

2-15-2024

## Efficiency Optimization Method for Data Sampling in Power Grid Topology Scheduling Simulation

Yingying Zhao

*State Grid Shanghai Electrical Power Research Institute, Shanghai 200437, China; State Grid Shanghai Municipal Electric Power Company, Shanghai 200125, China, zhaoyy\_sh@163.com*

Pusen Dong

*Beihang University, Beijing 100191, China*

Tianchen Zhu

*Beihang University, Beijing 100191, China*

Fan Li

*State Grid Shanghai Electrical Power Research Institute, Shanghai 200437, China; State Grid Shanghai Municipal Electric Power Company, Shanghai 200125, China*

*See next page for additional authors*

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the Artificial Intelligence and Robotics Commons, Computer Engineering Commons, Numerical Analysis and Scientific Computing Commons, Operations Research, Systems Engineering and Industrial Engineering Commons, and the Systems Science Commons

---

This Special Column: Big Models Enable Energy Internet Planning and Operation is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation. For more information, please contact [xtfzxb@126.com](mailto:xtfzxb@126.com).

---

# Efficiency Optimization Method for Data Sampling in Power Grid Topology Scheduling Simulation

## Abstract

**Abstract:** To address the large simulation computational workload and low simulation speed caused by the scale and complexity of the new power system, a simulation acceleration method for topology scheduling based on the distributed and quantization mechanisms is proposed. The parallelization of topology scheduling models is used to increase the scale of data simulation sampling in unit time. The introduced quantization operators accelerate the computation speed of the topology scheduling model operators, reduces the time cost of the every single simulation. Case studies confirm the effectiveness of the topology simulation acceleration, in which the available transfer capacity of the simulated power system increases by approximately 2 times in unit time and achieves the simulation acceleration by 1.5~2.5 times in various-scale power grid scenarios.

## Keywords

topology scheduling simulation, DRL, sampling efficiency optimization, model quantization, power grid digital twin

## Authors

Yingying Zhao, Pusen Dong, Tianchen Zhu, Fan Li, Yun Su, Zhenying Tai, Qingyun Sun, and Hang Fan

## Recommended Citation

Zhao Yingying, Dong Pusen, Zhu Tianchen, et al. Efficiency Optimization Method for Data Sampling in Power Grid Topology Scheduling Simulation[J]. Journal of System Simulation, 2024, 36(2): 283-295.

**编者按** 践行“碳达峰碳中和”战略,能源是主战场,电力是主力军,能源互联网是重要载体之一。近年来,随着电力能源对可再生能源消纳高效性、多网络耦合交互性、产消者参与调度实时性和经济性的需求不断提升,以及人工智能大模型时代的到来,亟需探究如何应用基于大模型和优化仿真技术适应输/配电网灵活建模、产消者分布式决策、数据收集和处理、需求侧管理、可再生能源发电等不同场景,解决能源互联网规划设计和智能交互、电力市场多能交易、多能潮流计算、区块链模式驱动的能源共享、新型电力系统互联协调和能源管理等研究课题,以促进能源互联网低碳、平稳、安全、经济运行,同时,推进仿真科学技术在能源领域深入应用及发展进步。

为此,《系统仿真学报》本期推出“大模型赋能能源互联网规划运行”专栏。专栏征稿过程中,受到清华大学电机系沈沉教授、武汉大学电气与自动化学院董旭柱教授、中国电力科学研究院技术战略研究中心闫华光主任、华北电力大学电气与电子工程学院陈艳波教授、山东大学电气工程学院赵浩然教授、天津大学电气自动化与信息工程学院葛磊蛟副教授、中国农业大学信息与电气工程学院张璐副教授、北京邮电大学计算机学院侯鲁洋特聘副研究员等多位专家的鼎力支持和指导,共收到了18篇高质量原创性成果投稿,经过学报编委会组织的严格同行评议和审查,最终精选出3篇高水平成果,以飨读者。

## 面向电网拓扑调度仿真的采样效率优化方法研究

赵莹莹<sup>1,2</sup>, 董普森<sup>3</sup>, 朱天晨<sup>3</sup>, 李凡<sup>1,2</sup>, 苏运<sup>1,2</sup>, 邵振赢<sup>3</sup>, 孙庆赞<sup>3</sup>, 凡航<sup>4\*</sup>

(1. 国网上海市电力公司电力科学研究院, 上海 200437; 2. 国网上海市电力公司, 上海 200125;  
3. 北京航空航天大学, 北京 100191; 4. 华北电力大学, 北京 100096)

**摘要:** 为解决新型电力系统的规模性与复杂性导致的仿真计算量较大, 仿真速度相对变缓的问题, 提出了一种基于分布式与量化机制的拓扑调度仿真加速方法。通过拓扑调度模型的并行, 增大了单位时间内数据仿真采样的规模; 引入了量化算子, 加快拓扑调度模型算子计算, 减少单次仿真的时间成本。算例分析验证了拓扑仿真加速的有效性, 在单位时间内将仿真电力系统的可用传输容量提升了约2倍, 并在不同规模电网算例实现了1.5~2.5倍的仿真加速。

**关键词:** 拓扑调度仿真; 深度强化学习; 采样效率优化; 模型量化; 电网数字孪生

中图分类号: TP391; TM769 文献标志码: A 文章编号: 1004-731X(2024)02-0283-13

DOI: 10.16182/j.issn1004731x.joss.23-0958

**引用格式:** 赵莹莹, 董普森, 朱天晨, 等. 面向电网拓扑调度仿真的采样效率优化方法研究[J]. 系统仿真学报, 2024, 36(2): 283-295.

**Reference format:** Zhao Yingying, Dong Pusen, Zhu Tianchen, et al. Efficiency Optimization Method for Data Sampling in Power Grid Topology Scheduling Simulation[J]. Journal of System Simulation, 2024, 36(2): 283-295.

收稿日期: 2023-07-29

修回日期: 2023-10-19

基金项目: 国网上海市电力公司科技项目(B3094022000D); 上海电力人工智能工程技术研究中心研究项目(19DZ2252800)

第一作者: 赵莹莹(1991-), 女, 工程师, 硕士, 研究方向为电力大数据和人工智能技术应用工作。E-mail: zhaoyy\_sh@163.com

通讯作者: 凡航(1993-), 男, 讲师, 博士, 研究方向为电力大数据与人工智能、电碳市场和隐私计算。E-mail: fanhang123456@163.com

## Efficiency Optimization Method for Data Sampling in Power Grid Topology Scheduling Simulation

Zhao Yingying<sup>1,2</sup>, Dong Pusen<sup>3</sup>, Zhu Tianchen<sup>3</sup>, Li Fan<sup>1,2</sup>, Su Yun<sup>1,2</sup>, Tai Zhenying<sup>3</sup>, Sun Qingyun<sup>3</sup>, Fan Hang<sup>4\*</sup>

(1. State Grid Shanghai Electrical Power Research Institute, Shanghai 200437, China; 2. State Grid Shanghai Municipal Electric Power Company, Shanghai 200125, China; 3. Beihang University, Beijing 100191, China; 4. North China Electric Power University, Beijing 100096, China)

**Abstract:** To address the large simulation computational workload and low simulation speed caused by the scale and complexity of the new power system, a simulation acceleration method for topology scheduling based on the distributed and quantization mechanisms is proposed. *The parallelization of topology scheduling models is used to increase the scale of data simulation sampling in unit time. The introduced quantization operators accelerate the computation speed of the topology scheduling model operators, reduces the time cost of the every single simulation.* Case studies confirm the effectiveness of the topology simulation acceleration, in which the available transfer capacity of the simulated power system increases by approximately 2 times in unit time and achieves the simulation acceleration by 1.5~2.5 times in various-scale power grid scenarios.

