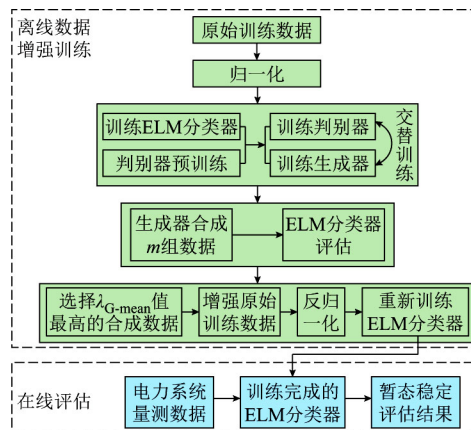


图 1 暂态稳定评估数据增强模型框架  
Fig.1 Framework of data augment model for transient stability assessment

#### 1) 离线数据增强训练

步骤 1: 首先将原始训练数据归一化,进而训练 ELM 分类器用于评估改进 CGAN 在训练过程中每一轮生成的数据。同时,针对 CGAN 存在训练困难的问题,进一步对判别器进行预训练,从而提高判别器性能,促进生成器的进化。

步骤 2: 依次交替训练判别器和生成器。为了有效训练改进 CGAN,采用首先训练判别器  $k$  次再训练生成器 1 次的训练方式,这个过程称为对改进 CGAN 的交替训练(1 轮训练)。每经过 1 轮训练,即生成一组失稳样本数目与原始数据稳定样本数目相同的合成数据,因此在  $m$  轮训练后,改进 CGAN 训练完成后可以生成  $m$  组合成数据。最后利用训练好的 ELM 分类器评估出  $\lambda_{G-mean}$  值最高的一组合成数据。

步骤 3: 取出步骤 2 中评估出的  $\lambda_{G-mean}$  值最高的合成数据中的所有失稳样本并融合至原始训练数据中,更新原始训练数据,从而达到对训练数据增强的效果。对增强后的数据反归一化,重新训练 ELM 分类器并应用于在线评估。

#### 2) 在线评估

离线数据增强训练之后,可以改善数据不平衡度,从而降低误判率。在线评估时,直接应用离线数据增强训练所得 ELM 分类器,对于不断获取的电力系统量测数据,分类器可以做出准确、快速的判断。

## 3 仿真分析

### 3.1 数据集生成

仿真硬件为 Intel® Core (TM) i5-7300HQ CPU@2.50 GHz, 8.00 GB 运行内存、内存为 4 GB 的 NVIDIA GEFORCE GTX 1050ti GPU 的 PC, 数

据增强模型基于深度学习框架 Tensorflow1.4<sup>[34]</sup> 和 Keras<sup>[35]</sup> 构建。

如附录 A 图 A5 所示,本文选择基准频率为 60 Hz 的新英格兰 10 机 39 节点典型电力系统<sup>[36]</sup> 作为测试电网。该测试系统包含 39 条母线、10 台发电机、19 个负荷和 34 条传输线,在 39 号母线上的发电机是外部电力系统的等值机,其功角作为其他发电机的参考值。用 PSS/E<sup>[37]</sup> 软件进行时域仿真,负荷设置为恒阻抗模型,发电机设置为 GENROU 模型,步长设置为 0.008 3 s。为了构造较为完备的样本空间,电力系统运行状态包括 70%,75%,80%,...,115%,120% 等 11 种负荷水平,发电机根据负荷水平做出相应改变。在这 11 种运行方式下进行潮流计算,如果潮流收敛,那么分别所有母线以及距离线路首端 20%,40%,60%,80% 处设置三相接地短路故障。故障持续时间为 0.1 s 或 0.3 s,设置仿真时长为 20 s,从而能够准确判断电力系统暂态稳定性。仿真结束时,如果电力系统任意两个发电机转子角度之差超过 360° 则判定为暂态失稳<sup>[16]</sup>,最后总共获得 3 848 个有效样本。为验证本文所提方法,根据文献<sup>[22]</sup> 失稳样本的占比设置,随机剔除一些失稳样本以模拟实际电力系统失稳样本稀少的情况,保证总样本数目和失稳样本数目比例为 20 : 1,最后得到其中 1 403 个样本用作训练集,剩下的 600 个样本作为测试集。

