

# 基于 GAN-RL 最优新能源消纳的运行方式样本智能生成方法

李帅虎, 李汉典, 邹谈, 王炜宇, 曹一家

(电网防灾减灾全国重点实验室(长沙理工大学), 长沙 411014)

**摘要:** 针对实际电网运行中仍存在新能源消纳能力不足的问题, 提出了一种基于生成对抗网络(generative adversarial networks, GAN)和强化学习(reinforcement learning, RL)最优新能源消纳的运行方式样本智能生成方法。该方法可以根据少量实际电网的运行方式样本集, 通过深度神经网络建立生成器和判别器模型, 并依托生成对抗网络的目标函数对生成器和判别器进行最大最小对抗训练, 使生成的虚假运行方式样本分布特征趋近真实样本。在生成器训练过程中, 定义了一个关于新能源消纳的奖惩反馈函数, 以对“生成动作”给予量化的优劣响应, 并在此基础上引入强化学习方法。在遵循使累积奖励最多的寻优原则下, 通过基于策略梯度的强化学习方法不断实时优化生成器参数, 智能生成具有最优新能源消纳特征的运行方式样本。所提方法在某地区实际电网进行了测试, 结果表明该方法能够有效地生成满足最优新能源消纳的运行方式样本, 为改善现阶段下新能源消纳不足的运行调度决策提供数据支撑。

**关键词:** 新能源消纳; 生成对抗网络; 强化学习; 智能运行方式生成

## Intelligent Generation Method for Operation Mode Samples of Optimal New Energy Consumption Based on GAN-RL

LI Shuaihu, LI Handian, ZOU Tan, WANG Weiyu, CAO Yijia

(State Key Laboratory of Disaster Prevention & Reduction for Power Grid, Changsha University of Science & Technology, Changsha 411014, China)

**Abstract:** Aiming at the insufficient new energy consumption capacity in actual power grid operation, an intelligent generation method is proposed for operating mode samples of optimal new energy consumption based on generative adversarial networks (GAN) and reinforcement learning (RL). The method is used to establish generator and discriminator models through deep neural networks with a small sample set of actual power grid operation modes. And based on the objective function of GAN, the generator and discriminator can be trained for maximum and minimum adversarial training to generate false operating mode samples, distribution characteristics of which approach the real samples. During the training process of the generator, a reward and punishment feedback function is defined for new energy consumption to provide a quantitative good or bad response to the “generation action”. On this basis, the reinforcement learning method is introduced. Following the optimization principle of accumulating the most rewards, the generator parameters are continuously optimized in real-time by the reinforcement learning method with policy gradients. Operating mode samples are intelligently generated with optimal new energy consumption characteristics. The proposed method is tested in the actual power grid of a certain region. The results show that it can effectively generate operation mode samples that meet the optimal new energy consumption, and provide data support for improving the current operation scheduling decision of insufficient new energy consumption.

**Key words:** new energy consumption; generative adversarial network; reinforcement learning; intelligent operation mode generation

**基金项目:** 国家自然科学基金联合基金资助项目(U23B200694); 湖南省自然科学基金资助项目(2023JJ30024)。

**Foundation item:** Supported by the Joint Funds of the National Natural Science Foundation of China (U23B200694); the Natural Science Foundation of Hunan Province of China (2023JJ30024).

## 0 引言

在能源低碳转型的发展趋势下,我国正在大力开发以风力、光伏为代表的新型绿色能源,并且预计在 2030 年风光发电量占比达到 20 % 以上,2060 年达到 50 % 以上<sup>[1]</sup>。然而由于以风、光为主体的新能源在出力方面具有随机性和波动性<sup>[2-4]</sup>,致使部分网架结构相对薄弱地区的电网新能源消纳能力不足问题愈加凸显。随之出现的弃风、弃光等问题受到社会各界广泛关注<sup>[5-6]</sup>。如何有效地获取电力系统最优新能源消纳的运行方式样本成为实际电网调度运行决策提升新能源消纳能力的基础问题。

在实际电网调度决策系统中,一般是通过调度员运行经验从可行的历史运行方式集对当前电网运行方式进行调整,并分析是否满足电网稳定运行条件,离线做出决策<sup>[7-8]</sup>。但是该方法效率低、主观性强,难以满足电力系统不确定因素显著增加趋势下需要随机生成更多可行运行方式样本,并快速获得最优目标样本的要求。

为了获取更多运行样本,一般采用通过对不确定性因素随机抽取得到大量可分析样本的各类蒙特卡洛法<sup>[9-11]</sup>,以及通过确定的约束条件和目标求解得到样本的生产模拟法<sup>[12-13]</sup>等。但前者面临难以针对性地生成需求样本,且生成效率不够高的问题,后者具有计算量大、耗时长等缺陷,同时又考虑到电力系统与人类生产活动具有强关联的耦合性以及非常复杂的非线性特征,基于深度学习(deep learning, DL)的样本驱动研究方法成为了重要的研究方向之一。其中, Ian Goodfellow 在 2014 年提出的一种无监督学习模型—生成对抗网络在样本生成方面就具有极大潜力<sup>[14]</sup>。作为样本驱动的人工智能生成模型, GAN 能通过神经网络较好地学习样本数据分布规律,从而大量生成类似样本,并且在面对分布复杂且容量巨大的样本数据时依然能保证足够高的学习效率,且生成的样本质量要优于 VAE、Pixel CNN 等模型<sup>[15]</sup>。文献[16]提出的 GAN 模型能够生成全新、有效的样本来扩充原始数据集,而不仅仅只是对原始样本进行简单的拼接。文献[17]提出的基于 GAN 和模型迁移的电力系统典型运行方式样本生成方法根据少量运行样本和微调获得大量的典型运行方式样本有效解决了样本稀缺的问题。文献[18]将 GAN 模型和核密度估计(kernel density

estimation, KDE)相结合,生成具备标签的用户习惯与负荷模式数据样本。文献[19]提出了一种基于 GAN 的短期调度策略框架,实现了调度策略的快速生成。文献[20]提出了一种基于 GRA-r ACGAN 的数据增强方法,实现了海上风机异常工况数据的扩充。这些研究结果都证实 GAN 在电力系统样本生成方面具有可观的应用性。

尽管 GAN 在电力领域已经取得了一定成效,但若获取带有明显目标性的样本还需在 GAN 生成样本集后进行繁琐的筛选过程。具体到生成新能源消纳的电力系统运行方式样本过程中,为了获取最优消纳的运行方式样本,不仅需要 GAN 生成海量的运行方式样本集,还需对集中各个样本的新能源消纳特征值进行比对,这种方法计算过程冗杂,甚至极端情况下存在最优样本并未在生成样本集中的可能。针对以上的问题,本文在 GAN 的基础上引入强化学习方法<sup>[21]</sup>,以便在 GAN 生成运行方式样本的实时过程中根据不同的生成结果给予生成器相应的环境反馈,刺激其通过不断“试错”寻找到最优生成“动作”,即生成实时系统下的最优目标样本。

相比于传统的筛选法以及模型驱动等方法,强化学习作为一种无模型方法在计算和建模方面都表现出极大优势<sup>[22]</sup>,并且已经在电力系统运行与控制等多领域得到了广泛应用<sup>[23-25]</sup>。在寻找最优策略方面,文献[26]利用强化学习算法训练控制器,寻找到了最优动作策略,优化了风储合作决策。文献[27]利用强化学习方法找到了优化“智能能源枢纽”运行的近似最优决策。文献[28]利用基于策略梯度的 Actor-Critic 算法求解了微电网的最优调度策略。文献[29]利用深度确定性策略梯度(deep deterministic policy gradient, DDPG)算法实现了建筑集群内共享能源的最优 24 h 调度。这些研究成果都表明强化学习在决策寻优上效果突出。因此,本文在 GAN 的基础上引入强化学习来对生成器动作策略进行优化,以实现最优新能源消纳样本的生成。

综上,为解决能源结构多元的新型电力系统新能源消纳不足以及运行方式样本稀缺的问题,本文将生成对抗网络和强化学习相结合,提出了一种智能生成方法,可以根据现有的少量运行方式样本,通过 GAN 学习实际运行样本的分布特征,通过基