**Keywords:** topology scheduling simulation; DRL; sampling efficiency optimization; model quantization; power grid digital twin

## 0 引言

随着新型电力系统中新能源占比的不断上升,电网智能调度系统作为电网运行控制的指挥中枢和“决策大脑”,面临源荷双侧不确定性显著增强、调度对象类型和数量指数级增加、运行特性更趋复杂等新挑战<sup>[1]</sup>。与此同时,数字孪生电网建立了从现实空间到虚拟空间的映射,通过在虚拟空间的建模、仿真、预测和控制,以虚控实,能够更有效地推动电网的稳定运行<sup>[2-3]</sup>。如何以数字孪生电网为基座,构建智能运行调度算法及其训练框架,实现能源系统能源供给侧与需求侧充分联动与经济效益优化<sup>[4-5]</sup>,并保证电网潮流运行的平稳性与安全性,对于新型电力系统具有重要的研究意义<sup>[6-7]</sup>。

对电网拓扑进行仿真建模与优化是一种有前景且经济高效的在线运行调度方法,最初的研究主要集中在构建模型驱动的电网仿真模型,并实现多个控制目标,包括成本最小化、电压控制和线路流量控制<sup>[8-9]</sup>。鉴于电网运营的复杂性和不确定性,类似于输电线路切换或母线分割这样的决策生成,需要求解一个非线性混合整数优化问题。早期针对电网

拓扑仿真优化的研究主要集中在如何设计近似算法,高效求解各类最优化模型。例如,文献[10]提出了一种混合整数线性规划模型,通过直流功率流对电网进行近似,并通过 CPLEX 广义优化求解器进行高效求解。文献[11]通过将拓扑优化问题分解为 2 个相对简单的子问题:组合优化和直流最优潮流,并通过 CPLEX 来求解该问题。文献[12]提出了一种启发式方法,通过直流最优潮流来近似解决问题。文献[13-14]提出了类似的方法,利用点估计来建模系统不确定性,并使用最优潮流进行验证。然而,由于电网结构愈发复杂、节点规模不断扩大,固有的组合性质和高度非线性等特点,使传统优化算法面临着非常大的挑战<sup>[15]</sup>。

不少研究人员在数字孪生电网的建模仿真层中引入机器学习、深度强化学习等前沿技术,建立数据驱动的仿真分析模型,并利用其强大的建模和逻辑推演能力来解决电网拓扑仿真优化等调度问题,辅助调度员进行在线决策<sup>[16]</sup>。例如,文献[2]提出了一种交/直流电网实时分析方法。文献[3]提出了一种状态估计方法,用来监测电网的实时运行状态,并对电网状态进行预测。在此基础

上,大量的工作利用以DQN为代表的深度强化学习算法来优化分布式发电系统的发电效率<sup>[17]</sup>,管理电力系统的经济效益<sup>[18]</sup>、可再生能源消纳及服务负荷能力等<sup>[19-20]</sup>。文献[21-22]在考虑各种实际约束的数字孪生电网仿真环境上,训练了一种基于深度强化学习的拓扑优化模型来最大化系统可用传输容量(available transfer capability, ATC)。

由于电网结构愈发复杂、节点规模不断扩大,数字孪生电网的实时仿真计算与分析优化面临着严峻的计算效率的挑战。为了适应新型电力系统的规模性与复杂性,电网的仿真与分析需要较大的仿真计算量。特别是对以DQN为代表的深度强化学习等数据驱动的仿真模型,其数据仿真互动过程需要动态生成大量高维的拓扑调度与潮流仿真数据,并进行反复次的“采样—试错”循环学习,以训练最优拓扑调度策略,因此,仿真分析过程耗时巨大、效率极低,严重限制了仿真的有效性。此外,低效的拓扑仿真优化模型在实时响应中容易导致次优的电网配置,从而破坏电网的稳定性。

因此,为了缓解数字孪生电网下基于深度强化学习的拓扑仿真优化模型的训练低效问题,本文优化了深度强化学习算法的数据采样与训练效率,避免实时场景下由于低效数据互动产生的次优调度策略,损害电力系统的稳态调度与可用传输能力。现有的研究表明,通过在不同的设备上与环境并行交互与学习训练,可以提高数据采样的效率,显著提升深度强化学习模型的训练效率与精度<sup>[23]</sup>。基于此,本文面向新型电力系统中广泛使用的DQN拓扑调度优化模型,提出了一个分布式量化优化的强化学习训练方法。一方面,引入基于分布式深度强化学习(distributed DRL, DDRL)<sup>[24-26]</sup>技术的分布式采样机制,提升数据采样的并行能力,进而提升DQN拓扑优化模型的训练速度;在此基础上,提出基于模型量化<sup>[27]</sup>的量化采样机制,数据互动时将强化学习模型中的参数等值从全精度格式(FP32)转换为较低精度的整数类型(Int8),有效减少存储空间的占用,降低单

次采样的时间成本,加速数字孪生电网下电网拓扑仿真模型的数据互动,进一步提升训练效率。

## 1 电网拓扑调度优化

### 1.1 问题建模

电网拓扑调度优化是指对固定时间窗口内各种不确定场景及组合进行推演,并给出电网网络拓扑调整,即母线的分裂与合母运行、线路的开断和合环以及两者的排列组合等决策。其主要目标是在各种运行场景下最大化电网的可用传输容量(ATC),每个场景包含多个连续时间断面,断面的时间间隔为5 min,优化问题需要考虑每日负荷变化、日前发电计划和实时发电计划调整、母线电压合格范围、检修计划等<sup>[21]</sup>。在控制过程中不考虑通过改变发电机出力 and 切负荷来增加系统可用传输容量。上述电网在线调度问题,通常能够建模为一个马尔可夫决策过程( $S, A, P, R, \gamma$ ),并利用深度强化学习的方法求解。其中, $S$ 为状态空间; $A$ 为动作空间; $P$ 为状态转移函数; $R$ 为某一状态采取某一动作从环境获得的奖励; $\gamma$ 为折扣因子。具体包含以下几项:

(1) 状态: 电网的拓扑配置(电线断开情况、变电站母线分配情况)、每条电线上的电流、每个节点消耗或者生产的电力等。

(2) 动作: 对变电站和电线进行操作,改变电网拓扑,包括母线的分裂与合母运行、线路的开断和合环以及两者的排列组合等。

(3) 转移函数: 采取动作后,电网生成新状态所用到的函数。

(4) 奖励函数: 智能体在某时刻 $t$ 动作后获得的奖励可以根据拓扑优化控制目标定义为

$$r_t = \begin{cases} -1, & \text{任务失败} \\ \frac{1}{N} \sum_{i=1}^N \max\{0, 1 - (P_i/P_i^{\text{lim}})^2\}, & \text{其他} \end{cases} \quad (1)$$

式中: $N$ 为电网的传输线路个数; $P_i$ 和 $P_i^{\text{lim}}$ 分别为线路 $i$ 的潮流值和线路限值。该奖励函数反映了电

力系统的ATC。

该马尔可夫决策过程的求解目标是得到控制策略 $\pi$ ，建立起系统状态与控制动作之间的匹配关系，从而使策略执行过程中得到的预期回报 $J(\pi)$ 最大，亦为马尔可夫决策过程的目标函数<sup>[15]</sup>：

$$J(\pi) = \max_{\pi} E \left[ \sum_{t=0}^T \gamma^t r_t \right] \quad (2)$$

式中： $E[\cdot]$ 为函数的期望； $T$ 为一组时序决策过程中所经历的步数；折扣因子 $\gamma \in [0, 1]$ 。

## 1.2 状态空间

某时段 $t$ 下的电网系统状态可具体描述为

$$\begin{aligned} s_t &= (P_T^1, P_T^2, \dots, P_T^N, Q_T^1, Q_T^2, \dots, Q_T^N, C_t, G, X) \\ P_T^i &= (P_{t-T}^i, P_{t-T+1}^i, \dots, P_{t-1}^i), \forall i \in N \\ Q_T^i &= (Q_{t-T}^i, Q_{t-T+1}^i, \dots, Q_{t-1}^i), \forall i \in N \\ C_t &= (C_t^1, C_t^2, \dots, C_t^N) \end{aligned} \quad (3)$$