### 3.2 结果讨论

#### 1) 改进 CGAN 训练过程中数据分布变化

改进 CGAN 的优点之一就是能够近似学习到原始数据的分布,可类比于一种“非线性插值”过程。如附录 A 图 A2 所示,为了分析改进 CGAN 训练过程中数据分布的变化情况,可利用 XGBoost 算法<sup>[38]</sup> 找到特征重要性最高的两个特征 F0 和 F130 来可视化变化过程(本文没有采用如主成分分析这类降维算法,是因为它会对原始数据空间变换而不利于发现数据分布变化规律)。

由于改进 CGAN 判别器和生成器均为多层神经网络,因此对抗训练采用小批量训练方式来提高模型的泛化性。通过网格搜索优化模型参数,设置比较重要的模型参数如附录 A 表 A1 所示。

在改进 CGAN 训练过程中,保持生成数据稳定样本数目及失稳样本数目和原始训练数据一致。从图 2 所示改进 CGAN 训练过程中生成数据分布变化的情况可以发现,在刚开始训练时,改进 CGAN 随机生成的数据和实际数据分布大相径庭,但是随着训练的深入,两个分布之间的距离越来越小,按照图 2 箭头指向顺序分别代表改进 CGAN 第 100,

4 100,7 900,9 600 代训练生成的数据,表明改进 CGAN 生成的数据分布和实际数据分布(见图 3(a)) 越来越接近。因为在训练过程中,Wasserstein 距离不断减小并最后保持在 0 附近波动,如图 3(b) 所示,这说明生成的数据分布非常接近实际数据的分布。相比于改进 CGAN,CGAN 的 Wasserstein 距离要相对大得多,因此改进 CGAN 能够更有效地学习到原始数据的分布特性。

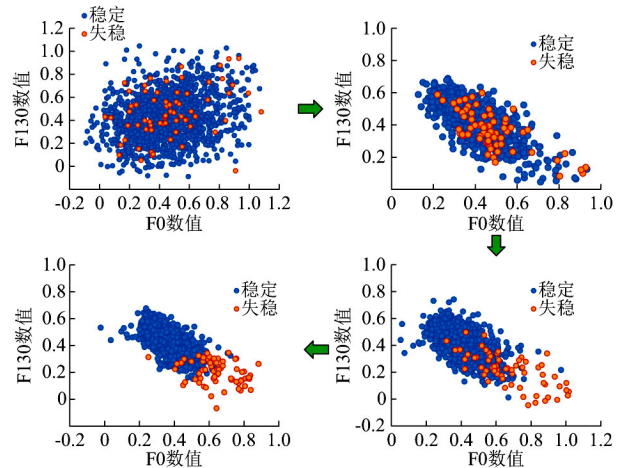


图 2 改进 CGAN 训练过程中的数据分布变化

Fig.2 Change of data distribution in improved CGAN training process

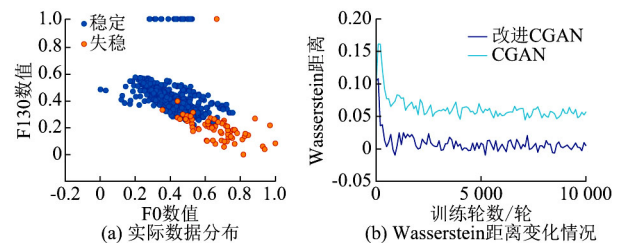


图 3 实际数据分布及 Wasserstein 距离变化情况

Fig.3 Real data distribution and change of Wasserstein distance

#### 2) 不同合成数据方法的性能对比

为了和改进 CGAN 进行对比,本文考虑了一些常用的数据合成方法,包括 SMOTE, ROS 和 ADASYN 算法。改进 CGAN 训练参数和 3.2 节 1) 中模型参数一致。所有方法均保证合成结果使得数据集达到绝对平衡,即最终得到的增强数据中稳定样本数目和失稳样本数目相等。所有方法在测试集上结果如表 1 所示。

原始数据由于失稳样本极少(占 5%),简单地将所有样本识别为稳定状态时所对应的分类器泛化能力最弱,即  $\lambda_{G\_mean}$  值最小。而使用广泛的数据合成方法如 SMOTE, ROS, ADASYN 均在一定程度上提升了分类器的泛化性能,但是仍然无法有效识

别失稳情况,其  $\lambda_{\text{TFR}}$  值均较小,因此漏警率高。

表 1 各种数据合成方法性能对比  
Table 1 Performance comparison of various data synthesis algorithms

