

# 生成对抗网络及其新能源数据质量应用研究综述

李洋<sup>1</sup>, 肖泽青<sup>2\*</sup>, 聂松松<sup>1</sup>, 曹军威<sup>2</sup>, 华昊辰<sup>2</sup>

(1. 国家电网有限公司大数据中心, 北京 100084; 2. 北京信息科学与技术国家研究中心, 清华大学, 北京 100084)

**摘要:** 由于新能源数据的随机性和波动性等特点, 并网数据存在的数据缺失、重复、异常以及分布不均匀等数据质量问题变得越来越突出, 针对数据质量的评估和治理等研究对新能源的发展具有重要的积极意义。传统数据质量研究手段不适用于解决新能源数据质量问题, 而人工智能算法在处理这方面问题上具有无可比拟的优点。生成对抗网络(generative adversarial networks, GAN)是人工智能领域近年来最热门的研究方向之一, 其出色的数据生成能力受到了广泛的关注。本文首先介绍了经典 GAN 的架构、优势、存在的问题及其改进综述; 接着对 GAN 在新能源方面的应用进行了综述; 然后阐述了 GAN 在新能源数据质量方面的应用; 最后对本文进行了总结, 并展望了未来 GAN 在新能源方面的可能应用。

**关键词:** 新能源; 数据质量; 生成对抗网络 (GAN)

## The Review of Research on Generative Adversarial Network and New Energy Data Quality

LI Yang<sup>1</sup>, XIAO Zeqing<sup>2\*</sup>, NIE Songsong<sup>1</sup>, CAO Junwei<sup>2</sup>, HUA Haochen<sup>2</sup>

(1. Large Data Center of State Grid Corporation of China, Beijing 100084, China; 2. Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing 100084, China)

**Abstract:** Due to the randomness and volatility of new energy data, data quality problems such as lack of data, repetition, abnormality and uneven distribution of grid-connected data have become more and more prominent. Studies on data quality evaluation and governance have important and positive significance for the development of new energy. Traditional data quality research methods are not suitable for solving new energy data quality problems, while artificial intelligence algorithms have incomparable advantages in dealing with this problem. Generative adversarial networks (GAN) is one of the hottest research directions in the field of artificial intelligence in recent years, and its excellent data generation ability has attracted wide attention. Firstly, this paper introduces the framework, advantages, existing problems and improvement of classic GAN. Then, the application of GAN in new energy is reviewed. Finally, this paper summarizes the paper and looks forward to the possible application of GAN in new energy.

**Key words:** new energy; data quality; generative adversarial networks (GAN)

### 0 引言

早期的数据质量研究主要集中于开发用于查询多个数据源和构建大型数据仓库的技术[1]。最初, 研究人员提出用元数据来表示数据质量以方便数据质量管理[2]。文献[3]认为数据质量是数据满足用户需求的能力。数据质量评估的核心在于如何具体地评估各个维度, 目前方法主要分成两类: 定性的策略和定量的策略[4]。到目前为止, 还没有关于数据质量的精确定义[5][6], 也没有形成系统化的数据质量评估指标与治理标准[5]。同时, 现有的数据质量治理方法存在精确性和效率不高等问题。

随着分布式新能源的蓬勃发展, 大规模新能源的监测数据日益丰富, 但由于风电/光伏等新能源并网监测数据具有间歇性、周期性、随机性和波动性等特点[7][8], 其并网数据存在数据缺失、重复、异常以及分布不均匀等数据质量问题, 严重影响了数据特征提取及数据挖掘的准确性和高效性[9]。因此, 开展大规模分布式新能源数据质量关键技术研究及应用对于引导和优化新能源规划建设, 规范引领新能源行业健康有序发展具有重要意义。

传统方法多靠阈值[10]、线性映射等方法分析数据, 在大规模数据质量治理方面存在严重不足, 如无法短时间内处理大规模信息、运算精度低等,

特别是针对具有随机性、波动性的分布式新能源数据，数据特性更复杂，传统数据治理手段更加不适用，而人工智能算法在处理这方面问题上具有无可比拟的优点。GAN 作为一种新兴的半监督或无监督学习技术而备受关注[11]。GAN 通过对数据的高维分布进行隐式建模来实现学习。GAN 提供了一种无需广泛注释训练数据就能学习深层表示的方法，通过一对网络的竞争过程来获得反向传播信号。2018 年《麻省理工科技评论》将 GAN 列入“全球十大突破性技术”之一[12]。清华大学孙岩团队首次证明量子计算能利用 GAN，其准确率达到 98.8%，创造了量子机器学习的新里程碑。GAN 在生成数据能力上的出色表现正成为新能源数据质量研究的热点。

本文围绕 GAN 在新能源数据质量方面的应用，首先介绍了经典 GAN 的原理；然后综述了 GAN 的理论改进；接着对新能源数据质量研究进行了文献概述；紧接着比较详细地阐述了 GAN 在新能源及新能源数据质量方面的应用；最后展望了 GAN 在新能源数据质量研究上的应用潜力，可为今后 GAN 应用到新能源数据质量研究中提供参考。

## 1 经典 GAN 基本原理

Goodfellow 于 2014 年首次提出的 GAN 是一种新兴的半监督和非监督学习技术[13]，其特点是训练一对相互竞争的网络。GAN 的架构如图 1 所示。G 称为生成器(Generator)，目的是制作逼真的样本，其输入为随机噪声向量(通常服从均匀分布或正态分布)；D 为鉴别器(Discriminator)，接收来自生成器伪造的样本和真实的样本，并旨在将它们区分开来，其输出为样本是真实的而不是虚假的概率。两者都是同时训练的，并且相互竞争(博弈)。当 D 不能确定数据是来自真实数据集还是来自 G 时，即达到所谓的纳什均衡，此时所得到的 G 已经学习到了真实数据的分布，从而掌握整体真实数据的分布情况。