式中： $P_T^i$ 为第 $i$ 个母线上过去 $T$ 个时段的有功功率； $Q_T^i$ 为第 $i$ 个母线上过去 $T$ 个时段的无功功率； $C_t$ 为 $t$ 时段所有母线上储能的荷电状态； $G$ 为电网拓扑结构信息； $X$ 为电线状态。系统状态空间 $S$ 包含538个特征，具体包括发电机的有功功率输出和电压设定值、母线等值负荷、线路状态、线路潮流、线路热稳限值、采样时间等。具体状态描述如表1所示。

表1 电网状态空间描述

Table 1 Description of power grid state space		
特征名	大小	描述
时间	6	年、月、周、日、时、分
发电机功率	$2 \times N_{\text{gen}}$	每个发电机的有功和无功功率
负载功率	$2 \times N_{\text{load}}$	每个负载的有功功率和无功功率
电线上的功率	$4 \times N_{\text{line}}$	每条电线两端的有功功率和无功功率
电线容量	$N_{\text{line}}$	每条电线的容量
拓扑信息	$N_{\text{gen}} + N_{\text{load}} + 2 \times N_{\text{line}}$	每个元件连接到了其变电站的哪个母线上
电线状态	$N_{\text{line}}$	每条电线的状态(连接/断开)

## 1.3 动作空间

智能体可以采取母线分配或线路开关2类动作，实现对电网拓扑的控制和优化。动作空间 $A$ 由线路开合、母线分裂运行与恢复以及两者的排列组合组成。母线分配的动作空间大小为 $\sum_{i=0}^{N_{\text{sub}}} 2^{\text{sub}(i)}$ ，

其中， $N_{\text{sub}}$ 代表变电站数量；函数 $\text{sub}(i)$ 代表第 $i$ 个变电站的元件数量。线路开关的动作空间大小为 $5 \times N_{\text{line}}$ ，其中， $N_{\text{line}}$ 代表电网中的线路数量。

## 1.4 约束条件

电力系统的前瞻调度需要满足以下约束。

(1) 支路越限约束：若任一支路的电流值超过其热稳限值，表示该支路电流越限。若支路电流越限但未超过热稳限值的135%，表示支路“软过载”。若支路电流超过热稳限值的135%，表示支路“硬过载”。任意支路连续时长 $t$ 发生“软过载”，则该支路停运。发生“硬过载”则支路立即停运。支路停运16个时间步之后，重新投运。

(2) 机组启停约束：主要涉及火电机组的停运规则。机组停运前机组有功出力必须调整至出力下限，再调整至0。机组停机后在连续时长 $t$ 内不允许重新启动。火电机组启动规则为机组开启前有功出力必须调整至出力下限。机组重新启动后在连续时长 $t$ 内不允许停机。新能源机组和平衡机没有关机状态，一直保持开机。火电机组的关机状态通过判断机组有功出力是否为0来确定。

(3) 拓扑连通约束：对于线路的开关等调度决策，需保持电网拓扑的连通性，不能使网络中出现一个不连通的子图。

(4) 负载电量约束：需保证负载所需的电量被充分满足，不能存在过多未连接的线路，导致发电机的出力没有被完全送达负载。

## 2 基于分布式量化加速的拓扑仿真效率优化方法

### 2.1 强化学习仿真模型的计算效率瓶颈

基于深度强化学习的电网拓扑仿真优化模型, 在仿真与训练过程中通常可以分为采样和优化 2 个阶段。在采样阶段, 智能体通常与电网仿真环境交互, 并从环境中采样电网状态数据, 在此基础上根据状态选择动作反馈到环境中, 进而获取奖励信号, 从而逐渐探索环境。在优化阶段, 智能体则需要基于采样阶段获取到的经验数据, 使用强化学习算法来更新和优化模型参数, 从而最小化损失函数。大多数电网拓扑仿真优化模型均通过 2 个阶段交替进行, 以实现智能体策略的不断迭代优化。事实上, 在一般强化学习模型的训练总时长中, 采样阶段约占总训练时间的 3/4<sup>[27]</sup>, 而在数字孪生电网这类具有组合、高维非线性等特点的复杂仿真环境中, 由于需要更加密集地与环境交互并采集大量数据, 因此, 这一时间比例相对普通环境将严重增加。本文在比较经典的 IEEE 14 算例上仿真并记录了 DQN 算法在训练过程中, 采样和优化阶段分别消耗的执行时间。结果显示采样阶段约占总训练时间的 95%, 如图 1 所示, 这表明在数字孪生电网上, 电网拓扑仿真优化的时间瓶颈在于强化学习的数据采样阶段。

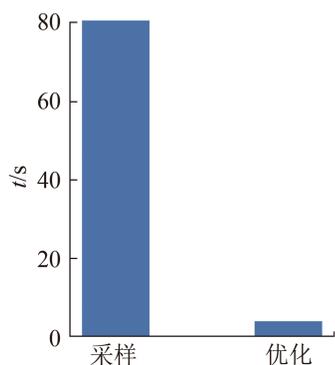


图1 IEEE 14 上 DQN 采样与优化阶段的占用时长分布  
Fig. 1 Distribution of duration for DQN sampling and optimization phases on IEEE 14

### 2.2 基于量化分布式强化学习的高效采样框架

为了加速基于深度强化学习的电网拓扑仿真优化模型的训练, 核心任务是提升算法的采样效率, 降低智能体单次采样的开销。因此, 本文提出了量化分布式强化学习训练架构, 通过分布式的数据采样(模型层)和采样过程的量化加速(计算层)协同优化强化学习算法的采样效率。其中模型层通过堆叠采样模型数量, 提高数据采样的并行能力; 计算层通过优化采样计算模式, 减少单次采样的时间成本。算法整体框架如图 2 所示。

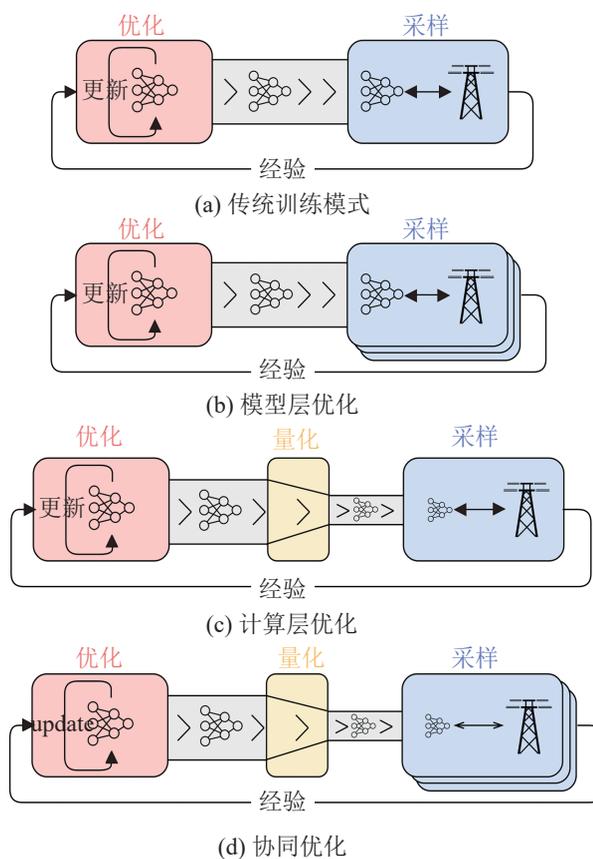


图2 基于量化分布式强化学习高效采样框架  
Fig. 2 Efficient sampling framework based on quantized distributed reinforcement learning