于策略梯度的强化学习方法实时优化生成器参数,不断调整生成策略,使生成模型往最优新能源消纳运行方式样本方向动作,生成当前电网架构下的目标样本。本文以提高风光消纳率为目标,以湖南省某地区电网为算例进行测试,算例结果表明该方法能够将GAN的生成能力和强化学习的寻优优势相结合,有效生成满足最优新能源消纳的运行方式样本,为解决弃风弃光问题提供理论依据与数据支撑。

## 1 基于GAN的电力系统运行方式样本生成

### 1.1 运行方式的特征提取

广义上,电力系统的运行方式指代电气设备运行的方法与形式;而在数学角度上,则可以被定义为各个节点的有功功率、无功功率、电压幅值和电压相角4个状态量的总集合。

本文借助以上4个状态量对运行方式进行特征提取。对 $N$ 个节点的电力系统将电网的某一完整运行方式具象为一个 $4 \times N$ 矩阵。而从潮流计算可知,电网节点被分为PQ节点、PV节点及平衡节点3类。对于包含这 $2N$ 个已知量的运行方式样本矩阵 $A$ ,如式(1),可以通过潮流方程(2)求出另外 $2N$ 个未知量,并根据安全运行约束式(3)~(4)去除不符合工况的极端样本,得到完整的 $4 \times N$ 运行方式矩阵。

$$A = \begin{bmatrix} (P_\alpha, Q_\alpha)^T \\ (P_\beta, V_\beta)^T \\ (V_\gamma, \theta_\gamma)^T \end{bmatrix} \quad (1)$$

式中: $\alpha$ 、 $\beta$ 、 $\gamma$ 分别为PQ节点、PV节点和平衡节点; $P_\alpha$ 、 $Q_\alpha$ 分别为PQ节点输入或输出的有功功率和无功功率; $P_\beta$ 、 $V_\beta$ 分别为PV节点输入或输出的有功功率和节点电压幅值; $V_\gamma$ 、 $\theta_\gamma$ 分别为平衡节点的电压幅值和相角。

$$\begin{cases} P_i = U_i \sum_{j=1}^N U_j (G_{ij} \cos \theta_{ij} + B_{ij} \sin \theta_{ij}), i = 1, 2, \dots, N \\ Q_i = U_i \sum_{j=1}^N U_j (G_{ij} \sin \theta_{ij} - B_{ij} \cos \theta_{ij}), i = 1, 2, \dots, N \end{cases} \quad (2)$$

式中: $P_i$ 和 $Q_i$ 分别为节点 $i$ 输入和输出的有功功率和无功功率; $U_i$ 和 $U_j$ 分别为节点 $i$ 和节点 $j$ 的电压幅值; $G_{ij}$ 和 $B_{ij}$ 分别为节点 $i$ 和节点 $j$ 之间的电导和电纳; $\theta_{ij}$ 为节点 $i$ 和节点 $j$ 之间的电压相位差。

$$\begin{cases} U_i^{\min} \leq U_i \leq U_i^{\max}, i = 1, 2, \dots, N \\ I_k \leq I_k^{\max}, k = 1, 2, \dots, b \end{cases} \quad (3)$$

式中: $I_k$ 为线路 $k$ 通过的电流; $I_k^{\max}$ 为线路 $k$ 所允许通过的最大电流; $b$ 为总线路数; $U_i^{\max}$ 和 $U_i^{\min}$ 分别为节点 $i$ 允许的最大和最小电压。

$$Q_{G, \min} \leq Q_G \leq Q_{G, \max} \quad (4)$$

式中: $Q_G$ 为发电机无功出力; $Q_{G, \max}$ 和 $Q_{G, \min}$ 分别为发电机无功出力的上下限。

满足上述各方程的样本定义为电力系统运行方式样本,本文将基于此类样本为解决电力系统目标样本生成问题。

### 1.2 基于GAN的运行方式样本生成

对于一实际电网,将各节点所包含的状态参数整合成矩阵 $A$ 后,本文利用GAN模型对 $A$ 中各参数的联合分布 $p_{\text{data}}$ 进行学习,进而生成类似 $A$ 分布规律的虚假运行方式样本,并将得到的各虚假样本代入潮流计算式,经由各类约束判别后得到完整的电网运行数据。区别于传统的“经验法”,GAN模型在生成效率以及生成样本质量上更具有保证,能够通过两个人工智能模型的对抗博弈,使输出样本达到理论上的最优值。

如图1所示,GAN模型的主体由生成器模型(generator, G)和对抗器模型(discriminator, D)两组深度神经网络组成。

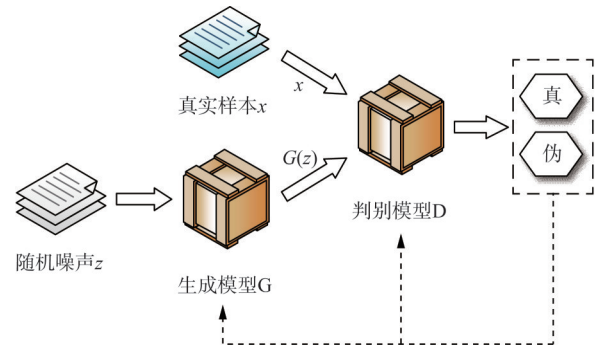


图1 GAN的基本结构

Fig. 1 Basic structure of GAN

其中,生成器模型通过一组随机噪声 $z$ 映射出对应的虚假样本 $G(z)$ ,判别器模型对生成器所生成的样本 $G(z)$ 和从属于真实样本集 $p_{\text{data}}$ 的真实样本 $x$ 进行真与假的判断。从两组模型的训练目标来看,判别器旨在对“真”与“伪”两组样本进行严格的分类,防止虚假样本被标记上真实样本的标签;生成器的优化目标在于生成能够使判别器给出“真”判断



的虚假样本,从而使判别器误判,难以区分生成样本的真假。生成模型与判别模型二者相互对抗,通过反复多次的博弈实现高效学习。

为保证生成器的学习效率以及生成样本的质量,同时也作为 GAN 的核心思想,生成器模型和判别器模型在不断的最大最小博弈中共同训练,直至达到纳什平衡,即虚假样本在分布规律上已几近真实样本,判别器对包括虚假样本在内的所有样本输出为完全随机的模糊判断。文献[14]将 GAN 的目标函数  $V(\cdot)$  定义为式(5)的形式。

$$\min_G \max_D V(D, G) = E_{x \sim p_{\text{data}}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log (1 - D(G(z)))] \quad (5)$$

式中:  $D$ 、 $G$  分别为判别器和生成器;  $x$  为从属于真实样本集  $p_{\text{data}}$  的真实样本; 随机噪声  $z$  从属于先验分布  $p_z$ , 一般选用标准正态分布;  $G(z)$  为生成器生成样本的输出结果;  $D(x)$ 、 $D(G(z))$  分别为判别器对真实样本和生成样本的输出;  $E_{x \sim p_{\text{data}}(x)}(\cdot)$  为真实数据被判别为真的概率的对数期望;  $E_{z \sim p_z(z)}(\cdot)$  为生成数据被判别为假的概率的对数期望。对于判别器网络  $D$ , 优化目标是希望它判别出真实数据的概率越大越好, 即对于服从  $p_{\text{data}}$  的样本  $x$ , 判别器的理想预测结果为  $D(x)=1$ , 即  $\log D(x)$  最大化。相反, 如果输入是生成数据  $D(G(z))$ , 判别器目标即为最小化  $D(G(z))$ , 希望它被判定为 0, 即希望  $\log(1-D(G(z)))$  越大越好。另一方面, 对于生成器网络  $G$ , 优化目标是生成样本能够被判别器识别为真, 即  $D(G(z))=1$ , 在式(4)中表现为希望  $\log(1-D(G(z)))$  越小越好, 此时目标函数达到最小值。

在电力系统运行方式样本生成过程中, 训练生成器和判别器一般采用梯度下降法。记生成器和判别器网络参数分别为  $\theta_g$  和  $\theta_d$ , 在训练时, 随机噪声  $z$  通过生成器  $G(z, \theta_g)$  实现  $p_z$  到虚假样本分布  $p_g$  的映射, 而判别器  $D(x, \theta_d)$  的输出结果表示输入的样本  $x$  从属于真实样本集  $p_{\text{data}}$  的概率。在对判别器参数进行反向传播更新时梯度方向需满足  $D(x, \theta_d)$  最大, 且  $D(G(z, \theta_g), \theta_d)$  最小, 以使判别器对真假样本之间的差异感知更加灵敏; 在对生成器参数更新时需向判别器输出  $D(G(z, \theta_g), \theta_d)$  最大的方向进行优化, 以使生成器和判别器在优化更新上满足最大最小的对抗, 从而实现对生成器的高效训练。