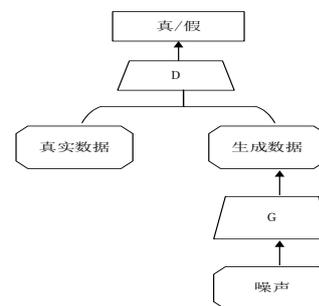


图 1 GAN 结构图[11]

需要强调的是，GAN 不再像传统的监督式深度学习那样需要海量带有类别标记的图像，不需任何图像标记即可训练，也就是进行无监督条件下的深度学习。

## 2 GAN 综述

### 2.1 GAN 的优势

GAN 这种竞争的方式不再要求一个假设的数据分布，而是直接进行取样，能够比其它模型产生更好的样本，从而真正达到理论上可以完全逼近真实数据。GAN 能训练任何一种生成器网络；生成对抗式网络能学习可以仅在与数据接近的细流形上生成点，不需要设计遵循任何种类的因式分解的模型，任何生成器网络和任何鉴别器都会有用。GAN 无需马尔科夫链，仅使用反向传播来获取梯度，学习过程中不需要推理，且模型中可融入多种函数。对抗的网络的另一优点是可表示很尖，甚至退化的分布，而基于马尔科夫链的方法为混合模式而要求模糊的分布[13]。

### 2.2 GAN 的劣势

GAN 无法学习是因为 G 开始退化，学习难以进行，方案之间振荡，或 G 有崩溃的倾向[14]，即 G 坍塌产生一小部分类似样品(部分崩溃)，最坏的情况下只产生一个样品(完全崩溃)[15]。一种解决 GAN 生成样本多样性不够好的方法是增加 G 的多样性，使用多个 GAN 来覆盖概率分布的不同模式[16]；另一种减轻模式崩溃的方法是改变用于比较统计分布的距离度量。

所有的理论都认为 GAN 应该在纳什均衡上有卓越的表现，但梯度下降只有在凸函数的情况下才能保证实现纳什均衡，且寻找纳什均衡解是一个难点[17][18]。当博弈双方都由神经网络表示时，在没有实际达到均衡的情况下，可以让博弈双方对自己策略进行调整来达到纳什均衡。GAN 不需要预先建模使得 GAN 太过自由了，解决办法之一是给 GAN 加约束[18]；还有一个想法就是不要让 GAN 一次完成全部任务，而是一次生成一部分，比如分多次生成一张完整的图片[19]。

### 2.3 GAN 改进综述

针对原始 GAN 的劣势，近年来出现了各种出色的 GAN 改进变体。本节阐述主流的 GAN 改进。

Salimans 等[20]提出的特征匹配、小批量的训练判别器、历史平均和类别标签平滑这几种稳定训练的技术，能够训练以前无法解决的模型。InfoGAN 处理的是图像外观中错综复杂的因素，包括姿态、光照和面部图像的情感内容的变化[21]。给 GAN 加一些约束，可得到条件生成对抗网络 ( Conditional Generative Adversarial Nets, CGAN ) [22]。CGAN 可以解决 GAN 太过自由、本身不可控的缺点。CGAN 模型可以生成以类标签为条件的 MNIST 数字，还可以使用此模型来学习多模态模型[22]。然而，使用一位有效编码，对包含数千个类的标签向量进行条件设置在内存开销和计算时间开销方面是不可行的[23]。

Radford 等人[24]提出的深度卷积 GAN(deep convolutional GAN, DCGAN)使得 G 和 D 都可以学习良好的上采样和下采样操作，有助于提高图像合成的质量。而且 G 表现得更加可控和稳定。DCGAN 将批量规范化(batch normalization, BN)用在了 G 和 D 上，能够避免崩溃。Radford 等人[24]研究表明，在 D 中间层之间使用泄漏 ReLU 激活函数比使用常规 ReLU 具有更好的性能。此外，DCGAN 还可以支持图像“语义”层面的加减。但 DCGAN 难以训练，训练过程只能是启发式的[25]。虽然 BN 能够在早期

阶段加速训练，但可能会对训练模型的质量和训练过程的稳定性产生负面影响。在此基础上，提出了一种新的权重归一化方法，该方法可以提高神经网络的重构能力和训练速度，特别是神经网络的稳定性，因此在 GAN 训练中应采用权重归一化方法来代替 BN[26]。

Akbari 等人[27]提出了用于序列数据生成的半递归混合 VAE-GAN 模型。提出的两种测试框架用来合成任意帧数的序列。在数据集 CIFAR-10 上进行的早期实验表明，与用于监督学习的网络相比，使用具有相同容量和表示能力的 CNNs 训练 G 和 D 网络更加困难。建立在 GAN 和 CGAN 基础上的拉普拉斯金字塔 GAN ( Laplacian Generative Adversarial Networks, LAPGAN ) 为这个问题提供了一个解决方案[19]。然而，训练 GAN 需要用连续的高维参数确定非凸博弈的纳什均衡。当用于寻求纳什均衡时，LAPGAN 算法可能无法收敛[20]。文献[28]提出了一个可通过对抗性训练进行训练的递归生成模型(recurrent generative model, GRAN)。Mou 等人[29]设计了一种具有 PReLU 的 GRU，它能够处理长期序列数据(比如高光谱序列)，并且更适合于少量训练样本。