#### 2.2.1 基于量化加速的采样效率优化

在采样模型的计算层面, 为了减少单次采样时间, 一种直观的想法便是通过牺牲数据精度来获得计算加速, 例如, 量化。量化是一个将数据

的连续取值近似为多个离散值的过程，本质上是通过信息压缩实现加速计算<sup>[28-29]</sup>。多数电网拓扑仿真优化模型在训练时均采用浮点型进行数值运算，例如，Float32 和 Float64，目的是获得更高的计算精度。通过将神经网络量化，能够使网络模型的权重、激活值等从连续高精度取值转化为离散低精度，比较常见的是从 Float32 转化为 Int8。Int8 由于数据量小，计算速度更快，整数运算比 Float32 运算快 4 倍。因此，本节提出通过量化深度神经网络来加速强化学习模型的采样效率。然而，量化会给模型带来误差，进而对模型性能造成损失，本文将训练后量化(post training quantization, PTQ)<sup>[29]</sup>应用于电网强化学习训练的采样阶段，建立量化采样机制，并记录精度损失情况。这一方法能对误差进行修正：量化后的模型在推理时会根据校准数据的数据分布动态地调整缩放因子，减小误差。动态量化的机制允许模型在运行时根据实际输入进行量化，因此，能够更好地适应不同的数据分布和变化。

本文使用标准统一仿射量化<sup>[30]</sup>执行训练后量化：

$$Q_n(W) = \text{round}(W/\delta) \quad (4)$$

$$\delta(W) = \frac{|\min(W, 0)| + |\max(W, 0)|}{2^n} \quad (5)$$

式中： $n$ 为量化后的位数； $W$ 为模型参数； $\text{round}(\cdot)$ 为取整函数； $\delta(\cdot)$ 函数用于计算量化时的步长。

反量化过程被定义为

$$D(W_q, \delta) = \delta(W_q) \quad (6)$$

式中： $W_q$ 为量化后的模型参数。

将电网拓扑仿真优化策略以标准的全精度进行训练；在训练完成后，将它们量化为 Int8 类型，策略网络的所有层都被量化到相同的精度，然后对它们进行反量化以模拟量化误差；使用量化后的拓扑优化策略模型与电网环境交互并采样数据，从而提升采样效率。

由于强化学习电网拓扑优化与其他深度学习任务(图像分类、检测等)存在训练方式、时序性等

差异，因此，模型量化操作对模型性能和收敛性的影响也可能存在差异。一般来说，强化学习任务需要在不同状态之间做出连续决策<sup>[31]</sup>，如果在一个状态下做出错误决策，可能会传播到随后的状态，而其他深度学习任务通常只需要一次性地处理输入数据进行预测或分类，不会产生误差传播。此外，强化学习电网拓扑仿真优化任务的模型输入数据中包含电网拓扑信息，而拓扑信息对于数据精度的敏感性较强，数据精度的压缩可能会影响拓扑信息的表达。为了观察量化采样机制对于模型收敛性的影响，本文在 IEEE 5、IEEE 118 电网标准算例集的强化学习训练中，分别测试了简单和复杂 2 类拓扑调度任务上的收敛性。该测试为所有电网算例均训练一个 3 层全连接网络，相邻全连接网络之间使用 ReLU 激活函数提供非线性映射。量化过程中，模型参数均被量化为相同的精度。图 3 显示了量化采样机制对于电网拓扑仿真优化强化学习模型收敛性的影响。结果表明：在使用 8 位精度的量化采样机制时，不仅不影响模型的收敛性，反而能够加速模型的收敛速度，甚至带来小幅的性能提升。造成该结果的原因，一方面是因为电网环境中特征的稀疏性导致量化产生的精度损失处于可以接受的范围，进而没有造成显著的奖励损失。另一方面，量化能够提供随机的扰动<sup>[28]</sup>，这些扰动对于电网环境的充分探索是积极的。

### 2.2.2 基于分布式训练的采样效率优化

本文除了从计算层面优化采样效率，还从模型层面提高采样效率，建立分布式采样机制，即通过增加采样模型的数量，通过提高采样时的并行度来增加相同时间内的采样次数。

本文采用的分布式强化学习算法以 Actor-Learner<sup>[32]</sup>训练框架为基础，如图 4 所示。该训练框架由一个或多个 Actor 根据当前的策略收集训练经验，然后将经验传递给 Learner 或经验回放缓冲区<sup>[33]</sup>。Learner 利用生成的经验和定义好的优化器

来更新自己的策略权重。Learner 定期将策略广播给 Actor, Actor 使用接收到的策略来更新自己内部的策略, 从而使模型同步。

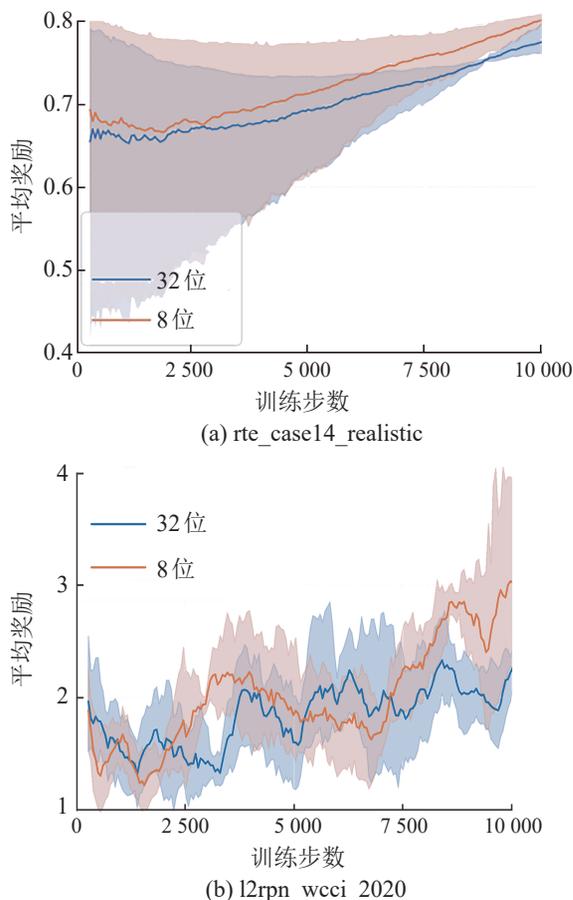


图3 基于量化的数据采样机制对强化学习收敛性的影响  
Fig. 3 Impact of quantized data sampling mechanism on convergence of reinforcement learning

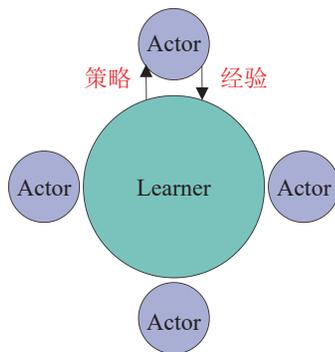


图4 Actor-Learner 训练框架  
Fig. 4 Actor-Learner training framework

在本文的分布式场景下, 这些组件在不同的计算节点上运行, 彼此之间不断通过通信传输采样数据并训练。本文算法主要包含4个组件: 低精度分布式采样器、全精度优化器、量化器、经验回放缓冲区, 如图5所示。低精度分布式采样器用于创建多个采样进程, 生成用于模型优化的经验数据, 每个采样进程被绑定到CPU的不同核心上。采样器接收量化器产生的低精度量化策略作为采样策略, 以提高采样效率。优化器在整个策略优化过程中则需保持全精度, 从而保证优化过程的收敛性。优化器会定期向量化器发送最新的全精度电网策略, 量化器接收到策略后会根据式(5)进行训练后量化, 生成量化后的低精度策略。采样器产生的经验数据被存储在经验回放缓冲区中, 该模块会根据每个经验的TD误差<sup>[34]</sup>赋予其一个优先级, 用于决定该经验被优化器选取的概率。