算法	$\lambda_{\text{TSR}}/\%$	$\lambda_{\text{TFR}}/\%$	$\lambda_{\text{G-mean}}/\%$	$\lambda_{\text{Acc}}/\%$
原始数据	100.00	0	0	95.17
SMOTE	98.95	75.86	86.64	97.83
ROS	99.65	68.97	82.90	98.17
ADASYN	96.50	89.66	93.01	96.17
改进 CGAN	98.60	96.55	97.60	98.50

本文对各种合成数据方法利用主成分分析算法降维成二维  $V_1$  和  $V_2$  并可视化,如图 4 所示,发现 ROS 只是简单地从原始数据中复制失稳样本,这样做的结果是合成数据中并没有新样本,样本之间互相重叠,极易导致过拟合问题;SMOTE 和 ADASYN 算法均为基于线性插值的方法,因此图 4(a) 和 (c) 显示失稳样本线性分布。而 ADASYN 算法可以自适应合成远离边界点,因此 ADASYN 算法不容易造成边界重叠问题,最终准确率和  $\lambda_{\text{G-mean}}$  值均高于 SMOTE 算法。本文所提出的改进 CGAN 方法是基于分布学习的方法,所以在数据分布特性上接近原始数据分布,可以合成原始数据从未出现过而又“真实”的数据;从合成失稳数据分布来看,改进 CGAN 属于非线性合成方法,它在 4 种数据合成方法中的  $\lambda_{\text{TFR}}$  值、 $\lambda_{\text{Acc}}$  值和  $\lambda_{\text{G-mean}}$  值最高,因此漏警率非常低,能够有效降低暂态失稳带来的风险。可见从数据合成方式和合成性能上比较,改进 CGAN 更加适合于电力系统失稳数据的平衡。

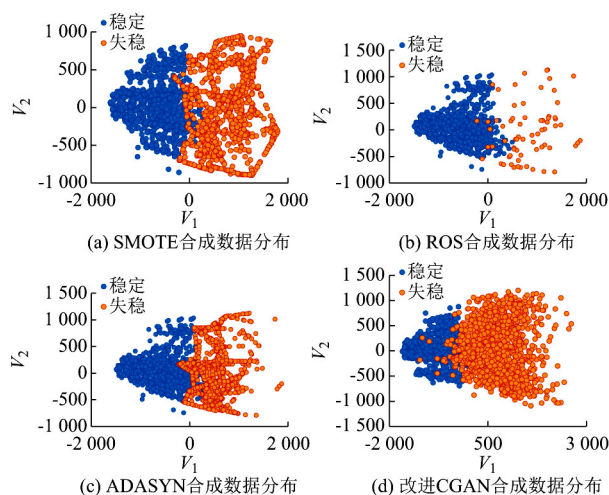


图 4 不同方法合成数据分布对比  
Fig.4 Comparison of data distribution with various synthesis methods

### 3) 模型对噪声的鲁棒性

在之前的分析中,假设所有量测数据都是精确

测量值,但考虑到电力系统 WAMS 一般存在噪声的干扰,可能对算法造成一定影响。根据 IEEE C37.118 标准对相量测量单元(PMU)所制定的标准,PMU 量测相量误差应小于 1%<sup>[39]</sup>,具体表达式为:

$$\left| \frac{\tilde{X} \angle \tilde{\theta} - X \angle \theta}{X \angle \theta} \right| < 1\% \quad (16)$$

式中:  $\tilde{X} \angle \tilde{\theta}$  为从 WAMS 获取的量测数据相量;  $X \angle \theta$  为 WAMS 量测数据相量减去噪声之后的真实值。

根据信噪比定义,可以采用 40 dB 的高斯白噪声来模拟量测误差,从而检验所提出方法的鲁棒性。改进 CGAN 训练参数和 3.2 节 1) 中模型参数一致,各种数据合成方法评估结果如表 2 所示。

表 2 考虑广域噪声下各种数据合成方法性能对比  
Table 2 Performance comparison of various data synthesis algorithms under wide-area noise

算法	$\lambda_{\text{TSR}}/\%$	$\lambda_{\text{TFR}}/\%$	$\lambda_{\text{G-mean}}/\%$	$\lambda_{\text{Acc}}/\%$
原始数据	100.00	0	0	95.17
SMOTE	82.31	86.21	84.23	82.50
ROS	90.19	72.41	80.82	89.33
ADASYN	81.09	89.66	85.26	81.50
改进 CGAN	96.32	82.76	89.28	95.67