面向电力系统运行方式样本生成时, 为了保证生成样本满足电力系统运行约束, 采用安全运行约束及潮流约束的生成判别模型反复训练, GAN 的具体结构见图 2。

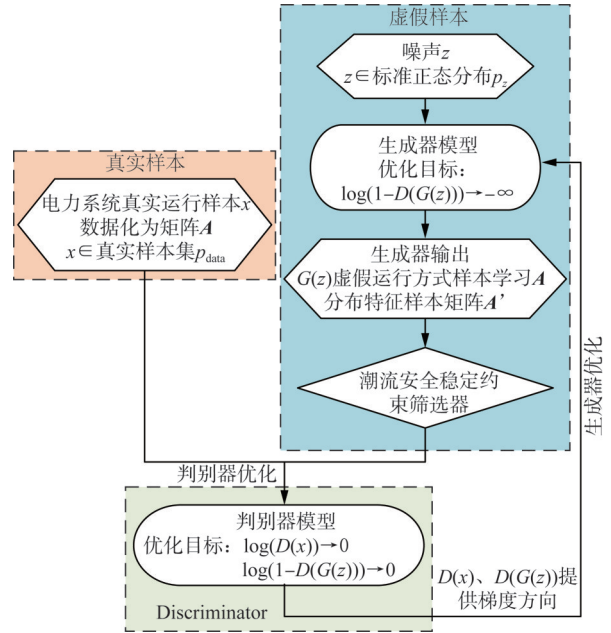


图2 生成电力系统运行方式样本的 GAN 架构

Fig. 2 GAN structure of generating power system operation mode samples

## 2 基于 GAN-RL 最优新能源消纳的运行方式样本生成

### 2.1 最优新能源消纳的目标函数

尽管 GAN 模型的样本生成能力已趋于成熟, 但只考虑模型本身时其生成样本仅满足真实样本的分布特征, 而并未定义“新能源消纳”标签, 因此生成的电力系统运行方式样本并不能直观地体现出新能源消纳能力的强弱。

为了量化电力系统对新能源的消纳能力, 本文考虑现阶段新能源发电的主要组成, 即风力和光伏发电, 并将目标函数选定为新能源消纳率  $\lambda$  最大, 即当前系统下的最优新能源消纳, 如式(6)所示。

$$\max \lambda = \frac{\sum_{s=1}^S P_{w,s} + \sum_{q=1}^Q P_{p,q}}{\sum_{s=1}^S P_{w,\max,s} + \sum_{q=1}^Q P_{p,\max,q}} \times 100\% \quad (6)$$

式中:  $s$  为风电场的序号;  $q$  为光伏电厂的序号;  $S$  为系统中风电场的数量 ( $1 \leq s \leq S$ );  $Q$  为光伏电站的

数量( $1 \leq q \leq Q$ );  $P_{w,s}$ 为系统对风电场 $s$ 消纳的实际风功率;  $P_{p,q}$ 为系统对光伏电场 $q$ 消纳的实际光电功率;  $P_{w, \max, s}$ 为风电场 $s$ 的最大可用发电量;  $P_{p, \max, q}$ 为光伏电场 $q$ 的最大可用发电量。

对于整个系统,除满足式(2)—(4)外,还需满足功率平衡约束,即:

$$\sum_{s=1}^S P_{w,s} + \sum_{q=1}^Q P_{p,q} + \sum_{m=1}^M P_{t,m} + \sum_{y=1}^Y P_{h,y} = P_L \quad (7)$$

式中:  $m$ 为火力发电厂的序号;  $y$ 为水电厂的序号;  $M$ 为系统中火力发电厂的数量( $1 \leq m \leq M$ );  $Y$ 为水电厂的数量( $1 \leq y \leq Y$ );  $P_{t,m}$ 为火电厂 $m$ 的实际输出功率;  $P_{h,y}$ 为水电厂 $y$ 的实际输出功率;  $P_L$ 为系统有功功率负荷。

对于风电场 $s$ 和光伏电场 $q$ ,其机组需满足运行约束,即:

$$\begin{cases} 0 \leq P_{w,s} \leq P_{w, \max, s} \\ 0 \leq P_{p,q} \leq P_{p, \max, q} \end{cases} \quad (8)$$

## 2.2 引入强化学习的GAN模型下最优新能源消纳运行方式样本生成

强化学习遵从累积“奖励”最大化原则,使agent即智能体在与环境的不断交互中根据不同的环境反馈而作出此条件下能获取最大收益的动作<sup>[19]</sup>。显而易见地,在强化学习训练过程中需要给予agent量化的反馈以定义当前决策的优劣程度。在本文中,由于新能源消纳被作为生成器样本生成的目标,因此式(6)定义的消纳率 $\lambda$ 需作为环境反馈的参数引入到强化学习中。

另一方面,如图3所示,生成器 $G$ 既是GAN的重要组成部分之一,同时又充当着强化学习中的agent,并且在每一次生成的电力系统运行方式样本中,消纳率 $\lambda$ 又不尽然相同,因此可将消纳率 $\lambda$ 作为桥梁,根据式(9)定义的奖惩函数来量化生成器每次生成样本时的环境反馈。

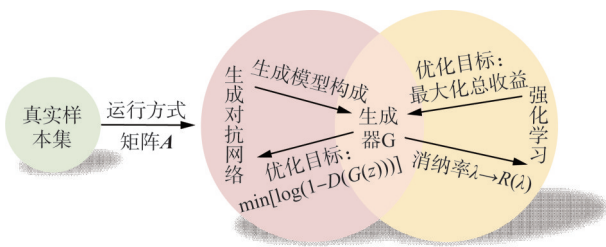


图3 GAN和强化学习中生成器职能

Fig.3 Functions of generator in GAN and RL

$$\begin{cases} R(\lambda_n) = \omega(\lambda_n - \lambda_{n-1}) + b, 2 \leq n \leq o \\ \lambda = \{\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_o\} \end{cases} \quad (9)$$

式中:奖惩反馈 $R(\cdot)$ 为消纳率集 $\lambda$ 的函数,其中 $\lambda_n$ 为生成器第 $n$ 次生成的运行方式样本新能源消纳率, $\lambda_{n-1}$ 为第 $n-1$ 次生成的样本新能源消纳率; $\omega(\cdot)$ 为奖惩函数的线性系数,反映生成器对奖惩反馈的灵敏程度( $\omega > 0$ );  $b$ 为偏置,由于策略寻优时只对比不同反馈间的相对值,因此偏置 $b$ 可任意选取,本文将其设置为0;  $o$ 为样本生成量。

$R(\lambda)$ 作为当前生成样本的消纳率与上次生成样本的消纳率之差的函数,它的意义在于衡量生成器当前生成动作与上次动作相比是否更优,即在消纳率上是否有正优化。

除环境反馈,即奖励(reward)外,强化学习的基本参数还包括状态(state)、动作(action)及策略(policy)。具体到本文面向电力系统运行样本生成的GAN模型中,各参数意义如表1所示。

表1 强化学习各参数意义

Tab.1 Significances of reinforcement learning parameters

| 强化学习参数     | 在面向电力系统的GAN模型中对应意义                                   |
|------------|------------------------------------------------------|
| 状态(state)  | 生成器输出 $S=G(z, \theta_g)$ ,即生成的各虚假电力系统运行方式样本 $s_n$    |
| 动作(action) | $\theta_g$ 基于目标函数式(12)的梯度调整,所有 $\theta_g$ 调整向量构成动作空间 |
| 奖励(reward) | 量化的奖惩反馈值 $r_n$ ,如式(11)                               |
| 策略(policy) | 生成器根据状态选择动作的方式,初始策略可定义为单独的GAN生成器策略                   |

而在训练方法上,强化学习又可分为基于值的强化学习<sup>[30]</sup>和基于策略的强化学习<sup>[31-32]</sup>。基于策略的强化学习以最大化累积奖励为训练目标,找到累积奖励上升方向,即其随策略参数变化的梯度,从而实现对策略参数的优化。