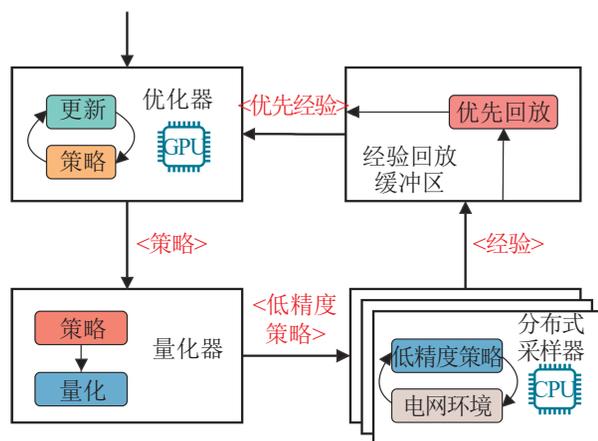


图5 基于分布式训练的强化学习拓扑优化模型训练框架  
Fig. 5 Training framework for reinforcement learning-based topology optimization model using distributed training

本文从计算和模型2个层面优化强化学习模型的采样效率, 进而提升电网数字孪生模型的动态仿真分析效率。在计算层面用标准统一仿射量化法改进了DQN模型, 替代了传统DQN高精度的浮点数表示; 在模型层面增加模型并行数量, 扩大采样频率。此外, 所提出的DQN采样加速算

法同样可以泛化到其他异策略深度强化学习方法中，例如，DDPG(deep deterministic policy gradient)、ACER(actor-critic with experience replay)等。因为这些方法的数据采样与利用方式与DQN模型类似，所以可以采用相似的方法对其进行量化，以提高计算和存储效率，并尽量减少量化误差。

### 3 实验验证

为了验证基于量化分布式采样的深度强化学习的加速效果，本文在3个电网标准算例集上的仿真实验中开展了强化学习模型的加速比、收敛性等测试与分析。

#### 3.1 实验设置

智能体的训练环境采用开源电网仿真软件Grid2op<sup>[22]</sup>，该仿真环境建立在用于模拟电网的Pypowernet等开源工具的基础之上。Grid2op能够对大型电网进行建模并仿真各种系统运行状态，同时支持交流和直流潮流求解。Grid2op的框架基于Linux操作系统，接口可以通过自定义方式进行设置，从而可以与强化学习智能体进行交互。通过与Grid2op的大量交互，使用Python脚本训练和调整智能体。Grid2op提供了多个标准算例集，数据集代表了实际的时间序列运行条件。本文以IEEE 5、IEEE 14、IEEE 118这3个标准算例作为研究对象，并使用开源数据集进行测试。在IEEE 5上使用rte\_case5\_example<sup>[22]</sup>开源数据集，该数据集包含20个场景共7天的连续电网运行断面数据，每个场景包含2 016个连续断面；在IEEE 14上使用l2rpn\_case14\_sandbox<sup>[22]</sup>开源数据集，该数据集包含1 000个场景共28天的连续电网运行断面数据，每个场景包含8 065个连续断面；在IEEE 118上使用l2rpn\_wcci\_2020<sup>[22]</sup>开源数据集，该数据集包含240个场景共1年的连续电网运行断面数据，每个场景包含105 120个连续断面。3个数据集中断面

的时间间隔均为5 min。电网算例的详细统计信息如表2所示。

表2 电网标准算例的详细信息  
Table 2 Detailed information of power grid standard test cases

电网算例	变电站数量	线路数量	发电机数量	负载数量	状态空间维度	是否允许调整发电值
IEEE 5	5	8	2	3	74	否
IEEE 14	14	20	6	11	194	是
IEEE 118	36	59	22	37	590	是

本文采用Double Dueling DQN算法<sup>[35-36]</sup>作为基线算法，值函数和优势函数用两个隐藏层大小为1 024的全连接网络来表示。在采样阶段，本文采用Epsilon-greedy<sup>[37]</sup>方法来鼓励探索，初始的探索概率设置为0.99，代表与环境交互时，有99%的概率随机选取动作，1%的概率选取由策略生成的动作。

实验所用的硬件包括1块128核的Intel(R) Xeon (R) Platinum 8338C CPU，1块NVIDIA GeForce RTX 4090 GPU，具体硬件环境需求如表3所示。

表3 实验所用硬件环境  
Table 3 Hardware environment used for experiments

硬件	信息类型	信息
CPU	型号	Intel(R) Xeon(R) Platinum 8338C CPU
	架构	x86_64
	核数	128
GPU	型号	NVIDIA GeForce RTX 4090
	显存/MB	24 564
内存	总大小/MB	515 415

#### 3.2 加速效果分析

本文在IEEE 5、IEEE 14、IEEE 118这3个仿真环境上测试了所提算法的加速效果。图6分别记录了在应用经典DQN算法、PPO算法与本文的加速算法下，智能体训练过程中的“时间-奖励”收敛曲线。由图6可以看出，本文提出的量化分布式采样机制相比经典DQN加速比提供约1.5~2.5倍。

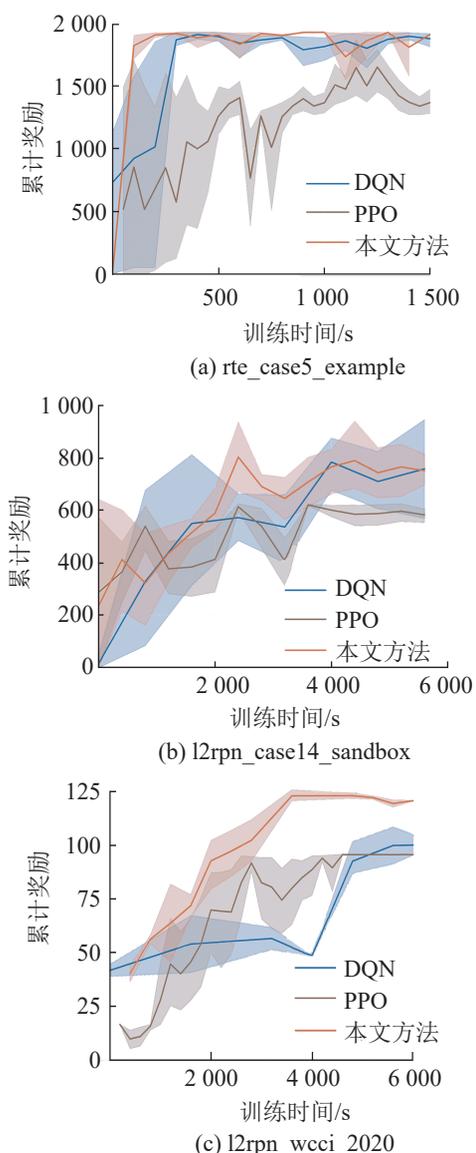


图 6 “时间-奖励”收敛性分析  
Fig. 6 Convergence analysis of "time-rewards"

除此之外, 在 3 个仿真环境上, 本文方法可以在相同的训练时间下, 获得更高的累积奖励, 如表 4 所示。结果表明: 本文方法可在相同的训练时间下, 显著提升电网系统的 ATC。