从表 2 结果可以看出,所有数据合成方法在噪声干扰下  $\lambda_{\text{Acc}}$  和  $\lambda_{\text{G-mean}}$  值均有所变化,但是 SMOTE 和 ADASYN 等基于线性插值的方法鲁棒性较差,  $\lambda_{\text{Acc}}$  和  $\lambda_{\text{G-mean}}$  值最小,这是由于这两种数据合成方法是根据特征空间的分布进行线性插值来工作的,没有考虑到噪声这类异常点的影响,因此极易在噪声点和正常点之间合成异常数据;ROS 方法在受到噪声干扰后性能反而有进一步提高,噪声的存在使得分类器不会过分关注在重复的失稳样本上。在所有结果中,虽然改进 CGAN 的  $\lambda_{\text{TFR}}$  值低于 SMOTE 和 ADASYN 方法,但不论是  $\lambda_{\text{Acc}}$  还是  $\lambda_{\text{G-mean}}$  值,改进 CGAN 的评估性能均为最优异,在噪声干扰下依然具有较低的漏警率。这个结果表明改进 CGAN 对噪声的干扰相对不敏感,鲁棒性较强,也就是说即使存在噪声,改进 CGAN 也能够通过对抗的训练方式减小噪声的干扰,学习到原始数据的分布特征,从而生成和原始数据特性接近的失稳数据,保证暂态稳定评估结果的可靠性。

### 4) 高维数据的影响

为了进一步验证本文所提方法对电力系统高维数据的处理能力,在美国南卡罗来纳州实际电网上进一步测试了该算法的性能。该电网由 56 台发电机、500 条母线组成,单线图如附录 A 图 A6 所示。

该系统故障持续时间设置为 0.1, 0.3, 0.5 s 共 3 种, 潮流水平考虑基准值的 70%, 80%, 90%, 100%, 110% 共 5 种, 故障位置为所有母线以及每条线路距离母线首端的 20%, 40%, 60%, 80%, 最后共产生 34 725 个样本, 其中包含 1 243 个失稳样本。最终 24 308 个样本用于训练, 剩下的样本用于测试。

算法特征输入仍然由发电机功角、发电机有功及无功、线路电压、线路功率、负荷有功及无功组成, 但是总维度达到了 3 956 维, 远高于新英格兰 39 节点电力系统算例的特征维度。对于算法学习而言, 高维数据容易造成过拟合问题, 因此本文利用这个大规模实际电网来测试改进 CGAN 算法在高维空间下的性能。改进 CGAN 和其他数据合成算法结果如表 3 所示。

表 3 各种数据合成算法在高维数据下的性能对比  
Table 3 Performance comparison of various data synthesis algorithms with high dimensional data

算法	$\lambda_{TSR}/\%$	$\lambda_{TFR}/\%$	$\lambda_{G-mean}/\%$	$\lambda_{Acc}/\%$
原始数据	99.96	0.81	8.98	96.42
SMOTE	85.95	90.59	88.24	86.12
ROS	93.89	84.14	88.88	93.53
ADASYN	89.45	93.01	91.21	89.58
改进 CGAN	98.43	93.01	95.68	98.24

如表 3 结果所示, SMOTE, ROS, ADASYN 三种方法对高维数据均非常敏感, 而本文所提出的改进 CGAN 方法对高维数据依然有强大的泛化性能, 因此在  $\lambda_{G-mean}$  和  $\lambda_{Acc}$  值上均高于 3 种传统的数据合成方法, 特别是在稳定样本的正确识别率上优势明显。这是因为 SMOTE 和 ADASYN 是两种基于线性插值的方法, 造成高维数据方差和平均欧式距离减小<sup>[40]</sup>, 从而减小了数据稳定样本和失稳样本的差异性, 甚至大幅降低稳定样本的评估正确率, 因此无法有效平衡数据。ROS 算法则因为其简单重复样本造成过拟合。而改进 CGAN 采用对抗训练的神经网络在深度学习中提取了有效的高阶特征, 因此高维数据下依然能够对其分布特性进行学习。

#### 4 结语

针对电力系统暂态失稳情况稀少造成的样本不平衡问题, 本文提出了一种基于改进 CGAN 的暂态稳定样本非线性数据增强方法, 能够有效学习原始数据分布特性, 应用于暂态稳定评估能够有效降低误判率。在新英格兰 10 机 39 节点电力系统和美国南卡罗来纳州实际电网上进行了仿真计算, 得到如下结论。

