

RESEARCH

Open Access



Power flow adjustment for smart microgrid based on edge computing and multi-agent deep reinforcement learning

Tianjiao Pu¹, Xinying Wang¹, Yifan Cao², Zhicheng Liu², Chao Qiu² , Ji Qiao^{1*} and Shuhua Zhang¹

Abstract

In current power grids, a massive amount of power equipment raises various emerging requirements, e.g., data perception, information transmission, and real-time control. The existing cloud computing paradigm is stubborn to address issues and challenges such as rapid response and local autonomy. Microgrids contain diverse and adjustable power components, making the power system complex and difficult to optimize. The existing traditional adjusting methods are manual and centralized, which requires many human resources with expert experience. The adjustment method based on edge intelligence can effectively leverage ubiquitous computing capacities to provide distributed intelligent solutions with lots of research issues to be reckoned with. To address this challenge, we consider a power control framework combining edge computing and reinforcement learning, which makes full use of edge nodes to sense network state and control power equipment to achieve the goal of fast response and local autonomy. Additionally, we focus on the non-convergence problem of power flow calculation, and combine deep reinforcement learning and multi-agent methods to realize intelligent decisions, with designing the model such as state, action, and reward. Our method improves the efficiency and scalability compared with baseline methods. The simulation results demonstrate the effectiveness of our method with intelligent adjusting and stable operation under various conditions.

Keywords: Power flow adjustment, Edge intelligence, Microgrid management, Machine learning, Edge computing

Introduction

With the continuous evolution and innovation of current power construction, millions of smart devices collaborate in the power grid to support various services, and a considerable number of heterogeneous data will be generated and transmitted. The diversified demand for data analysis and processing lists serious requirements to the power system. In the face of rapid response and real-time interaction, traditional fixed allocation of resources has a series of shortcomings in scalability, utilization efficiency and deployment cost. Unlike centralized cloud computing suffered from various pressures in data collection, analysis

and processing, edge computing can realize rapid perceptual response and support regional autonomy, which has become a promising way of following the trend of intelligent power grids.

As an important research issue of smart grid, power flow calculation determines the steady-state parameters of the power system according to the given structure and operating values, which can evaluate the impact of power supply and demand changes on safety. However, this problem will encounter a non-convergence situation under different conditions, and previous solutions rely on both expert experience and human resources. Further, an intelligent power grid can dynamically adjust its setting when the environment changes, and different power units have a variety of optional configurations, which brings more restrictions and uncertainties for the management of microgrids.

*Correspondence: qiaojj@epri.sgcc.com.cn

¹China Electric Power Research Institute, Haidian District, 100192 Beijing, China
Full list of author information is available at the end of the article

Edge intelligence is a combination of edge computing and artificial intelligence [1], which has been applied to power networks with positive results by some studies. In terms of edge computing, some works advocate that enabling smart grid with edge computing to overcome the defects of bandwidth and latency in cloud computing, and produce a large number of application basis and design ideas. In terms of artificial intelligence, some studies focus on how to apply feature engineering or expert systems to manage power flow. However, some of them have shortcomings in scalability and performance. How to carry out power flow adjustment with edge intelligence still needs to be considered. In this paper, we consider the problem of power adjustment and propose the framework of multi-agent deep reinforcement learning and edge computing for distributed power control in microgrids. Firstly, we analyze the typical service requirements of power calculation in the microgrid and propose the entire framework with three different aspects. Then, we model the power flow adjustment problem with Markov processes and design a learning-based adjustment algorithm for microgrids. Finally, with the IEEE 39 bus system simulated by the tool Pandapower [2], the experimental results demonstrate that the proposed framework can effectively obtain solutions.

The main work presented in this paper is summarized as follows: 1) We present a comprehensive framework for smart grid management and control, which enables the data sensing, processing and controlling of smart grids to realize the functions of real-time response and local autonomy. 2) A learning-based distributed algorithm is presented for power flow adjustment, considering system requirement and current state. The simulation results demonstrate that our framework can obtain successful adjustment results under various power conditions.

After introducing the research background, we summarize some related works in “[Related works](#)” section and propose our framework with learning-based decision algorithm in “[The framework of power flow adjustment based on edge intelligence](#)” and “[Automatic adjustment of power flow convergence based on DRL](#)” sections. Then, we present the configuration and evaluation results of simulation experiments in “[Numerical results](#)” section. Finally, we conclude the paper and detail further work in “[Conclusion](#)” section.

Related works

Power flow adjustment is considered as an emerging problem in smart microgrids. As a dynamic decision problem under uncertainty, emergency control of power systems is generally regarded as the last safety net for grid resiliency [3]. Due to the complexity of power demand and supply, the stability of a power system is dependent on multiple

adjustable power devices, which mathematically is essentially the solution of nonlinear equations. Previous works have carried out some studies on power flow control. However, applying edge intelligence to the adjustment of power flow still needs to be addressed.