文献[30]通过强制 D 网络输出类标签，将 GAN 扩展到半监督上下文。该方法可用于建立一个更高效的数据分类器，并能生成比常规 GAN 更高质量的样本。Wu 等人[31]利用体积卷积网络和 GAN 的最新进展，研究了三维对象生成问题，提出了三维 GAN(3D-GAN)。文献[32]采用 Wasserstein 距离的近似，替代了 GAN 的成本函数，提出了 Wasserstein GAN (WGAN)。与原始的 GAN 成本函数不同，WGAN 更可能提供对更新 G 有用的梯度。虽然 WGAN 在稳定训练方面取得了进展，但有时仍然只能生成较差的样本或无法收敛。Gulrajani 等人[33]提出了一种裁剪权值的替代方法，比标准 WGAN 表现更好，并且能够在几乎没有超参数调整的情况下对各种 GAN 架构进行稳定的训练。基于梯度的

随机裁剪能够准确地恢复潜在向量，对噪声具有较强的鲁棒性。随机裁剪比标准裁剪更精确、更健壮[34]。采用全卷积神经网络 (fully convolutional network, FCN) [35]能让 GAN 自己学习更合适的空间下采样方法，应用在图像彩色化和图像检测等方面取得了较好的效果[36]。为了解决生成的图像无效的问题，Heusel 等人[37]使用真实和生成图像的激活之间的 Wasserstein-2 距离作为 Frechet 初始距离得数。

Mescheder 等人[38]提出的统一变分自编码器 (variational autoencoders, VAEs)，以对抗型的变分贝叶斯 (adversarial variational bayes, AVB) 框架的形式进行对抗性训练，使推理模型更加灵活，保留了标准变分自编码器的大部分优点，易于实现。Ian Goodfellow 的 NIPS 2016 教程[39]也提出了类似的想法。AVB 尝试优化与 VAEs 相同的标准，但使用对抗训练目标而不是 Kullback-Leibler (K-L)散度。AVB 明确规定了要使用散度量，因此无法应用 WGAN[38]。

在条件熵的框架下，文献[40]提出了对抗性和非对抗性两种方法来学习无监督和有监督任务的理想匹配联合分布，将各种 GAN 模型统一理解为节点匹配的观点来解决双向对抗性网络的不可识别性问题，稳定了无监督双向对抗性学习方法的学习。然而，这种模式与 AVB 和对抗自动编码器 (adversarial autoencoder, AAE)有很多共同之处。这些类似于 VAEs 的自动编码器的潜在空间是通过使用对抗训练而不是编码样本和先验样本之间的 K-L 散度来调节的[41]。

Gurumurthy 等人[42]提出将潜在的生成空间作为混合模型重新参数化，建模为高斯混合，并与 GAN 模型一起学习混合模型的参数。这种基于 GAN 的新的体系结构——DeLiGAN 尽管只使用有限的数据进行训练，但生成的样本具有多样性，这并不一定意味着给定的所有图像都具有良好的质量。如果生成器在一个类中为所有类合成了各种各样但

完美的图像，那么得分就会失败[43]。

文献[44]研究表明基于生成式卷积网络的 STGConvNet 不仅可以自动合成动态纹理，同时还可以合成声音。文献[45]介绍了利用任意一种预先训练的 GAN 将图像样本投影到潜在空间的技术，并给出了计算图。该方法以期望的输出为优化目标，虽然非常简单，但速度很慢[46]。Dumoulin 等人[14]引入了对抗学习推理(adversarially learned inference, ALI)模型，该模型利用逆过程对生成网络和推理网络进行联合学习。使用双向 GAN(bidirectional generative adversarial networks, BiGAN)作为学习逆映射的一种方法，所得到的学习特征表示对于辅助监督辨别任务是有用的[47]。然而，观察到的重建结果往往只是模糊地类似于输入，而且往往只是语义上的，而不是术语上的像素值[38]。

Sønderby 等[48]在超分辨率技术上，应用 GAN 对结果进行优化，通过在将样本送入 D 之前向样本添加噪声来挑战 D，得到了比对比方法更好的结果，能还原打码图片。但该超分辨率技术只能为低分辨率图像添加有限的细节，不能像 StackGAN[49]那样纠正较大的缺陷。在数据样本中添加噪声以稳定训练的过程后来被 Arjovsky 等人[17]正式证明是正确的。文献[50]提出了一种直接从数据中提取策略的通用框架，就好像它是通过反强化学习后的强化学习得到的一样。

### 3 GAN 在新能源方面的应用

场景生成有助于模拟可再生能源发电的不确定性和可变性，是可再生能源高渗透率电网决策的重要工具[51][52]。文献[51]将 GAN 应用于可再生资源的场景生成，利用深度神经网络和大量历史数据的能力来执行直接生成符合相同历史数据分布的场景的任务。应用该模型建立的实例研究表明，GAN 对风力发电和太阳能发电都能很好地进行场景生成。这是一种数据驱动的、无模型的方法，它直接从历史数据生成新的场景，而不需要显式地拟合概率模

型。这种方法易于扩展，并且优于现有的基于模型的方法，特别是在多个相关可再生发电机的设置中。但如果在历史数据中包含几种不同的模式（例如，大风和低风），则生成的场景可能包含这些模式的混合，而这些“混合”场景可能不适合随后的优化问题，需要通过额外的后处理步骤进行过滤。为了克服这一挑战，基于贝叶斯 GAN [53]，文献[52]扩展了[51]的结果，成功捕获了场景集群中的固有多样性并重现它们中的每一个。