1) 改进 CGAN 的训练方式能够有效学习到原

始数据的分布, 生成接近原始数据特性的数据, 其优势在于能够不断平衡数据; 相比于 SMOTE 等数据线性合成算法, 改进 CGAN 属于非线性数据合成方法。

2) 基于改进 CGAN 的电力系统数据增强框架能够有效平衡电力系统失稳数据, 在评估准确率和  $\lambda_{G-mean}$  值上较现有常用的传统数据合成方法提升效果更为明显。

3) 由于改进 CGAN 是基于对抗的训练方式, 因此抗噪声干扰和处理高维数据的能力相比其他数据合成方法要更强, 适合应用于大电网这样的高维、非线性的复杂高阶系统暂态稳定数据的增强。

本文提出的方法是针对单一时刻的电力系统暂态失稳数据进行增强, 随着目前可靠性更高的自适应暂态稳定评估方法的研究不断深入以及深度强化序列学习的持续突破, 如何生成适用于所有运行方式下的电力系统失稳序列, 并应用于自适应暂态稳定评估以提高数据平衡性将是未来研究的重点。

附录见本刊网络版 (<http://www.aeps-info.com/aeps/ch/index.aspx>)。

#### 参考文献

- [1] KUNDUR P, PASERBA J, AJJARAPU V, et al. Definition and classification of power system stability [J]. IEEE Transactions on Power Systems, 2004, 19(3): 1387-1401.
- [2] ZHANG C, LI Y, YU Z, et al. A weighted random forest approach to improve predictive performance for power system transient stability assessment [C]// Power and Energy Engineering Conference (APPEEC), April 15-17, 2016, Suzhou, China; 1259-1263.
- [3] KUNDUR P, BALU N J, LAUBY M G. Power system stability and control[M]. New York: McGraw-hill, 1994: 848-859.
- [4] XUE Y, VAN CUTSEM T, RIBBENS-PAVELLA M. A simple direct method for fast transient stability assessment of large power systems[J]. IEEE Transactions on Power Systems, 1988, 3(2): 400-412.
- [5] 王锡凡. 现代电力系统分析[M]. 北京: 科学出版社, 2003: 354-362.  
WANG Xifan. Analysis on modern power system[M]. Beijing: Science Press, 2003: 354-362.
- [6] ZHANG Yingchen, MARKHAM P, XIA Tao, et al. Wide-area frequency monitoring network (FNET) architecture and applications[J]. IEEE Transactions on Smart Grid, 2010, 1(2): 159-167.
- [7] DE LA R J, CENTENO V, THORP J S, et al. Synchronized phasor measurement applications in power systems[J]. IEEE Transactions on Smart Grid, 2010, 1(1): 20-27.
- [8] 于之虹, 黄彦浩, 鲁广明, 等. 基于时间序列关联分析的稳定运行规则提取方法[J]. 中国电机工程学报, 2015, 35(3): 519-526.  
YU Zhihong, HUANG Yanhao, LU Guangming, et al. A time series associative classification method for the operation rule