本文采用基于策略梯度优化更新算法的基本思想在于生成器进行对抗训练学习真实样本分布规律的实时过程中,将在各状态下生成器选择不同动作得到的奖励式(9)作为环境反馈信号,通过增加生成网络参数获取高额奖赏动作,即生成更高新能源消纳率运行方式样本出现的概率,减少得到惩罚动作,即生成更低新能源消纳率运行方式样本出现的概率。

记agent,即生成器生成 $i$ 个样本的参数调整轨

迹  $\tau$  为:

$$\tau = [(s_1, a_1, s_2, r_1), (s_2, a_2, s_3, r_2), \dots, (s_{i-1}, a_{i-1}, s_i, r_{i-1})] \quad (10)$$

式中:  $a_i$  为第  $i$  次训练时选择的动作;  $r_i$  为第  $i$  次动作获得的奖励或惩罚;  $s_i$  为经过  $i-1$  次训练后的状态。

对于  $r_n (n=1, 2, \dots, i-1)$  有:

$$r_n = R(\lambda_{n+1}) = \omega(\lambda_{n+1} - \lambda_n) + b \quad (11)$$

基于策略梯度方法的在线策略优化, 其目标函数  $J(\theta_g)$  可以被定义为:

$$\begin{aligned} J(\theta_g) &= \sum_{t=0}^T \gamma^t R_t = \sum_{t=0}^T \gamma^t R_{\tau_t} \\ &= R_{\tau_0} + \gamma R_{\tau_1} + \gamma^2 R_{\tau_2} + \dots + \gamma^T R_{\tau_T}, \gamma \in [0, 1] \end{aligned} \quad (12)$$

式中:  $\gamma$  为折扣因子;  $t$  为时间步骤;  $T$  为总训练时间;  $R_t$  为时间步骤  $t \rightarrow t+1$  内获得的总奖励, 倘若将每个时间步骤当作一个单独的样本轨迹  $\tau_t$ , 则每个时间步骤获得的奖励  $R_t$  即为每条样本轨迹获得的奖励  $R_{\tau_t}$ 。目标函数可被定义为多条样本轨迹的实际总奖励, 在系统和连续地利用多条轨迹, 即多次迭代后, 可以逐渐逼近真实的期望值。

定义了具体的奖惩机制和目标函数, 则可以通过梯度训练来更新参数。为了获得最大总奖励的策略, 通常采用梯度上升的训练方法, 在对生成器进行更新时梯度方向需满足  $J(\theta_g)$  最大, 以使获得高额奖励行为被选中的概率增大。通过不断地迭代训练, 生成器策略从初始的 GAN 单独生成策略  $\pi_0$  转化为逐渐专注生成更优消纳率运行样本的策略。

另外, 本文中 GAN 模型和强化学习相互交叉, 当生成器  $G$  每生成一个新样本, 转移到另一状态时, 即评估当前生成样本的优劣来定义此时生成器动作的好坏, 而非等 GAN 完全训练结束后再进行评估, 有效提高了计算效率。本文在传统 GAN 模型上引入基于策略梯度的强化学习进行目标样本生成, 其基本架构如图 4 所示。

在模型训练过程中, 生成器输出的运行方式样本经过系统安全约束筛选后不仅输入到数据集, 最终经由判别器进行二分类判断, 同时又作为强化学习的状态输入。在基于策略梯度的优化更新中, 奖惩模块基于生成器参数更新后下一步动作所产生的样本, 与上一次动作所产生的样本进行比较, 给予生成器一个往奖励梯度更新的趋向, 推动生成网络

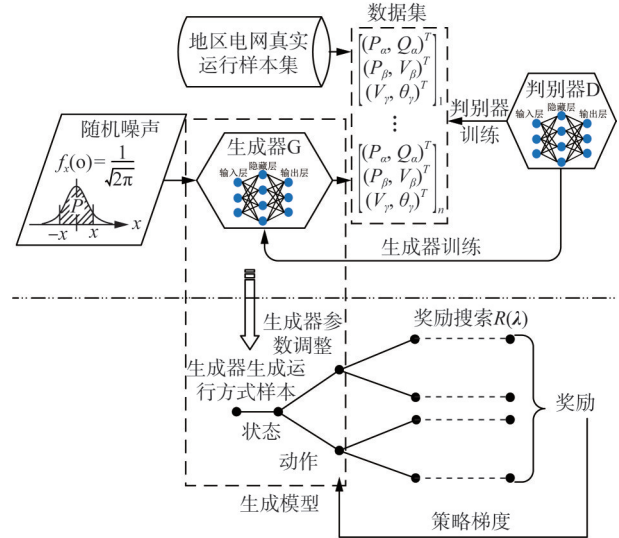


图 4 引入策略梯度的改进式 GAN 结构图

Fig. 4 Structure graph of the improved GAN introducing strategy gradient

往既定目标方向进行动作。

为了实现生成器在拟合真实运行方式样本分布规律的过程中, 往奖励梯度方向更新, 需要对 GAN 和强化学习两部分的目标函数进行耦合优化, 即:

$$\begin{cases} \min V(D, G_{\theta_g}) = E_{z \sim p_z(z)} [\log(1 - D(G(z)))] \\ \max J(\theta_g) = \sum_{t=0}^T \gamma^t R_t \end{cases} \quad (13)$$

在 GAN-RL 达到一定的训练次数时, 生成器参数在两种优化方法的先后更新下取得一个平衡点, 即在一定程度上既不会偏离真实分布规律又保证了真实工况下的最优消纳特征。

### 3 整体流程

本文提出的基于 GAN-RL 最优新能源消纳的运行方式样本智能生成具体流程如图 5 所示。

首先, 在信息交互层通过对电网实际运行方式进行特征提取和预处理, 构建真实运行方式样本集。在生成对抗网络层判别器以真实运行方式样本集和生成器在潮流安全稳定约束下输出的虚假运行方式样本为输入, 在最大最小博弈中通过目标函数式(5)驱动, 与生成器不断进行对抗性训练, 并反之为生成器提供反向传播参数, 促使生成器在学习真实样本分布规律和数据特征上趋向成熟。同时, 强化学习层以当前生成器输出  $G(z)$  为输入, 以每