文献[54]首先应用 GAN 以学习历史可再生能源发电数据中的内在模式，然后通过解决优化问题，快速生成大量真实、高质量的未来场景。该方法无需任何显式的模型构建即可捕获可再生能源的时空行为。为了考虑风光不确定性给微网运行带来的风险，针对独立型微网的容量优化配置，李康平等[55]提出了一种基于 GAN 场景模拟和条件风险价值 (conditional value at risk, CVaR) 的容量随机优化配置模型。仿真算例表明，相比于仅采用典型年风光资源数据进行配置的传统方法，文中提出的模型对于规划周期内可能出现的运行场景适应性更好。

#### 4 新能源数据质量研究文献概述

国内外对新能源电网数据质量做了大量研究 [56-71]。文献[56]对建立供电企业的数据质量指标体系进行了综述。针对数据的失真问题，叶锋等[72]结合负荷预测特点的总加值数据多源处理技术，将单测点的多源处理推广到多测点的计算公式中，并提出了能量管理系统 (energy management system, EMS) 中的“动态多源技术”，有效地改进了最终的负荷预测精度。在电力系统数据质量的研究方面，对数据质量检测，错误数据修正与精度，以及整体数据质量评估的综合研究尚不完善 [73]。结合电力数据对数据质量评估的需求，文献[74]利用层次分析法确定评估指标的权值，建立了一个统计数据质量评估模型。但评估结果存在较大的主观性。

在电力系统不良数据/坏数据研究方面，国内外

学者做了一些工作 [66] [68][70][73] [74][75]。目前的研究主要集中于对不良数据的检测和辨识 [57][64][76]，其方法大致分为：1) 传统的不良数据检测方法，如加权残差法 [63]、量测量突变检测法 [62]、目标函数极值检测法 [61] 等；2) 相对较新的检测方法，如模糊等价矩阵聚类分析法 [60]、神经网络法 [59] 和云技术 [66] 等。虽然基于残差类的方法对不良数据的检测与辨识有明显的效果，但存在残差污染等弊端 [75]。文献 [75] 采用基于量测量相关性的方法对电网不良数据进行检测与辨识。基于多源多时空信息的电压数据质量综合检测和修正方法能有效检测出不满足精度要求的电压数据，并对其进行准确修正，同时可为运行人员评估数据质量提供参考 [73]。文献 [65] 利用极限学习机 (extreme learning machine, ELM) 对微电网历史数据进行学习以提取数据特征；进而利用密度聚类算法 (density-based spatial clustering of applications with noise, DBSCAN) 聚类算法分析特征量以识别不良数据，提出了一种基于 ELM 和 DBSCAN 的微电网不良数据检测方法。

虚假数据注入 (false data injection, FDI) 攻击通过破坏网络数据完整性以干扰控制决策，是一种典型的网络攻击方式 [77]。Singh S K 等 [70] 提出的一种基于联合变换的虚假数据注入 (false data injection, FDI) 攻击实时检测方案对 FDI 攻击具有较高的检测概率。但该检测方案预训练的时间较深度学习长 [78]。Sedghi 等人 [67] 提出了一种基于母线相位角马尔可夫图的分散式 FDI 检测方案，其补救措施是第一个全面检测这种复杂攻击，它不需要额外的硬件。而且，无论被攻击子集的大小，它都是成功的。此外用于 FDI 检测的方法还包括稀疏优化方法 [79]、双层模型 [80] 等。

对缺失数据进行数据填充保证了电网业务数据的真实性、完整性、准确性，促进了智能电网大数据的发展 [81]。马尔科夫链蒙特卡罗算法可用于电网调度控制系统缺失数据填补 [81] [82]，最佳数据参数的优化算法包括遗传算法 [81] 和混沌遗传优化

算法[82]等。文献[83]提出了一种基于自适应 BP 神经网络的模型来填充缺失的风电数据，解决了风机数据丢失的问题，有效地提高了数据的质量。

电网地理信息系统(geographic information system, GIS)基础数据质量是所有数据处理和分析的前提，对开展 GIS 业务应用、提高系统实用化水平有重大意义[84]。方国锋[85]分别就 GIS 基本理论、电网 GIS 数据采集的内容、特性、质量控制需求等方面进行分析，提出了一套电网 GIS 数据采集质量评价体系，实现了电网 GIS 数据采集工作的质量控制和自动化处理。文献[86]提出了一种基于电网 GIS 平台的电网资源图形数据质检工具，能快速有效地定位图形问题所在、跟踪数据清洗质量，提升数据清洗效率的同时，保证数据成果质量。

新能源数据通常价值密度较低，异常数据，比如故障数据很少，但却是很多算法学习所需的关键。解决小样本问题一般是通过改进算法来使其更适应于小样本环境[87]。另一种思路是对小样本数据进行扩充，其传统方法包括 Bootstrap 方法[88]、支持向量机[89] [90]和专业软件仿真[91]等。文献[91]选择在电网仿真软件 BPA (Bonneville Power Administration)中搭建某城市真实电网的仿真模型来进行仿真数据的生成，并对文中提出的基于深度强化学习的能源互联网无功决策优化算法在仿真数据上进行了实验验证。

## 5 GAN 在新能源数据质量方面的应用研究

智能电网中的大数据虽然体量巨大，但是质量往往不高[92]。电力大数据通常价值密度较低，即所采集的绝大部分数据为正常样本数据，异常数据很少但却是深度学习所需的关键[93]。因此，解决小样本学习问题是深度学习在智能电网中深入应用的关键。杨懿男等[87]提出了一种基于改进 GAN 训练样本生成的技术，以解决小样本数据问题。通过对已有样本数据的学习，利用 FCN 强大的特征提取能力提高 GAN 的学习效果，最终能够自主生成全