- extracting of transient stability[J]. Proceedings of the CSEE, 2015, 35(3): 519-526.
- [9] 黄天罡, 薛禹胜, 陈国平, 等. 暂态稳定算例的高效剔除[J]. 电力系统自动化, 2018, 42(8): 83-91. DOI: 10.7500/AEPS20180302011.  
HUANG Tiangang, XUE Yusheng, CHEN Guoping, et al. An efficient stable case screening algorithm for transient stability assessment[J]. Automation of Electric Power Systems, 2018, 42(8): 83-91. DOI: 10.7500/AEPS20180302011.
- [10] 朱乔木, 陈金富, 李弘毅, 等. 基于堆叠自动编码器的电力系统暂态稳定评估[J]. 中国电机工程学报, 2018, 38(10): 2937-2946.  
ZHU Qiaomu, CHEN Jinfu, LI Hongyi, et al. Transient stability assessment based on stacked autoencoder [J]. Proceedings of the CSEE, 2018, 38(10): 2937-2946.
- [11] TANG Y, LI F, WANG Q, et al. Hybrid method for power system transient stability prediction based on two-stage computing resources[J]. IET Generation, Transmission & Distribution, 2018, 12(8): 1697-1703.
- [12] GUO T, MILANOVIC J V. Probabilistic framework for assessing the accuracy of data mining tool for online prediction of transient stability [J]. IEEE Transactions on Power Systems, 2014, 29(1): 377-385.
- [13] TSO S K, GU X P, ZENG Q Y, et al. An ANN-based multilevel classification approach using decomposed input space for transient stability assessment[J]. Electric Power Systems Research, 1998, 46(3): 259-266.
- [14] 尹雪燕, 闫炯程, 刘玉田, 等. 基于深度学习的暂态稳定评估与严重程度分级[J]. 电力自动化设备, 2018, 38(5): 64-69.  
YIN Xueyan, YAN Jiongcheng, LIU Yutian, et al. Deep learning based transient stability assessment and severity grading[J]. Electric Power Automation Equipment, 2018, 38(5): 64-69.
- [15] 叶圣永, 王晓茹, 刘志刚, 等. 基于支持向量机增量学习的电力系统暂态稳定评估[J]. 电力系统自动化, 2011, 35(11): 15-19.  
YE Shengyong, WANG Xiaoru, LIU Zhigang, et al. Power system transient stability assessment based on support vector machine incremental learning method [J]. Automation of Electric Power Systems, 2011, 35(11): 15-19.
- [16] WANG B, FANG B, WANG Y, et al. Power system transient stability assessment based on big data and the core vector machine[J]. IEEE Transactions on Smart Grid, 2016, 7(5): 2561-2570.
- [17] ZHANG Rui, XU Yan, DONG Zhaoyang, et al. Post-disturbance transient stability assessment of power systems by a self-adaptive intelligent system [J]. IET Generation Transmission & Distribution, 2015, 9(3): 296-305.
- [18] JAMES J Q, HILL D J, LAM A S, et al. Intelligent time-adaptive transient stability assessment system [J]. IEEE Transactions on Power Systems, 2018, 33(1): 1049-1058.
- [19] 吴为, 汤涌, 孙华东, 等. 基于广域量测信息的电力系统暂态稳定研究综述[J]. 电网技术, 2012, 36(9): 81-87.  
WU Wei, TANG Yong, SUN Huadong, et al. A survey on research of power system transient stability based on wide-area measurement information [J]. Power System Technology, 2012, 36(9): 81-87.
- [20] 李倩倩, 刘霄影. 多类类别不平衡学习算法: EasyEnsemble. M [J]. 模式识别与人工智能, 2014, 27(2): 187-192.  
LI Qianqian, LIU Xuying. EasyEnsemble. M for multiclass imbalance problem [J]. Pattern Recognition and Artificial Intelligence, 2014, 27(2): 187-192.
- [21] FAN W, STOLFO S J, ZHANG J, et al. AdaCost: misclassification cost-sensitive boosting [C]// Sixteenth International Conference on Machine Learning, June 27-30, 1999, Bled, Slovenia: 97-105.
- [22] ZHU L, LU C, DONG Z Y, et al. Imbalance learning machine-based power system short-term voltage stability assessment[J]. IEEE Transactions on Industrial Informatics, 2017, 13(5): 2533-2543.
- [23] BATISTA G E, 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.
- [24] 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.
- [25] CHEN Y, WANG Y, KIRSCHEN D, et al. Model-free renewable scenario generation using generative adversarial networks[J]. IEEE Transactions on Power Systems, 2018, 33(3): 3265-3275.
- [26] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein generative adversarial networks[C]// International Conference on Machine Learning, August 6-11, 2017, Sydney, Australia: 214-223.
- [27] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets [C]// Advances in Neural Information Processing Systems, December 8-13, 2014, Montreal, Canada: 2672-2680.
- [28] MIRZA M, OSINDERO S. Conditional generative adversarial nets[R/OL]. [2014-11-04]. <https://arxiv.org/abs/1411.1784>.
- [29] NAIR V, HINTON G E. Rectified linear units improve restricted Boltzmann machines[C]// Proceedings of the 27th International Conference on Machine Learning, June 21-24, 2010, Haifa, Israel: 807-814.
- [30] DAUPHIN Y, DE VRIES H, BENGIO Y. Equilibrated adaptive learning rates for non-convex optimization [C]// Advances in Neural Information Processing Systems, December 7-12, 2015, Montreal, Canada: 1504-1512.
- [31] HUANG G B, ZHU Q Y, SIEW C K. Extreme learning machine: theory and applications[J]. Neurocomputing, 2006, 70(1/2/3): 489-501.
- [32] XU Y, DONG Z Y, MENG K, et al. Real-time transient stability assessment model using extreme learning machine[J]. IET Generation Transmission & Distribution, 2011, 5(3): 314-322.
- [33] ZHANG Y, XU Y, DONG Z Y, et al. Intelligent early warning of power system dynamic insecurity risk: toward optimal accuracy-earliness tradeoff[J]. IEEE Transactions on Industrial Informatics, 2017, 13(5): 2544-2554.
- [34] ABADI M, BARHAM P, CHEN J, et al. TensorFlow: a