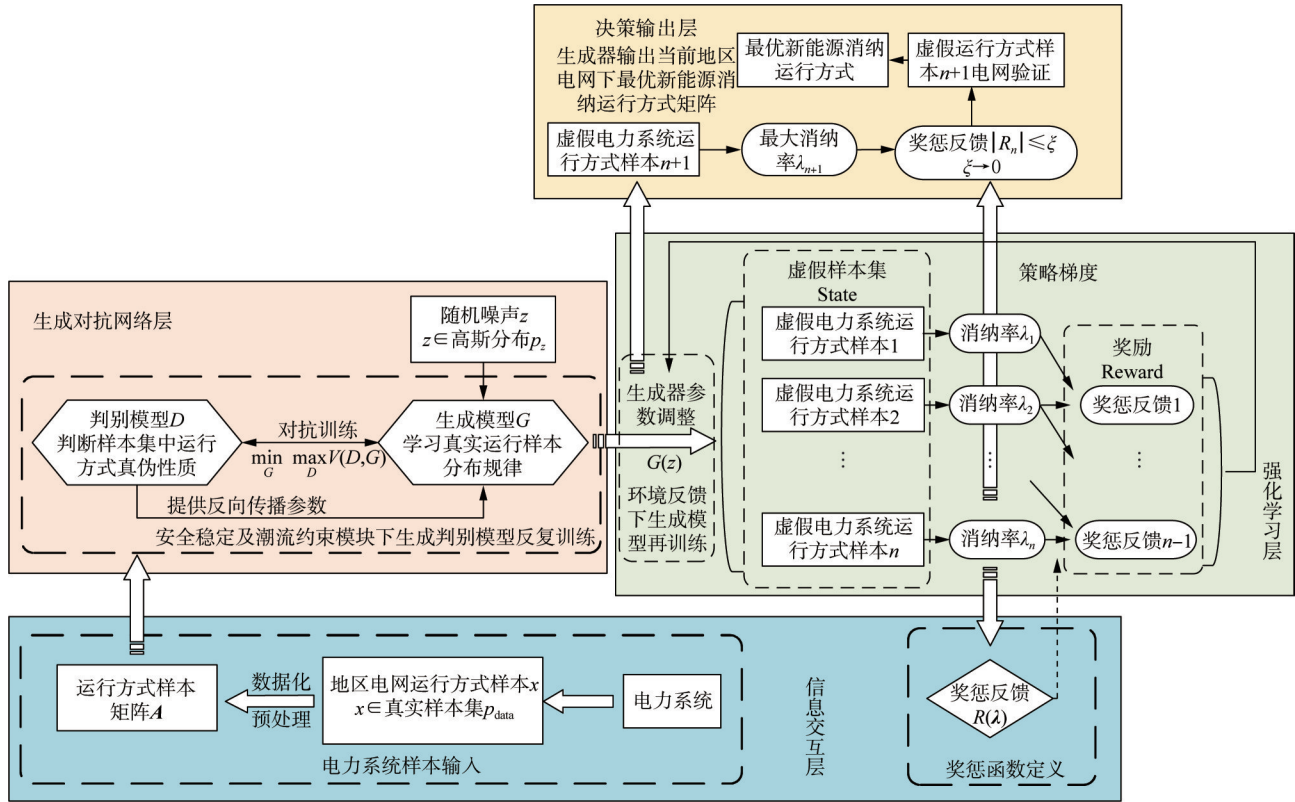


图5 基于GAN-RL的电力系统最优新能源消纳运行方式样本生成流程

Fig. 5 Sample generation process of optimal new energy consumption operation mode in power systems based on GAN-RL

一虚假运行方式样本下的新能源消纳率 $\lambda$ 为媒介,通过信息交互层定义的奖惩函数映射出当前生成器动作下的环境反馈,通过策略梯度优化增加生成器获取更高反馈奖励动作出现的概率,与生成对抗网络层实现对生成器网络参数的交替优化。当生成的运行方式样本达到系统条件下的最大消纳率,即反馈函数的范数趋近于一个极小值时,决策输出层即输出该最优新能源消纳的运行方式样本。

## 4 算例分析

### 4.1 算例设置

图6是本文所采用的湖南省某地区50节点电网算例。

本文以该地区电网调度人员所制定的实际运行方式为真实样本集。以基于GAN方法生成的运行方式样本特征值分布情况,与GAN-RL方法进行对比,验证后者在目标样本生成方面表现更优。以多次训练下GAN-RL方法生成的运行方式样本新能源消纳率和奖惩反馈值曲线的变化为指标,验证该方法在生成最优消纳样本上的可行性,并将最终生成

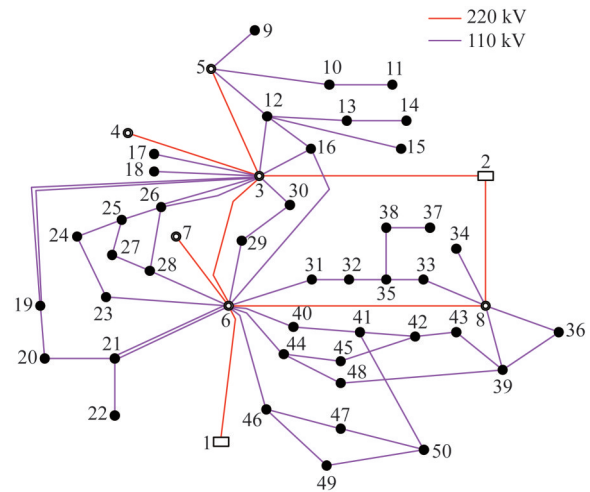


图6 湖南省某地区50节点电网地理接线图

Fig. 6 Geographical wiring diagram of 50-node power grid in a certain region of Hunan Province

的最优新能源消纳运行方式样本在GAN方法所生成的大量运行方式样本中进行最优性验证。

表2是该地区电网算例中各节点意义。其中,节点1和节点2分别为总装机容量为200 MW的风力发电站和总装机容量为150 MW的光伏发电站,

表2 算例地区电网中各节点意义

Tab. 2 Meanings of various nodes in the regional power grid of the example

| 节点          | 意义                |
|-------------|-------------------|
| 1,2         | 新能源电站             |
| 3,4,5,6,7,8 | 涵盖火力、水力发电厂的电力系统站点 |
| 其他          | 负荷                |

节点3、4、5、6、7、8均为涵盖火力和水力发电厂的电力系统站点。通过对调方给出的实际运行方式进行特征提取，得到30条有效的实际运行方式样本，分别记为 $A_1$ 、 $A_2$ 、 $\dots$ 、 $A_{30}$ ，记录该30条实际运行方式样本的消纳情况如图7所示。

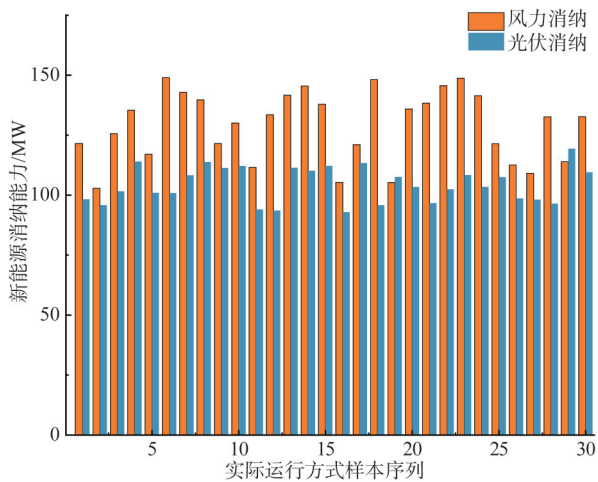


图7 各实际运行方式样本下新能源消纳能力

Fig. 7 New energy consumption capacities under various real operating mode samples

由图7可知，在当前已有的多种实际运行方式下新能源消纳能力还有较大的提升空间，风光资源都还未被有效利用。基于这种情况，一方面，需要考虑用高效率方法生成更多可行的运行方式样本，以补充现有的运行方式样本集；另一方面，除考虑补充样本集外还需要考虑通过寻优方法找到最优的生成策略以获取最优新能源消纳样本。

#### 4.2 基于GAN的运行方式样本生成分析

在未考虑通过强化学习寻找最优生成策略的前提下，针对上一节所获取的实际运行方式样本，以表2中各负荷节点的有功功率和无功功率和发电机节点的有功功率输出和机端电压为桥梁，将实际运行方式样本集数据化为 $\{A_i\}$ 。为保证后续生成器和判别器能够更好地训练，根据各状态参数最大最小值进行归一化更新运行方式样本矩阵 $A$ ，构建真实

样本集 $p_{data}$ 。

在包含50节点的算例中，对于任意一运行方式，可用100维的列向量表述。在基于11th Gen Intel(R) Core(TM) i5-11260H @ 2.60 GHz 2.61 GHz 和 NVIDIA GeForce RTX 3050 Laptop GPU 的配置以及Pytorch框架下，构建生成对抗网络架构。定义一个100维高斯分布，并通过随机采样得到生成器模型的输入，生成器网络包含3个线性层和激活函数ReLU，将100维的输入噪声向量映射为100维生成的虚假数据。判别器将100维向量作为输入，包含3个线性层和激活函数ReLU，并通过最后一层的Sigmoid函数对输入数据进行分类，输出一个介于0到1之间的概率值，表示输入数据为真实数据的概率。为避免出现过拟合问题，在各层间采用 $p=0.3$ 的Dropout层以提高模型的泛化能力。通过式(2)~(4)建立潮流安全稳定约束筛选器，在判别器判定前对生成器输出的批次向量进行过滤以增强样本质量。通过式(5)的损失函数，采用Adam优化器对生成器和判别器的参数进行优化。在对网络进行5000次训练后，通过生成器生成1000条虚假数据，还原为运行方式样本矩阵后，从中随机抽取5条运行方式样本，另外从实际运行方式样本集中筛选2条具有最高新能源消纳特征的真实样本，记为实际样本1和实际样本2，绘制运行方式状态参数特征值分布图8。