Smart grids based on edge computing have recently triggered an unprecedented upsurge, changing the model of power management in the past. Different from some general designs on edge computing [4–6], Trajano designs an edge computing-based architecture to support the implementation of smart grid applications, which provides a stable and low latency communication network to achieve an effective end-to-end power management [7]. With a hardware-implemented architecture, Barik adopts the concept of edge computing in smart grids to migrate task loads from the cloud, resulting in improved performance metrics in power consumption, storage requirements, and analysis capabilities [8]. Huang considers an edge computing-based framework to realize real-time monitoring with an efficient heuristic algorithm, which can significantly optimize the frame rate as well as the detection delay compared with cloud framework [9]. Similarly, Awadi considers detecting abnormal samples in electricity consumption records in advance through the collaboration of distributed devices based on edge computing. The paper tests the performance of the proposed model on service latency and network resilience [10]. To process, analyze and store power consumption information, Chen proposes a smart grid system based on IoT and mobile edge computing, and demonstrates that the proposed system supports substantial terminal management, real-time analysis and massive data processing [11]. The above works propose a series of architectures and frameworks for applying edge computing to smart grids. However, they do not specifically consider the application of edge intelligence to microgrids. Albataineh proposes a two-level solution that combines the advantages of cloud computing for power distribution and edge computing for power information processing, which a learning-based engine can establish the communication between the two levels. This engine is enable the system to load balance between the cloud and the edge, which can achieve a higher power grid throughput and power utilization [12].

Different from some papers on general resource management in edge computing [13–15], it is worth noting that this paper applies edge intelligence to distributed grids, but does not consider power flow calculations between microgrids. Along with power consumers’ increasing demand for power services, the microgrid framework is increasingly seen as a hot issue in current smart grids. Yang uses deep reinforcement learning to design an online scheduling strategy to manage energy dispatch in microgrids under uncertainties of energy gen-

eration [16]. Fang considers an economic dispatch problem in microgrids and proposes a learning-based cooperative auction algorithm, which has the advantages of avoiding single point of failure and strong scalability [17]. Ji proposes a learning-based microgrid scheduling strategy for economic energy management, which does not require an explicit model that requires predictors to estimate stochastic variables with uncertainties [18]. Etemad puts forward a learning-based charging strategy for microgrid batteries with renewable energy to improve electrical stability, power quality and the peak power load [19]. Liu proposes a collaborative reinforcement learning method to address a distributed scheduling problem in microgrid, which reduces the coupling of nodes in the microgrid and improves the efficiency of distributed scheduling [20]. Brida proposes a data-driven reinforcement learning method to generate optimal scheduling strategies for given system states [21]. Dabbaghjamanesh proposes a deep learning algorithm with gated recurrent unit to obtain the optimal decision of reconfigurable microgrids. The algorithm learns the network topology characteristics that vary with time and make real-time reconfiguration decisions [22]. The above works present a series of strategies and approaches for economic energy management and show that the application of edge intelligence to microgrid management can effectively improve various performance indicators. However, they do not specifically consider the dynamic configuration of microgrids.

For the problem itself, Ma discusses the application difficulties of deep learning in power flow calculation, proposes the network structure and training process of a deep neural network, as well as the method to solve the over-fitting problem [23]. Aiming at the non-convergence problem of power flow calculation in large-scale power grids, Wang combines professional experience with artificial intelligence to propose a learning-based power flow adjustment method [24]. To quantifying the impact of the wind speeds correlation among multiple wind power stations, Zhu proposes a probabilistic power flow calculation framework with a learning-based distribution estimation approach [25]. A learning-based approach is proposed by Yang to speed up the calculation process of probabilistic power flow problem. The performance differences among neural networks with various structures are compared, and three kinds of power bus systems are used for evaluation benchmark. Compared with the pure data-driven deep learning method, the proposed method can comprehensively improve the approximate accuracy and training speed [26]. Compared with the current situation that learning-based approaches are mostly proposed to identify and evaluate system situations, Su proposes a power system control method with deep belief network [27]. Huang proposes an adaptive emergency control scheme based on the feature extraction and nonlinear

generalization capabilities of deep reinforcement learning for complex power systems [28]. Some of the above works consider how to apply the deep learning method to the power flow calculation problem. However, the research on the application of edge intelligence to the problem of microgrids is still in the preliminary stage.

From the viewpoint of the literature, few research works have considered how to apply edge intelligence to the power flow calculation of microgrids. The existing methods have poor adaptability to the edge computing framework and are unable to deal with local autonomy, or lead to the failure of calculation result, thus leading to system instability. Different from the above works, our research proposes a power flow adjustment framework based on edge computing and multi-agent learning. Considering the complexity of the power flow, we focus on the situation that the system does not converge, proposes our learning-based distributed framework to tackle this problem.