新的样本数据。该方法不需要领域内其他信息辅助，模型能够自主学习样本特征，生成的样本更贴近真实数据，很好地解决了当前电网数据真实样本不足的问题。

文献[94]提出了一种基于 GAN 的负荷场景生成方法，为避免原始 GAN 训练不稳定的问题，采用基于 DCGAN 架构及 J-S 散度作为目标函数对模型进行训练。为测试生成负荷序列的质量，针对多样性以及锐度的要求，提出 TSTR 以及 TRTS 指标对序列进行评判，并基于支持向量回归算法(support vector regression, SVR)对上述指标进行测试，实验结果表明，随着训练的逐步进行，GAN 产生的数据质量逐渐提高，可以产生足够真实且满足多样性的负荷序列。

针对小样本环境下异构数据资源用于智能配电和利用的低利用率问题，文献[95]提出了一种基于 GAN 的智能配电与利用异构数据集成技术。该方法通过引入 GAN 理论，根据所有测量指标都已完成的目标样本，扩展了样本空间。然后，构建了一个新的峰聚类模型，实现了扩展样本空间的有限全覆盖，并修复了那些不完整的样本以消除异构特征。最后，使用训练有素的 GAN 鉴别器检验修复后的样本，实现异构数据的集成，实现智能配电和用电。文献[95]成功地将 GAN 学习和聚类理论相结合，提供了一种新颖的异构数据集成，为配电系统进一步的数据质量管理提供了一个新的视角。

基于能源互联网架构，针对自能源网络结构复杂且状态变量采样周期不同所造成的信息不完备等特点，孙秋野等[96]提出了一种基于 GAN 技术的数据-机理混合驱动方法对自能源模型进行参数辨识。将 GAN 模型中训练数据与专家经验结合进行模糊分类，解决了自能源在不同运行工况下的模型切换问题。通过应用含策略梯度反馈的改进 GAN 技术对模型进行训练，解决了自能源中输出序列离散的问题。

## 6 总结与展望

### 6.1 GAN 在新能源方面的应用展望

GAN 方法是数据驱动和无模型的,不需要任何特定的统计假设。它利用深度神经网络和大量历史数据的能力来执行直接生成符合相同历史数据分布的任务,而不需要对分布进行显式建模。因此它可以应用于电力系统中大多数感兴趣的随机过程。

由于 GAN 方法采用前馈神经网络结构,不需要对可能复杂的高维过程进行采样,可以方便地扩展到具有大量不确定性的系统。在未来的工作中,建议将 GAN 扩展到可再生能源发电的高渗透决策策略设计中[51][52]。此外,各种 GAN 的出色的变体还没有广泛应用到新能源数据质量研究的应用之中,跟踪 GAN 的改进并将其应用于新能源数据质量领域中是一种新颖而又有挑战性的尝试,对新能源数据质量的研究具有积极的推动作用。

### 6.2 总结

基于新能源并网的特点,开展新能源数据质量关键技术研究及应用,对于引导和优化新能源规划建设,规范引领新能源行业健康有序发展具有重要意义。传统数据治理手段不适用于解决新能源并网数据质量问题,而人工智能算法在处理这方面问题上具有无可比拟的优点。

GAN 作为近年来人工智能领域的新秀,其出色的数据生成能力在新能源数据质量研究中正在成为新的研究热点。本文首先介绍了 GAN 的经典理论及其优缺点,并综述了 GAN 的改进;接着对 GAN 在新能源领域的应用进行了总结;然后,对 GAN 在新能源数据质量研究方面的应用进行了综述;最后,讨论了 GAN 在新能源研究中的应用潜力。

**致谢** 本工作受到以下项目的资助:国家电网有限公司大数据中心项目(面向大规模分布式新能源并网数据的质量评估与治理技术研究);国家重点研发计划项目(项目编号 2017YFE0132100);北京信息科学与技术国家研究中心项目(项目编号

BNR2019TD01009)。在此一并致以衷心的感谢!

## 参考文献

- [1] Madnick S E, Wang R Y, Lee Y W, et al. Overview and Framework for Data and Information Quality Research[J]. *Journal of Data & Information Quality*, 2009, 1(1):1-22.
- [2] 郭志懋,周傲英. 数据质量和数据清洗研究综述[J]. *软件学报*, 2002, 13(11):2076-2082.
- [3] Cappiello C, Francelanci C, Pernici B. Data quality assessment from the user's perspective[C]// *International Workshop on Information Quality in Information Systems*. ACM, 2004.
- [4] 韩京宇,徐立臻,董逸生. 数据质量研究综述[J]. *计算机科学*, 2008, 35(2):1-5.
- [5] 黄刚,袁满,吴秀英, et al. 元数据驱动的数据质量评估体系架构研究[J]. *计算机工程与应用*, 2013, 49(8):114-119.
- [6] 杨筱娟,高丽颖,王青松. 在线式多晶制绒清洗设备数据质量控制[J]. *电子工业专用设备*, 2011(11).
- [7] 黄强,吕振华,韩华春,等. 光伏电站并网运行典型问题分析[J]. *电气开关*, 2018, 56(4):29-37.
- [8] 肖泽青,华昊辰,曹军威. 人工智能在能源互联网中的应用综述[J]. *电力建设*, 2019, 40(05):67-74.
- [9] 李伟花. 考虑时空分布特性的区域风电功率预测方法[D]. 2015.
- [10] 董述林,文福拴,陈亮. 电力负荷数据预处理的二维小波阈值去噪方法[J]. *电力系统自动化*, 2012, 36(2):101-105.
- [11] 张莹莹. 生成对抗网络模型综述[J]. *电子设计工程*, 2018.
- [12] 刘淑,周永春. 《麻省理工科技评论》发布 2018 年“全球十大突破性技术”[J]. *科技中国*, 2018.
- [13] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in *Advances in Neural Information Processing Systems*, 2014, pp. 2672–2680.
- [14] Dumoulin V, Belghazi I, Poole B, et al. Adversarially Learned Inference[J]. 2016.
- [15] Arora S, Ge R, Liang Y, et al. Generalization and Equilibrium in Generative Adversarial Nets (GANs)[J]. 2017.
- [16] Tolstikhin I, Gelly S, Bousquet O, et al. AdaGAN: Boosting Generative Models[J]. 2017.
- [17] M. Arjovsky and L. Bottou, “Towards principled methods for training generative adversarial networks,” *NIPS 2016 Workshop on Adversarial Training*, 2016.
- [18] Wang B, Liu K, Zhao J. Conditional generative adversarial networks for commonsense machine comprehension[C]// *International Joint Conference on Artificial Intelligence*. AAAI Press, 2017.
- [19] Denton E, Chintala S, Szlam A, et al. Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks[J]. 2015.
- [20] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, “Improved techniques for training gans,” in *Advances in Neural Information Processing Systems*, 2016, pp. 2226–2234.
- [21] X. Chen, Y. Duan, R. Houthoofd, J. Schulman, I. Sutskever, and P. Abbeel, “Infogan: Interpretable representation learning by information maximizing generative adversarial nets,” in *Advances in*