从图8可以看出，GAN在进行5000次训练后，已经能够较好地学习真实运行方式样本的分布特

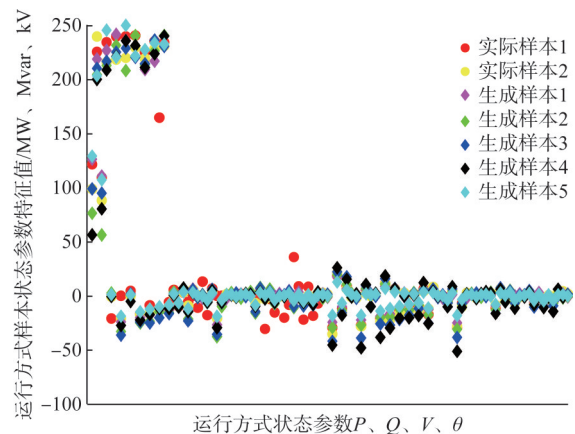


图8 基于GAN方法生成的虚假样本与实际样本特征值分布图

Fig. 8 Distribution map of fake and real sample eigenvalues generated based on GAN



征,从而生成符合当前系统工况的虚假样本,但这也是其局限性的体现。为了保证生成样本的“类真实性”,在不加目标的前提下生成模型很难打破最大程度拟合真实样本分布的规则,这也就使得当前生成的样本在消纳率上并未有明显提升。换言之,GAN在最优新能源消纳样本的生成上难以发挥出可观作用,因此下文结合了强化学习优化生成器模型,使其对既定目标进行进一步生成。

#### 4.3 基于GAN-RL的最优新能源消纳运行方式样本生成分析

在第2节中,为了解决GAN难以生成目标样本的问题,首先赋予了GAN模型中生成网络一个目标函数式(6),并通过含该目标 $\lambda$ 的反馈函数式(9)将量化的样本优劣情况反映给生成动作结束后已完成GAN参数更新的生成器,使其在基于策略梯度的优化算法下二次更新网络参数,以在下一步动作中完成既定目标的更新。在本算例中,考虑到新能源消纳率之差为百分比,将奖励信号 $R(\lambda)$ 中的参数 $\omega$ 设置为100,生成模型学习率设置为0.01,折扣因子 $\gamma$ 设置为0.99,通过最大化目标函数 $J(\theta_g)$ ,并将目标函数中每一个时间步骤定义为生成器执行一次动作,即生成一个虚假运行方式样本后观察到环境给予反馈的过程,最小化系统约束下的目标样本生成损失 $|R(\lambda)|$ ,以环境反馈为奖励信号反向传播更新生成器。

多目标耦合训练下生成器模型基于GAN的约束排除极端样本的生成动作,同步通过策略梯度优化算法二次更新生成网络参数,在模型优化过程中,逐步生成满足更优条件的运行方式样本。

图9是将4.2节中同样的实际运行方式样本作为真实样本集,将上述奖惩函数作为反馈信号,在GAN-RL训练过程中对目标函数式(13)进行归一化处理后的同步监测曲线。

由图9可以发现,在经过5000次训练后达到一个较好的平衡点。此时,目标函数1趋近于一个极小值,而目标函数2趋近于一个极大值。

同时,记录5000次训练过程中生成样本的新能源消纳率及奖惩反馈大小变化曲线如图10所示,并分别将100次、500次、2000次和5000次训练后的 $|R(\lambda)|$ 值记录于表3。

从图10和表3结果可以看出,当训练次数不断增加时,生成模型在样本生成方向上逐渐明确,特

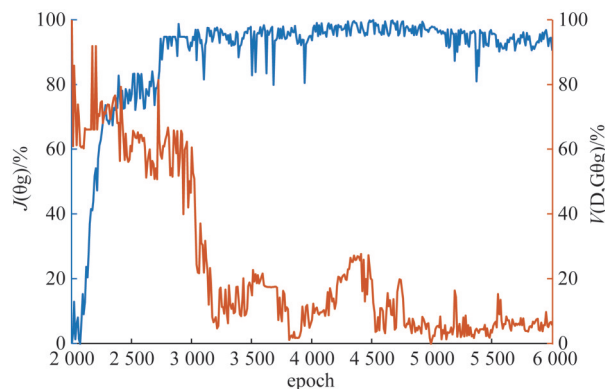


图9 GAN-RL训练过程中目标函数(13)监测曲线

Fig. 9 Monitoring curves of objective function (13) during GAN-RL training process

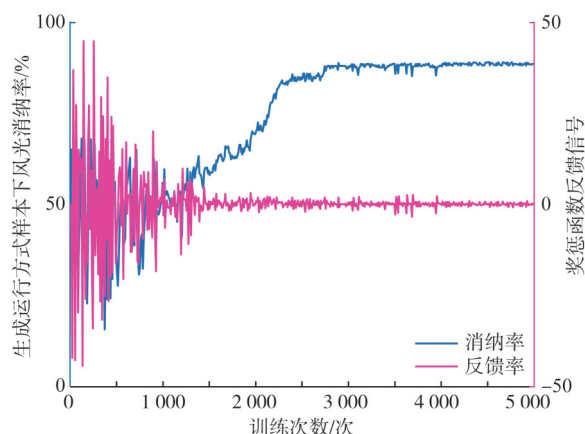


图10 训练过程中样本的消纳率与对应反馈大小变化曲线

Fig. 10 Consumption rates of samples and the corresponding feedback size change curves during the training process

表3 固定次数训练后 $|R(\lambda)|$ 值与累积收益

Tab. 3 The  $|R(\lambda)|$ -values and cumulative benefits after a fixed number of trainings

| 训练次数 | $ R(\lambda) $ | $\sum R(\lambda)$ |
|------|----------------|-------------------|
| 100  | 10.382         | -20.219           |
| 500  | 7.514          | -33.350           |
| 2000 | 1.152          | 4.025             |
| 5000 | 0.098          | 23.389            |

别经过5000次训练后, $|R(\lambda)|$ 已经在0附近波动,且风光消纳率在此时达到了一个近乎稳定的峰值。但当训练次数较少时,考虑GAN-RL生成尚不成熟,往往由于GAN模型训练次数较少并未有效拟合真实样本分布规律,或寻优机制并不明确而导致逆目标生成,即生成比前次样本消纳率更低的运行方式。

为了验证GAN-RL方法不仅具备GAN高效学

习真实样本特征的优势, 同时在目标样本生成上的表现较之传统 GAN 方法更优。图 11 将 GAN-RL 生成模型训练 5 000 次后所生成的运行方式样本与上一节中 GAN 模型训练 5 000 次后生成样本 4 的状态参数特征值分布进行了比较。

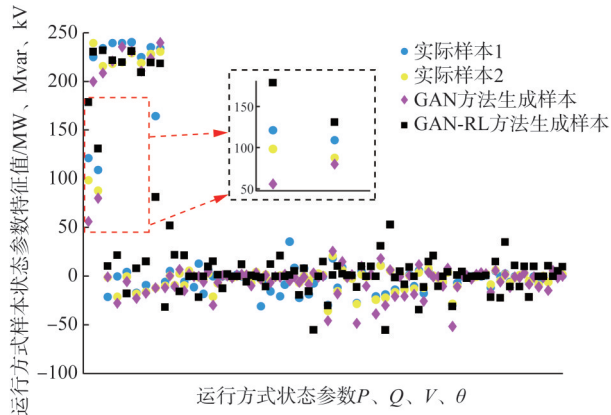


图 11 GAN-RL 生成模型训练 5 000 次后生成样本的特征值分布图

Fig. 11 Distribution map of sample eigenvalues generated by the GAN-RL generative model after 5 000 trainings

从图 11 看出, GAN-RL 方法生成的样本依然能够较好地拟合实际运行工况, 并且在新能源消纳特征值上较之 GAN 方法生成样本和实际运行方式样本都有所提高。换言之, GAN-RL 生成模型在得到充足的训练后能够较好改善新能源消纳不足问题。

对于图 11 中 4 条电力系统运行方式, 记录其新能源消纳率于表 4。

表 4 GAN 方法与 GAN-RL 方法生成样本的新能源消纳率

Tab. 4 New energy consumption rates of samples generated by GAN method and GAN-RL method

| 样本序列          | 风力消纳率/% | 光伏消纳率/% |
|---------------|---------|---------|
| 实际样本 1        | 60.727  | 72.923  |
| 实际样本 2        | 49.358  | 58.744  |
| GAN 方法生成样本    | 28.165  | 53.542  |
| GAN-RL 方法生成样本 | 89.480  | 87.439  |