### 3.3 量化采样机制的收敛性分析

为了研究量化采样机制对于电网强化学习训练收敛性的影响, 本文记录了训练过程中经典 DQN 算法与本文算法“步数-奖励”的收敛曲线, 如图 7 所示。对于较为简单的 IEEE 5 任务, 引入

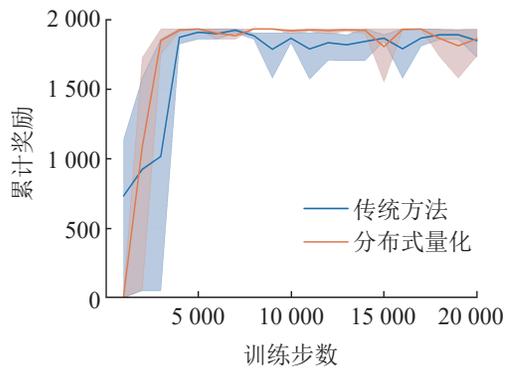
量化采样机制的 DQN 算法能够更快收敛, 并且获得的累积奖励更大。这是由于量化操作带来了可以被视为提升“探索多样性”的噪声, 使智能体可探索到一些未知状态, 从而找到一些更优的未探索过的新动作。对于 IEEE 14 和 IEEE 118 任务, 引入量化采样机制后, 模型收敛曲线上的阴影区域更大, 代表累积奖励的方差较大, 这是由量化带来的噪声引起的。因为量化后的策略模型具有更大的不确定性, 会使智能体尝试一些非常规动作, 可能会执行一些具有较高风险的动作, 导致获得奖励的不确定性和方差增大。具体来说, 模型量化额外引入的累积误差, 在特定的状态下会使智能体的策略分布产生较大变化, 引导智能体选择次优甚至恶劣的动作。在这个过程中智能体可能会遭受惩罚或低奖励, 也有可能获得超出寻常动作所能带来的奖励值, 从而导致结果的方差变大。总结而言, 尽管引入量化采样机制的强化学习电网拓扑仿真优化模型的训练稳定性轻微下降, 但是算法仍可以在 ATC 性能损失极小的情况下保持模型收敛。

表 4 有限训练时长下的累积奖励(ATC)对比  
Table 4 Comparison of accumulated rewards (ATC) in limited training time

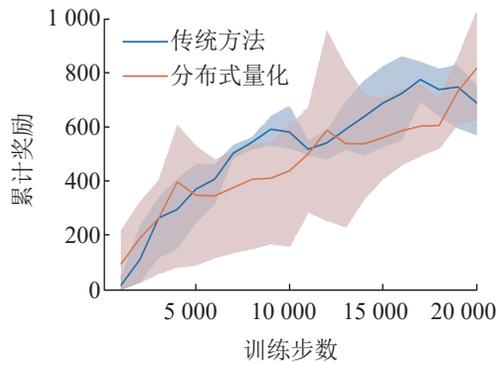
电网环境	方法	1/4 T	2/4 T	3/4 T	T
IEEE 5	DQN	254	430	842	1 029
	PPO	423	734	524	630
	本文方法	<b>653</b>	<b>932</b>	<b>1 420</b>	<b>1 924</b>
IEEE 14	DQN	215	465	560	564
	PPO	372	410	522	600
	本文方法	<b>217</b>	<b>463</b>	<b>564</b>	<b>783</b>
IEEE 118	DQN	43	48	52	58
	PPO	16	52	70	86
	本文方法	<b>52</b>	<b>72</b>	<b>93</b>	<b>120</b>

### 3.4 消融实验

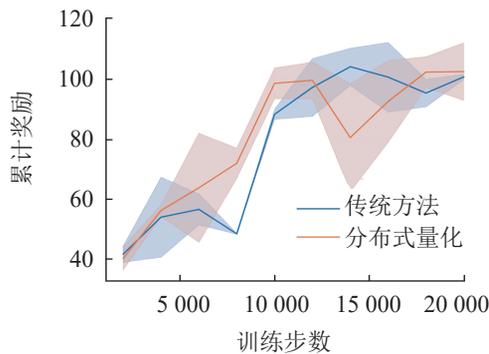
为了验证本文所提出的 2 个加速采样机制的有效性, 开展消融实验测试 2 个机制对强化学习电网拓扑仿真优化模型的加速效果。



(a) rte\_case5\_example



(b) l2rpn\_case14\_sandbox



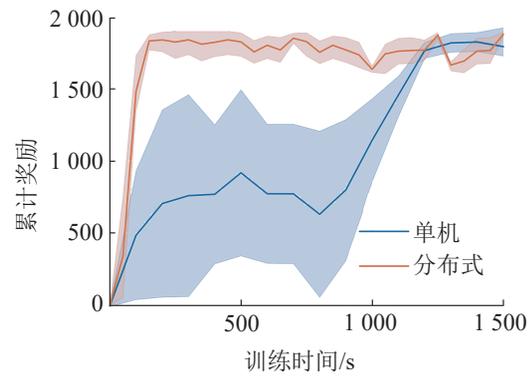
(c) l2rpn\_wcci\_2020

图7 “步数-奖励”收敛性分析

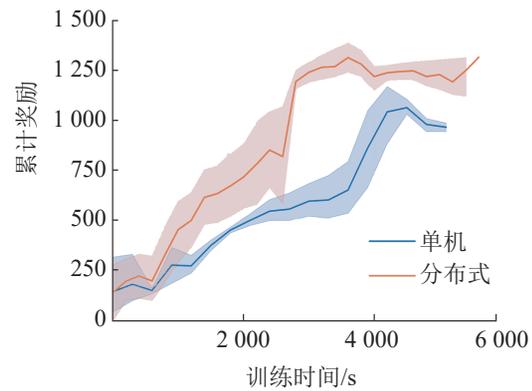
Fig. 7 Convergence analysis of "steps-rewards"

### 3.4.1 仅使用分布式采样机制的加速效果分析

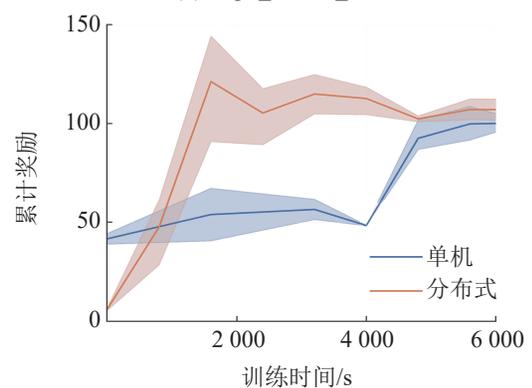
该实验分别在3个电网环境上对比了仅使用1个Actor(单机)和使用4个Actor(分布式)的训练加速效果。如图8所示,在较为简单的IEEE 5环境上,分布式采样的加速效果十分明显。而在任务更加复杂的IEEE 14和IEEE 118环境上,尽管加速效果不如在IEEE 5环境上明显,但仍有相当可观的加速效果,总体加速比约为1.5~2倍。



(a) rte\_case5\_example



(b) l2rpn\_case14\_sandbox



(c) l2rpn\_wcci\_2020

图8 仅采用分布式采样机制的加速效果对比

Fig. 8 Comparison of acceleration effect using only distributed sampling mechanism

### 3.4.2 仅使用量化采样机制的加速效果分析

该实验在3个电网仿真环境上对比了全精度采样(原始模型, FP32)和低精度采样(量化模型, Int8)2种情况下的加速比及训练效果。如图9所示,在较为简单的IEEE 5环境,由于收敛较快,量化采样机制并未明显地体现出加速效果。而任务较复杂、拓扑规模较大的IEEE 14以及IEEE 118