The framework of power flow adjustment based on edge intelligence

Framework overview

As shown in Fig. 1, we consider the power flow adjustment framework based on edge intelligence from the following three aspects: architecture, function and application. First, we introduce the framework based on edge intelligence to connect three kinds of computing entities, namely cloud server, edge node, and end device, using ubiquitous communication networks. The term cloud refers to the data center using cloud computing technology, which can uniformly manage multiple power regions, coordinate decision-making content between power regions, gather and analyze power sensing data. Although the cloud has powerful computing capabilities and extensive network coverage, the network distance to end devices results in a noticeable transmission overhead. The term end refers to the power sensing equipment that senses the environment and the power control equipment that executes the action in the power network, which can directly monitor, collect or perceive the running condition. As a key component of edge intelligence, edges realize nearby computation and data processing through edge nodes, play the role of connecting cloud and end architecturally. Edges are closer to underlying end devices than the cloud server, and can provide a better application experience for end devices through collaborative computing technology.

From the perspective of environment sensing, the process of power flow adjustment mainly includes three steps: data processing, task scheduling and system evaluation. Firstly, data processing, as the basic function of power flow adjustment, needs to perform multi-dimensional data collection from the power network and perform processing steps such as filtering, conversion, aggregation

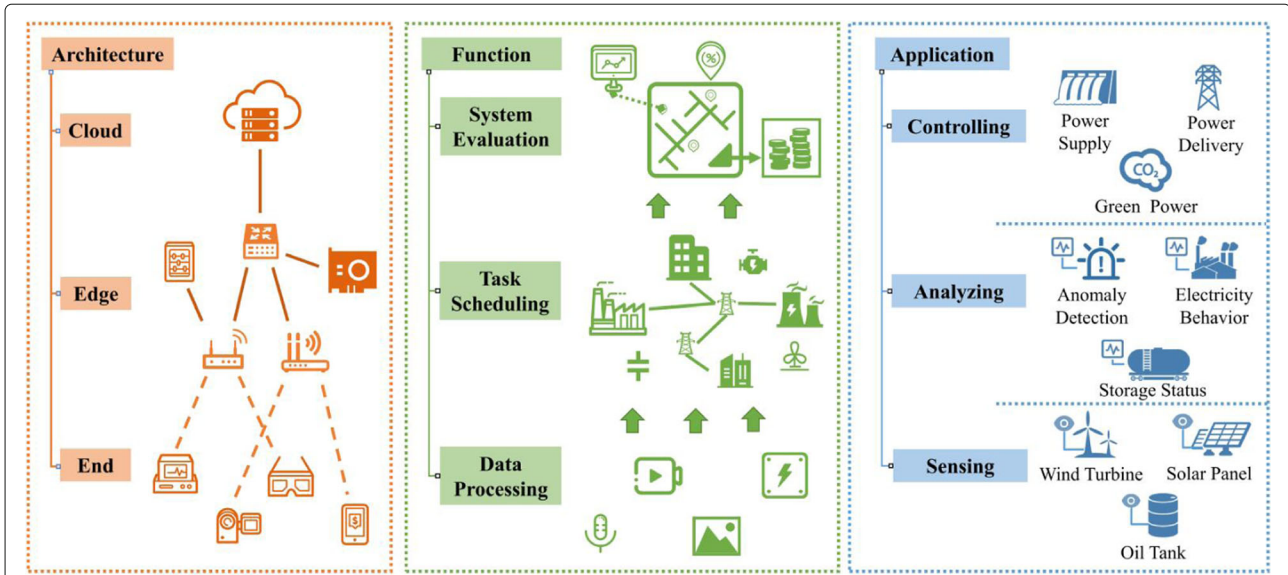


Fig. 1 The Framework of Power Flow Adjustment based on Edge Intelligence

and packaging. At the same time, the processing function also needs to have detailed configuration options, which should be compatible with multiple operation modes, so as to facilitate the agile deployment and application improvement for technical personnel. Finally, as the execution result of the adjustment, system evaluation can analyze the results of the decision-making process in time, then realize the dynamic and adaptive strategy adjustment, which can continuously optimize the application business, e.g., decision-making accuracy, system stability and task latency. The purpose of the power adjustment framework based on edge intelligence is to support power applications more efficiently, comprehensively and flexibly.