以表 4 结果中运行方式样本的消纳率为指标分析, 可以明显看出, GAN-RL 方法生成的运行方式极大程度地提高了风光消纳水平。特别在风力方面较之消纳水平较高的实际运行方式 1 也提高了接近 1.5 倍的风力消纳; 而在光伏消纳上比较其他 3 条运行方式都有着显著的提升。

另外, 为了验证 GAN-RL 生成模型所生成的样

本质量优于 GAN 生成模型, 统计 4.2 节 GAN 生成模型训练 5 000 次后生成的 1 000 条样本, 以及 GAN-RL 生成模型 1 000 次、1 000~1 500 次、1 500~2 000 次以及 2 000~5 000 次训练内生成样本的消纳率, 通过对比各消纳率区间生成样本数在总样本数中的占比比较二者在目标样本上的生成效率, 如图 12 所示。

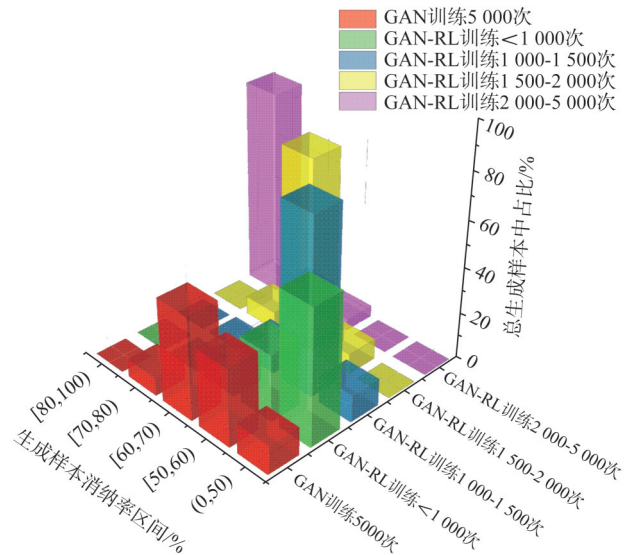


图 12 GAN 生成样本与 GAN-RL 生成样本消纳率对比

Fig. 12 Comparison of new energy consumption rates of samples generated by between GAN and GAN-RL

对图 12 分析可知, 传统 GAN 模型相比 GAN-RL 模型, 在生成高消纳率样本上难有显著成效, 尽管在少部分样本中确实提高了消纳, 但由于学习模型的严格约束, 其提升空间并不可观。而 GAN-RL 生成模型, 以 2 000 次训练为分水岭, 在训练次数较少时, 由于生成机制尚未成型, 在目标生成上极不稳定; 而当训练次数足够多时生成器最优动作策略逐渐成熟, 目标生成愈发稳定, 已经能够保持足够高的生成效率。

最后, 为了验证 GAN-RL 生成模型在经过 5 000 次训练后所生成样本的最优性, 将该运行方式样本置于 GAN 生成模型经过 5 000 次训练后所生成的 1 000 条虚假样本中进行比对, 证明了其在新能源消纳率上确实满足当前样本集下的最优。

比较传统的经验法, 基于深度神经网络的 GAN 能够生成更优、更泛的电力系统运行方式样本; 与 GAN 样本生成方法相比, 本文提出的 GAN-RL 智能生成方法能够在 GAN 的基础上, 基于量化

的反馈使生成模型产生更优样本的趋向,在多次训练后生成当前目标下的最优样本,为电力系统运行调度决策提供一定的数据支撑。

## 5 结论

本文所提出的基于GAN-RL智能生成方法能够在最优新能源消纳运行方式样本的生成上得到较好应用。具体来说,先通过给定的实际运行方式构建真实样本数据集,然后利用GAN中判别模型对生成模型进行对抗性训练以学习真实运行方式样本的分布特征。同时定义一个环境的奖惩反馈函数对生成器每一次动作所产生结果进行量化的“优劣”判定,以衡量该次动作所生成运行方式样本的新能源消纳能力提高程度,使得生成模型在学习真实运行方式样本的基础上实时根据不同的反馈进行参数调整,以在一次次策略的调整中找到获取最大“奖励”,即生成更优新能源消纳样本的动作策略,最终使生成方向与目标方向一致化。通过算例中地区电网的验证结果可知,GAN-RL生成方法在给定的奖励策略下能够有效、高质地生成当前电力系统下的最优新能源消纳运行方式样本,为电网调度人员提供数据参考。

另外,由于本文所提出的智能生成方法目前仅在小样本的地区电网中进行验证,在其面对节点更多、拓扑更复杂的多区域互联大电网时的生成质量与样本最优性还尚未检验,但由于人工智能的泛应用性以及高适用性,基于机器学习的数据驱动方法将是未来大电网的研究重点之一。

## 参考文献

- [1] 辛保安,单葆国,李琼慧,等. “双碳”目标下“能源三要素”再思考[J]. 中国电机工程学报, 2022, 42(9): 3117–3126.  
XIN Bao'an, SHAN Baoguo, LI Qionghui, et al. Rethinking on the "three elements of energy" under the "dual carbon" goal[J]. Proceedings of the CSEE, 2022, 42(9): 3117–3126.
- [2] 刘俊磊,刘新苗,姜源媛,等. 考虑源荷储不确定性的新型电力系统随机生产模拟方法[J]. 南方电网技术, 2024, 18(6): 36–42.  
LIU Junlei, LIU Xinmiao, LOU Yuanyuan, et al. A stochastic production simulation method for a new power system considering source load storage uncertainty[J]. Southern Power System Technology, 2024, 18(6): 36–42.
- [3] 何耿生,陈晖,张世旭,等. 储能提升高比例新能源基地低碳-经济-可靠性能的效益分析[J]. 南方电网技术, 2024, 18(4): 120–129.  
HE Gengsheng, CHEN Hui, ZHANG Shixu, et al. Benefit analysis of low carbon, economical, and reliable performance of high proportion new energy base improved by energy storage[J]. Southern Power System Technology, 2024, 18(4): 120–129.
- [4] 李建林,丁子洋,游洪灏,等. 构网型储能支撑新型电力系统稳定运行研究[J]. 高压电器, 2023, 59(7): 1–11.  
LI Jianlin, DING Ziyang, YOU Honghao, et al. Research on grid-forming energy storage supporting the stable operation of new power systems[J]. High Voltage Apparatus, 2023, 59(7): 1–11.
- [5] 舒印彪,张智刚,郭剑波,等. 新能源消纳关键因素分析及解决措施研究[J]. 中国电机工程学报, 2017, 37(1): 1–9.  
SHU Yinbiao, ZHANG Zhigang, GUO Jianbo, et al. Analysis of key factors in the consumption of new energy and research on solutions[J]. Proceedings of the CSEE, 2017, 37(1): 1–9.
- [6] 刘显苗,王皓怀,邓韦斯,等. 考虑不同时间尺度下的新能源跨区域消纳策略[J]. 南方电网技术, 2023, 17(2): 37–46, 56.  
LIU Xianzhuo, WANG Haohuai, DENG Weisi, et al. Consider cross regional consumption strategies for new energy at different time scales[J]. Southern Power System Technology, 2023, 17(2): 37–46, 56.
- [7] 孙宏斌,黄天恩,郭庆来,等. 面向调度决策的智能机器调度员研制与应用[J]. 电网技术, 2020, 44(1): 1–8.  
SUN Hongbin, HUANG Tian'en, GUO Qinglai, et al. Automatic operator for decision-making in dispatch: research and applications[J]. Power System Technology, 2020, 44(1): 1–8.
- [8] 朱思婷,管霖,徐尧燚,等. 电力调度智能操作的知识模型和实现技术[J]. 南方电网技术, 2022, 16(12): 98–108, 133.  
ZHU Siting, GUAN Lin, XU Yaoyi, et al. Knowledge model and implementation technology for intelligent operation of power dispatch[J]. Southern Power System Technology, 2022, 16(12): 98–108, 133.
- [9] 马光,张伊宁,陈哲,等. 含大规模风电的交直流混联系统风险评估方法[J]. 电网技术, 2019, 43(9): 3241–3252.  
MA Guang, ZHANG Yining, CHEN Zhe, et al. Risk assessment method for AC/DC hybrid systems containing large-scale wind power[J]. Power System Technology, 2019, 43(9): 3241–3252.
- [10] 王强强,姚良忠,徐箭,等. 基于切片采样-马尔科夫链蒙特卡罗模拟的高比例新能源电力系统等效惯量概率评估[J]. 电网技术, 2024, 48(1): 140–152.  
WANG Qiangqiang, YAO Liangzhong, XU Jian, et al. Probability evaluation of equivalent inertia in high proportion new energy power systems based on slice sampling Markov chain Monte Carlo model[J]. Power System Technology, 2024, 48(1): 140–152.
- [11] 马燕峰,骆泽榕,赵书强,等. 基于改进蒙特卡罗混合抽样的含风光电力系统风险评估[J]. 电力系统保护与控制, 2022, 50(9): 75–83.  
MA Yanfeng, LUO Zerong, ZHAO Shuqiang, et al. Risk assessment of wind and solar power systems based on improved Monte Carlo mixed sampling[J]. Power System Protection and Control, 2022, 50(9): 75–83.
- [12] 朱俊澎,施凯杰,李强,等. 考虑输电潮流约束的时序生产模拟及新能源消纳能力评估[J]. 电网技术, 2022, 46(5): 1947–1955.