- Neural Information Processing Systems, 2016.
- [22] Mirza M , Osindero S . Conditional Generative Adversarial Nets[J]. Computer Science, 2014:2672-2680.
- [23] Mao X , Li Q , Xie H , et al. Least Squares Generative Adversarial Networks[J]. 2016.
- [24] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in Proceedings of the 5th International Conference on Learning Representations (ICLR) - workshop track, 2016.
- [25] Arjovsky M, Bottou L. Towards Principled Methods for Training Generative Adversarial Networks[J]. Stat, 2017, 1050.
- [26] Xiang S, Li H. On the effect of Batch Normalization and Weight Normalization in Generative Adversarial Networks[J]. 2017.
- [27] Akbari M , Liang J . Semi-Recurrent CNN-based VAE-GAN for Sequential Data Generation[J]. 2018.
- [28] Im D J , Kim C D , Jiang H , et al. Generating images with recurrent adversarial networks[J]. 2016.
- [29] Mou L, Ghamisi P, Zhu X X. Deep Recurrent Neural Networks for Hyperspectral Image Classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(7):3639-3655.
- [30] Odena, Augustus. Semi-supervised learning with generative adversarial networks. arXiv preprint arXiv:1606.01583, 2016.
- [31] Wu J , Zhang C , Xue T , et al. Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling[J]. 2016.
- [32] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," in Proceedings of The 34th International Conference on Machine Learning, 2017.
- [33] Gulrajani I , Ahmed F , Arjovsky M , et al. Improved Training of Wasserstein GANs[J]. 2017.
- [34] Lipton Z C, Tripathi S. Precise Recovery of Latent Vectors from Generative Adversarial Networks[J]. 2017.
- [35] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 39(4):640-651.
- [36] 林家骏, 诸葛晶晶, 张晴. 基于像素级生成对抗网络的复杂场景灰度图像彩色化[J]. 计算机辅助设计与图形学学报, 2019, 31(03):93-100.
- [37] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In Advances in Neural Information Processing Systems, pages 6626–6637, 2017. 6,7
- [38] Mescheder L, Nowozin S, Geiger A. Adversarial Variational Bayes: Unifying Variational Autoencoders and Generative Adversarial Networks[J]. 2017.
- [39] Goodfellow I. NIPS 2016 Tutorial: Generative Adversarial Networks[J]. 2016.
- [40] C. Li, H. Liu, C. Chen, Y. Pu, L. Chen, R. Henao, and L. Carin, "Towards understanding adversarial learning for joint distribution matching," in Advances in Neural Information Processing Systems, 2017.
- [41] Creswell A, White T, Dumoulin V, et al. Generative Adversarial Networks: An Overview[J]. IEEE Signal Processing Magazine, 2017, 35(1):53-65.
- [42] Gurumurthy S, Sarvadevabhatla R K, Babu R V. DeLiGAN: Generative Adversarial Networks for Diverse and Limited Data[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE Computer Society, 2017.
- [43] Upadhyay Y, Schrater P. Generative Adversarial Network Architectures For Image Synthesis Using Capsule Networks[J]. 2018.
- [44] Xie J, Zhu S C, Wu Y N. Synthesizing Dynamic Patterns by Spatial-Temporal Generative ConvNet[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017:1061-1069.
- [45] Creswell A , Bharath A A . Inverting The Generator Of A Generative Adversarial Network[J]. IEEE Transactions on Neural Networks and Learning Systems, 2016.
- [46] Junyu Luo, Yong Xu, Chenwei Tang, and Jiancheng Lv. Learning inverse mapping by autoencoder based generative adversarial nets. In Derong Liu, Shengli Xie, Yuanqing Li, Dongbin Zhao, and El-Sayed M. El-Alfy, editors, Neural Information Processing, pages 207–216, Cham, 2017. Springer International Publishing.
- [47] Donahue J , Krähenbühl, Philipp, Darrell T . Adversarial Feature Learning[J]. 2016.
- [48] C. K. Sønderby, J. Caballero, L. Theis, W. Shi, and F. Huszár, "Amortised map inference for image super-resolution," in International Conference on Learning Representations, 2017.
- [49] Han Z, Tao X, Li H, et al. StackGAN: Text to Photo-Realistic Image Synthesis with Stacked Generative Adversarial Networks[C]// 2017 IEEE International Conference on Computer Vision (ICCV). 2016.
- [50] Ho J , Ermon S . Generative Adversarial Imitation Learning[J]. 2016.
- [51] Y. Chen, Y. Wang, D. S. Kirschen, and B. Zhang, "Model-free renewable scenario generation using generative adversarial networks," IEEE Transactions on Power Systems, 2018.
- [52] Chen Y , Li P , Zhang B . [IEEE 2018 52nd Annual Conference on Information Sciences and Systems (CISS) - Princeton, NJ, USA (2018.3.21-2018.3.23)] 2018 52nd Annual Conference on Information Sciences and Systems (CISS) - Bayesian renewables scenario generation via deep generative networks[J]. 2018:1-6.
- [53] Y. Saatchi and A. Wilson, "Bayesian GAN," in Advances in Neural Information Processing Systems, 2017, pp. 3625–3634.
- [54] Chen Y, Wang X, Zhang B. An Unsupervised Deep Learning Approach for Scenario Forecasts[J]. 2017.
- [55] 李康平, 张展耀, 王飞, 等. 基于 GAN 场景模拟与条件风险价值的独立型微网容量随机优化配置模型[J]. 电网技术, 2019, 43(5):1717-1725.
- [56] 乡立. 浅谈供电企业的数据质量指标体系研究[J]. 东方企业文化, 2013(4) : 17-19 .
- [57] 雷霆, 朱传柏, 黄太贵, 等. 基于数据平台的调度数据质量辨识方法[J]. 电力系统自动化, 2012, 36(12) : 71-75 .  
Lei Ting ,Zhu Chuanbai ,Huang Taigui ,et al . Quality identification methods for dispatching data based on data platform[J] . Automation of Electric Power Systems , 2012 , 36(12) : 71-75(in Chinese) .
- [58] 张亮. 电力调度数据中心数据质量问题研究[J]. 华东电力, 2009, 37(3) : 403-406 .  
Zhang Liang . Research on data quality for electric power dispatch data centers[J] .East China Electric Power ,2009 ,37(3) :403-406(in Chinese) .
- [59] 蒋德珑, 王克文, 王祥东. 基于模糊等价矩阵聚类分析的不良数据辨识[J]. 电力系统保护与控制, 2011, 39(21) : 1-6 .