- system for large-scale machine learning [C]// OSDI, November 2-4, 2016, Savannah, USA: 265-283.
- [35] CHOLLET F. Keras: Theano-based deep learning library[EB/OL]. [2018-05-09]. <https://github.com/keras-team/keras>.
- [36] PAI M A. Energy function analysis for power system stability [M]. New York: Springer Science & Business Media, 1989: 223-227.
- [37] SIEMENS. PSS/E [EB/OL]. [2018-05-09]. <https://www.siemens.com/global/en/home/products/energy/services/transmission-distribution-smart-grid/consulting-and-planning/pss-software/pss-e.html>.
- [38] CHEN T, HE T, BENESTY M. XGBoost[EB/OL]. [2018-05-09]. <http://xgboost.readthedocs.io/en/latest/>.
- [39] HE M, VITTAL V, ZHANG J. Online dynamic security assessment with missing PMU measurements: a data mining approach[J]. IEEE Transactions on Power Systems, 2013, 28(2): 1969-1977.
- [40] BLAGUS R, LUSA L. SMOTE for high-dimensional class-imbalanced data[J]. BMC Bioinformatics, 2013, 14(1): 106.
- 谭本东(1994—),男,硕士研究生,主要研究方向:数据挖掘在电力系统中的应用。E-mail: 15072475536@163.com
- 杨军(1977—),男,通信作者,教授,主要研究方向:电力系统运行与控制、电动汽车等。E-mail: JYang@whu.edu.cn
- 赖秋频(1995—),男,硕士研究生,主要研究方向:数据挖掘在电力系统中的应用。E-mail: 846472189@qq.com
- (编辑 章黎)

## Data Augment Method for Power System Transient Stability Assessment Based on Improved Conditional Generative Adversarial Network

TAN Bendong<sup>1</sup>, YANG Jun<sup>1</sup>, LAI Qiupin<sup>1</sup>, XIE Peiyuan<sup>2</sup>, LI Jun<sup>2</sup>, XU Jian<sup>1</sup>

(1. School of Electrical Engineering, Wuhan University, Wuhan 430072, China;

2. State Grid Hunan Electric Power Company, Changsha 410000, China)

**Abstract:** Data-driven transient stability assessment method has become the focus of research in the field of power network security. However, transient unstable situation in the actual power system is very rare, which brings great difficulties to the data acquisition method for judging the instability. This paper proposes a data augment method for the synthesis of unstable samples in the transient stability assessment. It enhances the adaptability of training methods for conditional generative adversarial network (CGAN) to improve their learning stability and uses the improved CGAN training generators and discriminators during offline training to learn the distribution characteristics of raw data. Then, the extreme learning machine (ELM) classifier is used to filter out the generated samples with the highest G-mean value among the multiple sets of samples generated by the improved CGAN. The unstable samples are used to augment the original unstable samples, and the augmented original samples are used to train the classifier to achieve online transient stability assessment. The simulation results show that the proposed method can effectively learn the distribution characteristics of the original data by the improved CGANs. The method has the advantages of strong anti-noise interference and good robustness to high-dimensional data, and it can effectively balance the unstable data of power system.

This work is supported by National Key R&D Program of China (No. 2017YFB0902900).