- ZHU Junpeng, SHI Kaijie, LI Qiang, et al. Time series production simulation and new energy consumption capacity evaluation considering power flow constraint in transmission network [J]. Power System Technology, 2022, 46(5): 1947 – 1955.
- [13] 赵书强, 索瑞, 许朝阳, 等. 考虑断面约束的多能源电力系统时序性生产模拟[J]. 电力自动化设备, 2021, 41(7): 1 – 6.
- ZHAO Shuqiang, SUO Qian, XU Chaoyang, et al. Sequential production simulation of multi-energy power system considering section constraints [J]. Electric Power Automation Equipment, 2021, 41(7): 1 – 6.
- [14] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets [C]//Advances In Neural Information Processing Systems 27, December 8 – 13, 2014, Montreal, Canada. Cambridge: NIPS, 2014: 2672 – 2680.
- [15] 林懿伦, 戴星原, 李力, 等. 人工智能研究的新前线: 生成式对抗网络[J]. 自动化学报, 2018, 44(5): 775 – 792.
- LIN Yilun, DAI Xingyuan, LI Li, et al. A new frontier in AI research: generative adversarial networks [J]. Acta Automatica Sinica, 2018, 44(5): 775 – 792.
- [16] 杨懿男, 齐林海, 王红, 等. 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.
- YANG Yin'an, QI Linhai, WANG Hong, et al. Research on small sample data generation technology based on generative adversarial network [J]. Electric Power Construction, 2019, 40(5): 71 – 77.
- [17] 兰健, 郭庆来, 周艳真, 等. 基于生成对抗网络和模型迁移的电力系统典型运行方式样本生成[J]. 中国电机工程学报, 2022, 42(8): 2889 – 2900.
- LAN Jian, GUO Qinglai, ZHOU Yanzhen, et al. Sample generation of typical operation mode of power system based on generative adversarial network and model migration [J]. Proceedings of the CSEE, 2022, 42(8): 2889 – 2900.
- [18] EL K S, SRIKANTHA P. A data-driven approach for generating synthetic load patterns and usage habits [J]. IEEE Transactions on Smart Grid, 2020, 11(6): 4984 – 4995.
- [19] GUAN Jinyu, TANG Hao, WANG Jiye, et al. A GAN-based fully model-free learning method for short-term scheduling of large power system [J]. IEEE Transactions on Power Systems, 2022, 37(4): 2655 – 2665.
- [20] 魏书荣, 殷世杰, 闫梦飞, 等. 基于改进生成对抗网络的海上风电机组故障数据增强及诊断[J]. 电力系统保护与控制, 2025, 53(1): 114 – 124.
- WEI Shurong, YIN Shijie, YAN Mengfei, et al. Enhancement and diagnosis of offshore wind turbine fault data based on improved generative adversarial networks [J]. Power System Protection and Control, 2025, 53(1): 114 – 124.
- [21] SUTTON R S, BARTO A G. Reinforcement learning: an introduction [M]. Cambridge, Mass.: MIT press, 2018.
- [22] XU Bowen, GUO Fanghong, ZHANG Wenan, et al. E<sup>2</sup> DNet: an ensembling deep neural network for solving nonconvex economic dispatch in smart grid [J]. IEEE Transactions on Industrial Informatics, 2022, 18(5): 3066 – 3076.
- [23] 彭刘阳, 孙元章, 徐箭, 等. 基于浓度强化学习的自适应不确定性经济调度[J]. 电力系统自动化, 2020, 44(9): 33 – 42.
- PENG Liuyang, SUN Yuanzhang, XU Jian, et al. Self-adaptive uncertainty economic dispatch based on deep reinforcement learning [J]. Automation of Electric Power Systems, 2020, 44(9): 33 – 42.
- [24] 曾仪, 周毅, 陆继翔, 等. 基于多智能体安全深度强化学习的电压控制[J]. 中国电力, 2025, 58(2): 111 – 117.
- ZENG Yi, ZHOU Yi, LU Jixiang, et al. Voltage control based on multi-agent secure deep reinforcement learning [J]. Electric Power, 2025, 58(2): 111 – 117.
- [25] LI Yanying, XI Lei, GUO Yiguo, et al. Automatic generation control based on the weighted double Q-delayed update learning algorithm [J]. Proceedings of the CSEE, 2022, 42(15): 5459 – 5470.
- [26] 刘国静, 韩学山, 王尚, 等. 基于强化学习方法的风储合作决策[J]. 电网技术, 2016, 40(9): 2729 – 2736.
- LIU Guojing, HAN Xueshan, WANG Shang, et al. Collaborative decision-making of wind storage based on reinforcement learning methods [J]. Power System Technology, 2016, 40(9): 2729 – 2736.
- [27] RAYATI M, SHEIKHI A, RANJBAR A M. Applying reinforcement learning method to optimize an energy hub operation in the smart grid [C]//2015 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference, February 18 – 20, 2015, Washington, DC, USA. New York: IEEE, 2015: 1 – 5.
- [28] WEI Yifei, ZHANG Zhiqiang, YU F. Richard, et al. Power allocation in HetNets with hybrid energy supply using Actor-Critic reinforcement learning [C]//GLOBECOM 2017 – 2017 IEEE Global Communications Conference, December 4 – 8, 2017, Singapore, Singapore. New York: IEEE, 2018: 1 – 5.
- [29] ODKOR P, LEWIS K. Control of shared energy storage assets within building clusters using reinforcement learning [C]//International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, August 26 – 29, 2018, Quebec, Canada. New York: ASME, 2018: 86 – 94.
- [30] 孟安波, 许炫淙, 陈嘉铭, 等. 基于强化学习和组合式深度学习模型的超短期光伏功率预测[J]. 电网技术, 2021, 45(12): 4721 – 4728.
- MENG Anbo, XU Xuancong, CHEN Jiaming, et al. Ultra short term PV power prediction based on reinforcement learning and combined deep learning model [J]. Power System Technology, 2021, 45(12): 4721 – 4728.
- [31] LILLICRAP T P, HUNT J J, PRITZEL A, et al. Continuous control with deep reinforcement learning [EB/OL]. (2017 – 07 – 05) [2023 – 12 – 24]. <https://doi.org/10.48550/arXiv.1509.02971>.
- [32] HEES N, TB D, SRIRAM S, et al. Emergence of locomotion behaviours in rich environments [EB/OL]. (2017 – 07 – 10) [2023 – 12 – 24]. <https://doi.org/10.48550/arXiv.1707.02286>.

收稿日期: 2024-01-16; 网络首发日期: 2024-06-24

作者简介:

李帅虎(1981), 男, 教授, 博士生导师, 博士, 研究方向为电力系统电压稳定分析与控制、智能电网的自愈调控方法以及大规模储能建模与优化控制方法, [lishuaihu2010@126.com](mailto:lishuaihu2010@126.com);

李汉典(2000), 男, 通信作者, 硕士研究生, 研究方向为基于人工智能的新型电力系统运行优化, [2839364349@qq.com](mailto:2839364349@qq.com);

邹谈(2000), 男, 硕士研究生, 研究方向为新型电力系统电压控制, [1664299755@qq.com](mailto:1664299755@qq.com)。