- Jiang Delong , Wang Kewen , Wang Xiangdong . Clustering method of fuzzy equivalence matrix to bad-data detection and identification[J] . Power System Protection and Control , 2011 , 39(21) : 1-6(in Chinese) .
- [60] 王兴志, 严正, 沈沉, 等. 基于在线核学习的电网不良数据检测与辨识方法[J]. 电力系统保护与控制, 2012, 40(1): 50-55 .  
Wang Xingzhi , Yan Zheng , Shen Chen , et al . Power grid bad-data detection and identification based on online kernel learning method [J] . Power System Protection and Control , 2012 , 40(1) : 50-55 (in Chinese) .
- [61] 卢志刚, 王浩锐, 孙继凯. 基于灵敏度分析的数据最优筛选与不良数据辨识[J]. 电网技术, 2011, 35(2): 38-42 .  
Lu Zhigang , Wang Haorui , Sun Jikai . Optimal data screening and bad data identification based on sensitive analysis[J] . Power System Technology , 2011 , 35(2) : 38-42(in Chinese) .
- [62] 周小宝, 罗滇生, 毛田, 等. 一种实用的高压配电网不良数据处理方法[J]. 计算机工程与应用, 2014, 50(8): 255-259 .  
Zhou Xiaobao , Luo Diansheng , Mao Tian , et al . Practical method of high voltage distribution network bad data processing[J] . Computer Engineering and Applications , 2014 , 50(8) : 255-259(in Chinese) .
- [63] 卢志刚, 程慧琳, 冯磊, 等. 基于证据融合理论的多不良数据辨识[J]. 电网技术, 2012, 36(1): 123-128 .  
Lu Zhigang , Cheng Huilin , Feng Lei , et al . Multi bad data identification based on evidence fusion theory[J] . Power System Technology , 2012 , 36(1) : 123-128(in Chinese) .
- [64] 刘莉, 翟登辉, 姜新丽. 电力系统不良数据检测与辨识方法的现状与发展[J]. 电力系统保护与控制, 2010, 38(5): 143-147 .  
Liu Li , Zhai Denghui , Jiang Xinli . Current situation and development of the methods on bad-data detection and identification of power system[J] . Power System Protection and Control , 2010 , 38(5) : 143-147(in Chinese) .
- [65] 熊晓琪, 黄鹤鸣, 郝亮亮, et al. 基于 ELM 与 DBSCAN 的微电网不良数据检测方法[J]. 电测与仪表, 2018.
- [66] 王莹. 探究云技术在电力系统不良数据处理中的应用[J]. 赤峰学院学报(自然科学版), 2017(01):42-43.
- [67] Sedghi H , Jonckheere E . Statistical structure learning to ensure data integrity in smart grid[J] . IEEE Transactions on Smart Grid , 2015 , 6(4) : 1924-1933 .
- [68] Po-Yu Chen , Shusen Yang , McCann J A . et al . Detection of false data injection attacks in smart-grid systems[J] . IEEE Communications Magazine , 2015 , 53(2) : 206-213 .
- [69] Panigrahi B K , Khanna K , Joshi A . Data Integrity Attack in Smart Grid: Optimized Attack to Gain Momentary Economic Profit:[J]. Iet Generation Transmission & Distribution, 2016, 10(16):4032-4039.
- [70] Singh S K , Khanna K , Bose R , et al . Joint transformation based detection of false data injection attacks in smart grid[J]. IEEE Transactions on Industrial Informatics, 2017, 14(1):89-97.
- [71] Khanna K , Panigrahi B K , Joshi A . AI-based approach to identify compromised meters in data integrity attacks on smart grid[J]. IET Generation, Transmission & Distribution, 2018, 12(5):1052-1066.
- [72] 叶锋, 何桦, 顾全, et al. EMS 中负荷预测不良数据的辨识与修正[J]. 电力系统自动化, 2006(15).