**Key words:** power system; transient stability assessment; data augment; conditional generative adversarial network (CGAN); G-mean value




Journal of Modern Power  
Systems and Clean Energy

<http://www.mpce.info>  
<http://www.springer.com/40565>

For further information, please contact:  
 State Grid Electric Power Research Institute Press  
 No.19 Chengxin Avenue, Nanjing 211106, Jiangsu Province, China  
 Tel: 86-25-81093060 Fax: 86-25-81093040 E-mail: mpce@alljournals.cn

Sponsored by : State Grid Electric Power Research Institute, China




## 附录 A

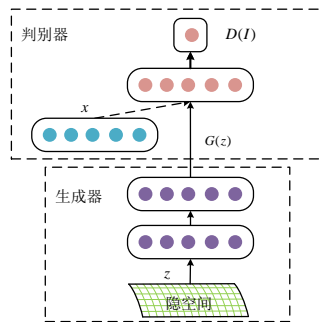
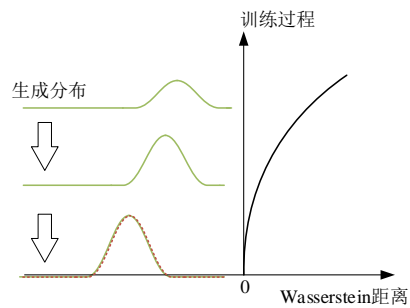


图 A1 对抗生成神经网络

Fig.A1 Generative adversarial network



绿色实线为生成器生成分布，黄色虚线为实际数据分布

图 A2 对抗生成神经网络训练过程示意图

Fig.A2 Generative adversarial network training process diagram

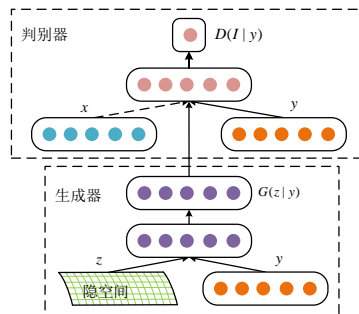


图 A3 改进条件生成对抗神经网络

Fig.A3 Improved Conditional generative adversarial network

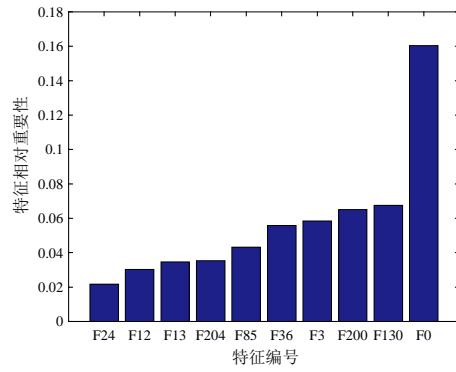


图 A4 特征相对重要性分布

Fig.A4 Feature relative importance distribution

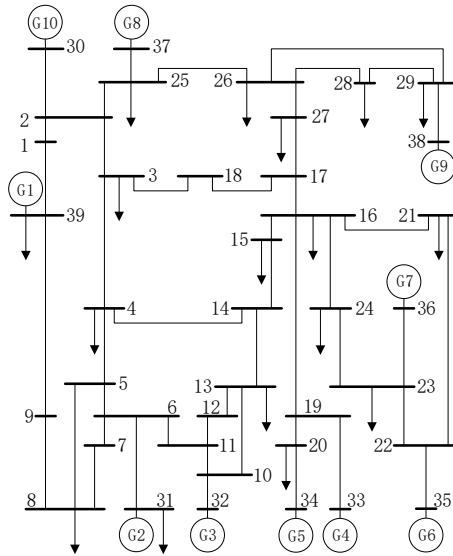


图 A5 新英格兰 10 机 39 节点电力系统

Fig.A5 New England 10-machine-39-bus power system

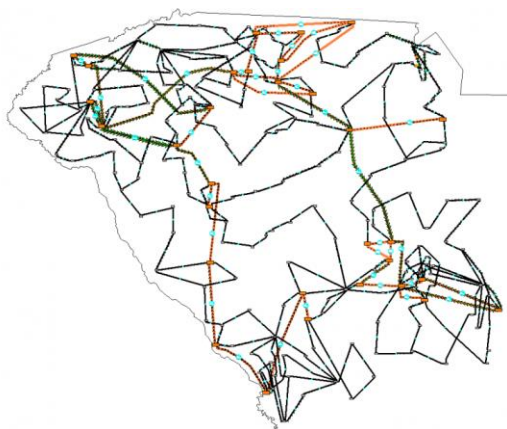


图 A6 南卡罗来纳州 500 节点系统接线图

Fig.A6 SouthCarolina 500-Bus System wiring diagram

表 A1 模型重要参数设置

Table A1 Model important parameter settings

模型参数	数值
判别器预训练次数	100
每轮判别器训练次数	5
每轮生成器训练次数	1
小批量训练样本数目	128
学习率	5e-4
训练轮数	10000
ELM 隐含层节点数	40