环境上, 使用低精度采样(Int8)有较为明显加速效果, 加速比约为 1.2~1.5 倍。

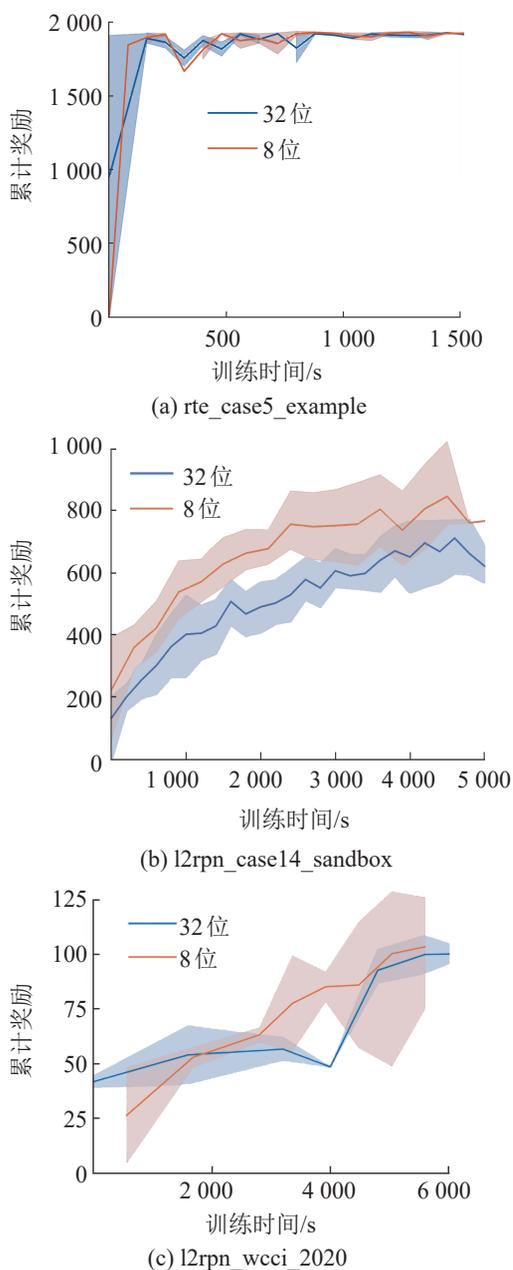


图 9 仅采用量化采样机制的加速效果对比  
Fig. 9 Comparison of acceleration effect using only quantized sampling mechanism

根据消融实验的结果, 本文总结出以下结论:  
①本文所提的 2 个加速采样机制在不同难度的任务中均能体现出 1.5 倍左右的加速效果, 其中, 分布式采样机制相较量化采样机制的加速效果更为显著。  
②同时使用 2 种加速机制协同优化模型采

样比单独使用分布式或单独使用量化的加速效果更为显著。

## 4 结论

数字孪生电网的拓扑智能调度仿真过程面临训练时间长、资源消耗大等问题, 严重制约仿真分析的性能。本文分析了电网仿真建模分析过程中, 以强化学习算法为代表的驱动模型的训练性能瓶颈, 并提出优化强化学习数据采样速度是提升拓扑智能调度仿真效率的关键。本文提出了一种基于分布式与量化机制的 DQN 高效数据采样与训练方法, 用标准统一仿射量化法改进了 DQN 模型, 替代了传统 DQN 高精度的浮点数表示, 有效解决 DQN 一类的强化学习模型与电网机理模型间数据互动低效的问题。不同规模算例上的仿真结果表明: 在有限的训练时长下, 本文所提出的方法能够大幅提升系统 ATC 性能。同时在保持或小幅提升 ATC 的前提下, 可将基于强化学习的拓扑仿真分析模型的训练效率和收敛速度提升 1.5~2.5 倍。在完整的数字孪生电网运行调度仿真中, 除了离散空间下的拓扑调度, 还涉及连续空间下的功率、电压调度等广泛场景, 如何将所提算法迁移到此类场景中是未来的一项研究课题。另外, 本文方法是在软件层面上加速仿真计算的, 考虑如何融合异构硬件计算等软硬结合的方式加速也是一项亟待研究的问题。此外, 本文所提算法在 DQN 等同策略强化学习算法上具有较强的泛用性, 如何在 PPO 等异策略强化学习算法上应用量化机制以实现采样加速仍有待研究。

## 参考文献:

[1] 董伟杰, 刘科研, 王义龙, 等. 高比例分布式发电接入配电网自适应控制方法[J]. 系统仿真学报, 2020, 32(10): 2052-2058.  
Dong Weijie, Liu Keyan, Wang Yilong, et al. Adaptive Control Method of High Proportion Distributed Generation Connected to Distribution Network[J]. Journal of System Simulation, 2020, 32(10): 2052-2058.  
[2] Zhou Mike, Yan Jianfeng, Zhou Xiaoxin. Real-time