- [73] 刘科研, 张剑, 陶顺, et al. 基于多源多时空信息的配电网 SCADA 系统电压数据质量检测与评估方法[J]. 电网技术, 2015, 39(11):3169-3175.
- [74] 秦璇. 电力统计数据的质量评估及其异常检测方法研究[D]. 长沙理工大学, 2013.
- [75] 周小宝. 电力系统状态估计不良数据检测与辨识方法研究与应用[D]. 湖南大学, 2013.
- [76] 刘莉, 翟登辉, 姜新丽. 电力系统不良数据检测与辨识方法的现状与发展[J]. 电力系统保护与控制, 2010, 38(5):143-147.
- [77] 王琦, 邵伟, 汤奕, 等. 面向电力信息物理系统的虚假数据注入攻击研究综述[J]. 自动化学报, 2019, 45(1):72-83.
- [78] Wang H , Ruan J , Wang G , et al . Deep Learning Based Interval State Estimation of AC Smart Grids against Sparse Cyber Attacks[J]. IEEE Transactions on Industrial Informatics, 2018 , 14(11): 4766 - 4778.
- [79] Liu L , Esmalifalak M , Ding Q , et al . Detecting False Data Injection Attacks on Power Grid by Sparse Optimization[J]. IEEE Transactions on Smart Grid, 2014, 5(2):612-621.
- [80] Khanna K , Panigrahi B K , Joshi A . Bi-level modelling of false data injection attacks on security constrained optimal power flow[J]. IET Generation, Transmission & Distribution, 2017, 11(14):3586-3593.
- [81] 王一蓉, 王瑞杰, 陈文刚, et al. 基于遗传优化的调控系统缺失数据填补算法[J]. 电力系统保护与控制, 2016, 44(21):182-186.
- [82] 唐良瑞, 王瑞杰, 吴润泽, et al. 面向全景调控统一数据模型的缺失数据填补算法 [J]. 电力系统自动化, 2017(1):25-30,87.
- [83] 许子龙, 邢作霞, 马仕昌. 基于自适应 BP 神经网络的风电数据缺失数据处理[C]// 第十五届沈阳科学学术年会. 0.
- [84] 唐凯, 钟一俊, 吴惠芬, 等. 电网 GIS 基础数据质量管理工具设计与应用[J]. 电力信息与通信技术, 2016, 14(2):98-101.
- [85] 方国锋. 电网 GIS 数据采集质量控制系统设计研究 [D]. 华北电力大学 , 2015.
- [86] 凌笑, 易衍孜. 基于电网 GIS 平台的电网资源图形数据质检工具[J]. 电力信息与通信技术, 2017(2):27-30.
- [87] 杨懿男, 齐林海, 王红, et al. 基于生成对抗网络的小样本数据生成技术研究[J]. 电力建设, 2019, 40(5):71-77.
- [88] 汪晓洋. 小样本下可靠性试验方法与数据处理的研究[D]. 电子科技大学, 2012.
- [89] 葛乐飞. 基于小样本磁链特性的开关磁阻电机位置预估方法研究[D]. 西北工业大学, 2016.
- [90] KAN Yingnan , YANG Zhaojun , LI Guofa . Bayesian reliability modeling and assessment solution for CN machine tools under small-sample data [ J ] . Chinese Journal of Mechanical Engineering , 2015 , 28( 6 ) : 1229-1239 .
- [91] CAO J , ZHANG W , XIAO Z , et al . Reactive power optimization for transient voltage stability in energy Internet via deep reinforcement learning approach [ J ] . Energies , 2019 , 12: 1556 .
- [92] 刘世成, 张东霞, 朱朝阳, 等. 能源互联网中大数据技术思考 [ J ] . 电力系统自动化 , 2016 , 40( 8 ) : 14-21 , 56 .
- [93] 尚宇炜, 马钊, 彭晨阳, 等. 内嵌专业知识和经验的机器学习方法探索(一): 引导学习的提出与理论基础 [ J ] . 中国电机工程学报 , 2017 , 37( 19 ) : 5560-5571 , 5833 .

- 
- [94] 张宇帆, 艾芊, 李昭昱, et al. 基于生成对抗网络的负荷序列随机场景生成方法[J]. 供用电, 2019, 36(01):35-39+98.
- [95] Tan, Y.; Liu, W.; Su, J.; Bai, X. Generative Adversarial Networks Based Heterogeneous Data Integration and Its Application for Intelligent Power Distribution and Utilization. *Appl. Sci.* **2018**, *8*, 93.
- [96] 孙秋野, 胡旌伟, 杨凌霄, et al. 基于GAN技术的自能源混合建模与参数辨识方法[J]. 自动化学报, 2018, v.44(05):136-149.

作者简介:

李洋, 男, 工程师, 研究生, 研究方向为电力大数据、人工智能等;

肖泽青 (1983), 男, 通信作者, 助理研究员, 博士, 研究方向为大数据、人工智能、智能电网、能源互联网,

Email: [xiaozeqing@mail.tsinghua.edu.cn](mailto:xiaozeqing@mail.tsinghua.edu.cn), Tel: 15807711752

聂松松, 男, 工程师, 研究生, 研究方向为电力大数据、人工智能等。

---

收稿日期: 2019-08-