- Online Analysis of Power Grid[J]. CSEE Journal of Power and Energy Systems, 2020, 6(1): 236-238.
- [3] Song Xinya, Cai Hui, Jan Kircheis, et al. Application of Digital Twin Assistant-system in State Estimation for Inverter Dominated Grid[C]//2020 55th International Universities Power Engineering Conference (UPEC). Piscataway, NJ, USA: IEEE, 2020: 1-6.
- [4] 陈强, 王意, 李康顺. 双重需求响应的虚拟电厂建模与调度研究[J]. 系统仿真学报, 2023, 35(4): 822-832.  
Chen Qiang, Wang Yi, Li Kangshun. Research on Modeling and Scheduling of Virtual Power Plant with Dual Demand Response[J]. Journal of System Simulation, 2023, 35(4): 822-832.
- [5] 沈沉, 曹仟妮, 贾孟硕, 等. 电力系统数字孪生的概念、特点及应用展望[J]. 中国电机工程学报, 2022, 42(2): 487-498, 中插4.  
Shen Chen, Cao Qianni, Jia Mengshuo, et al. Concepts, Characteristics and Prospects of Application of Digital Twin in Power System[J]. Proceedings of the CSEE, 2022, 42(2): 487-498, 中插4.
- [6] 冯昌森, 沈佳静, 赵崇娟, 等. 基于合作博弈的智慧能源社区协同运行策略[J]. 电力自动化设备, 2021, 41(4): 85-93.  
Feng Changsen, Shen Jiajing, Zhao Chongjuan, et al. Cooperative Game-based Coordinated Operation Strategy of Smart Energy Community[J]. Electric Power Automation Equipment, 2021, 41(4): 85-93.
- [7] 王珂, 姚建国, 余佩遥, 等. 基于深度强化学习的电网前瞻调度智能决策架构及关键技术初探[J]. 中国电机工程学报, 2022, 42(15): 5430-5438, 中插4.  
Wang Ke, Yao Jianguo, Yu Peiyao, et al. Architecture and Key Technologies of Intelligent Decision-making of Power Grid Look-ahead Dispatch Based on Deep Reinforcement Learning[J]. Proceedings of the CSEE, 2022, 42(15): 5430-5438, 中插4.
- [8] Glavitsch H. Switching as Means of Control in the Power System[J]. International Journal of Electrical Power & Energy Systems, 1985, 7(2): 92-100.
- [9] Mazi A A, Wollenberg B F, Hesse M H. Corrective Control of Power System Flows by Line and Bus-bar Switching[J]. IEEE Transactions on Power Systems, 1986, 1(3): 258-264.
- [10] Fisher E B, O'Neill R P, Ferris M C. Optimal Transmission Switching[J]. IEEE Transactions on Power Systems, 2008, 23(3): 1346-1355.
- [11] Khodaei A, Shahidehpour M. Transmission Switching in Security-constrained Unit Commitment[J]. IEEE Transactions on Power Systems, 2010, 25(4): 1937-1945.
- [12] J David Fuller, Raynier Ramasra, Amanda Cha. Fast Heuristics for Transmission-line Switching[J]. IEEE Transactions on Power Systems, 2012, 27(3): 1377-1386.
- [13] Dehghanian P, Wang Yaping, Gurunath Gurralla, et al. Flexible Implementation of Power System Corrective Topology Control[J]. Electric Power Systems Research, 2015, 128: 79-89.
- [14] Alhazmi M, Dehghanian P, Wang Shiyuan, et al. Power Grid Optimal Topology Control Considering Correlations of System Uncertainties[J]. IEEE Transactions on Industry Applications, 2019, 55(6): 5594-5604.
- [15] 魏利胜, 杨奔奔, 孙瑞霞. 基于新型BBO算法的微电网优化调度研究[J]. 系统仿真学报, 2023, 35(5): 1075-1085.  
Wei Lisheng, Yang Benben, Sun Ruixia. Research on Optimal Scheduling of Microgrid Based on NBBO Algorithm[J]. Journal of System Simulation, 2023, 35(5): 1075-1085.
- [16] Duan Jiajun, Shi Di, Diao Ruisheng, et al. Deep-reinforcement-learning-based Autonomous Voltage Control for Power Grid Operations[J]. IEEE Transactions on Power Systems, 2020, 35(1): 814-817.
- [17] Xi Lei, Yu Lu, Xu Yanchun, et al. A Novel Multi-agent DDQN-AD Method-based Distributed Strategy for Automatic Generation Control of Integrated Energy Systems[J]. IEEE Transactions on Sustainable Energy, 2020, 11(4): 2417-2426.
- [18] 马苗苗, 董利鹏, 刘向杰. 基于Q-learning算法的多智能体微电网能量管理策略[J]. 系统仿真学报, 2023, 35(7): 1487-1496.  
Ma Miaomiao, Dong Lipeng, Liu Xiangjie. Energy Management Strategy of Multi-agent Microgrid Based on Q-learning Algorithm[J]. Journal of System Simulation, 2023, 35(7): 1487-1496.
- [19] 王新迎, 赵琦, 赵黎媛, 等. 基于深度Q学习的电热综合能源系统能量管理[J]. 电力建设, 2021, 42(3): 10-18.  
Wang Xinying, Zhao Qi, Zhao Liyuan, et al. Energy Management Approach for Integrated Electricity-heat Energy System Based on Deep Q-learning Network[J]. Electric Power Construction, 2021, 42(3): 10-18.
- [20] 李嘉文, 余涛, 张孝顺, 等. 基于改进深度确定性梯度算法的AGC发电功率指令分配方法[J]. 中国电机工程学报, 2021, 41(21): 7198-7211, 中插2.  
Li Jiawen, Yu Tao, Zhang Xiaoshun, et al. AGC Power Generation Command Allocation Method Based on Improved Deep Deterministic Policy Gradient Algorithm[J]. Proceedings of the CSEE, 2021, 41(21): 7198-7211, 中插2.
- [21] 周毅, 周良才, 丁佳立, 等. 基于深度强化学习的电网拓扑优化及潮流控制[J]. 上海交通大学学报, 2021, 55(增2): 7-14.

- Zhou Yi, Zhou Liangcai, Ding Jiali, et al. Power Network Topology Optimization and Power Flow Control Based on Deep Reinforcement Learning[J]. Journal of Shanghai Jiaotong University, 2021, 55(S2): 7-14.
- [22] Antoine Marot, Benjamin Donnot, Gabriel Dulac-Arnold, et al. Learning to Run a Power Network Challenge: a Retrospective Analysis[C]//Proceedings of the NeurIPS 2020 Competition and Demonstration Track. Chia Laguna Resort, Sardinia, Italy: PMLR, 2021: 112-132.
- [23] 李宏仲, 王磊, 林冬, 等. 多主体参与可再生能源消纳的 Nash 博弈模型及其迁移强化学习求解[J]. 中国电机工程学报, 2019, 39(14): 4135-4149.
- Li Hongzhong, Wang Lei, Lin Dong, et al. A Nash Game Model of Multi-agent Participation in Renewable Energy Consumption and the Solving Method Via Transfer Reinforcement Learning[J]. Proceedings of the CSEE, 2019, 39(14): 4135-4149.
- [24] Mehdi Ahrarinouri, Mohammad Rastegar, Karami K, et al. Distributed Reinforcement Learning Energy Management Approach in Multiple Residential Energy hubs[J]. Sustainable Energy, Grids and Networks, 2022, 32: 100795.
- [25] Qin Yude, Ke Ji, Wang Biao, et al. Energy Optimization for Regional Buildings Based on Distributed Reinforcement Learning[J]. Sustainable Cities and Society, 2022, 78: 103625.
- [26] 林永君, 陈鑫, 杨凯, 等. 含多微网的主动配电网双层分布式优化调度[J]. 系统仿真学报, 2022, 34(11): 2323-2336.
- Lin Yongjun, Chen Xin, Yang Kai, et al. Bilevel Distributed Optimal Dispatch of Active Distribution Network with Multi-microgrids[J]. Journal of System Simulation, 2022, 34(11): 2323-2336.
- [27] Krishnan S, Lam M, Chitlangia S, et al. QuaRL: Quantization for Fast and Environmentally Sustainable Reinforcement Learning[EB/OL]. (2022-11-14) [2023-06-30]. <https://arxiv.org/abs/1910.01055>.
- [28] Christos Louizos, Matthias Reisser, Blankevoort T, et al. Relaxed Quantization for Discretized Neural Networks [EB/OL]. (2018-10-03) [2023-06-30]. <https://arxiv.org/abs/1810.01875>.
- [29] Nagel M, Fournarakis M, Amjad R A, et al. A White Paper on Neural Network Quantization[EB/OL]. (2021-06-15) [2023-06-30]. <https://arxiv.org/abs/2106.08295>.
- [30] Krishnamoorthi R. Quantizing Deep Convolutional Networks for Efficient Inference: A Whitepaper[EB/OL]. (2018-06-21) [2023-06-30]. <https://arxiv.org/abs/1806.08342>.
- [31] Lillicrap T P, Hunt J J, Pritzel A, et al. Continuous Control with Deep Reinforcement Learning[EB/OL]. (2019-07-05) [2023-06-30]. <https://arxiv.org/abs/1509.02971>.
- [32] Horgan D, Quan J, Budden D, et al. Distributed Prioritized Experience Replay[J]. (2018-03-02) [2023-06-30]. <https://arxiv.org/abs/1803.00933>.
- [33] Zhang Shangdong, Richard S Sutton. A Deeper Look at Experience Replay[EB/OL]. (2018-04-30) [2023-06-30]. <https://arxiv.org/abs/1712.01275>.
- [34] Sutton R S, Barto A G. Introduction to Reinforcement Learning[M]. Cambridge: MIT Press, 1998.
- [35] Wang Ziyu, Schaul T, Hessel M, et al. Dueling Network Architectures for Deep Reinforcement Learning[C]// Proceedings of the 33rd International Conference on Machine Learning. Chia Laguna Resort, Sardinia, Italy: PMLR, 2016: 1995-2003.
- [36] Hado van Hasselt, Guez A, Silver D. Deep Reinforcement Learning with Double Q-learning[C]// Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. Palo Alto, CA, USA: AAAI Press, 2016: 2094-2100.
- [37] Alexandre dos Santos Mignon, Ricardo Luis de Azevedo da Rocha. An Adaptive Implementation of  $\epsilon$ -greedy in Reinforcement Learning[J]. Procedia Computer Science, 2017, 109: 1146-1151.