The primary application of the framework is the perception of a power network, i.e., to obtain the real-time state of everything in the power system, including the state of supply equipment, storage equipment and consumption equipment. The sensing information, as an important factor of decision-making processes, can effectively support the intelligence of the decision-making process. Further, the framework can analyze the status or behavior pattern of power equipment, e.g., a failure happened if some power unit parameters fluctuate considerably. Additionally, it can also analyze the adjustment strategy and stability capability of one grid region and then summarize the enabling state of non-renewable and renewable energy to identify efficient behavior strategies and even obtain model descriptions that are easy for professionals to understand. i.e., the learning-based strategy can be beneficial to human analysis. Power flow adjustment needs to dynamically adjust the control equipment in the power system, so how to determine the strategy of power supply

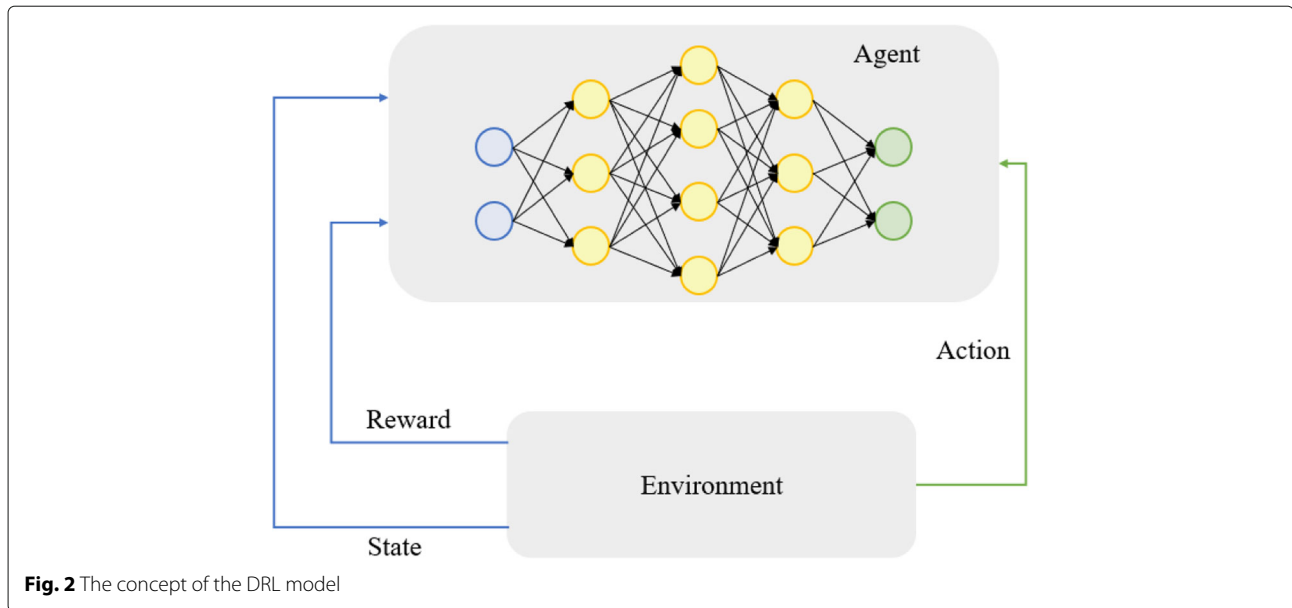
and power distribution becomes a crucial problem. If the calculation process does not converge, it is necessary to adjust the system parameters with actual operating steps. In addition, the control of carbon emissions has become an emerging problem in recent years. Therefore, the control of renewable resources should be taken into consideration in the process of power flow control, which is promising to improve the utilization of new energy and reduce the use of non-renewable resources.

Deep reinforcement learning

A tuple (S, A, T, r) is used to define a reinforcement learning task, as shown in Fig. 2. At each time-step t , the agents observe the state $s_t \in S$ of the environment and take actions $a_t \in A$ to transform themselves into a new state and receive a reward r . $T = p(s_{t+1}|s_t, a_t)$ is a mapping from state-action pairs (s_t, a_t) to a probability distribution of the next state s_{t+1} . The goal of an agent is to maximize its expected return during iterations, which is given by $R = \sum_{t=0}^{\infty} R_t = \sum_{t=0}^{\infty} \gamma^t r_t$, where $\gamma \in [0, 1]$ is the future discount factor. The state-action value function is defined as $Q^\pi(s, a) = \mathbb{E}[R_t | s_t = s, a_t = a, \pi]$, which means the expected discounted return based on the current state and action (s_t, a_t) . The following Bellman equation is used to express the optimal Q function Q^* under the suitable action:

$$Q^*(s, a) = \mathbb{E}_{s' \sim p(\cdot | s, a)} \left[r(s, a) + \gamma \max_{a' \in A} Q^*(s', a') \right].$$

In addition, each DRL agent has a target network. It has the same structure as the Q -network. Due to the unstable training process and poor performance with non-stationary targets, the target network's goal is to fix the Q



value targets. During every fixed number of iterations, the target network θ^- updates the parameters with that of the Q-network θ . The loss function is presented as follows:

$$L(\theta) = \mathbb{E}_{s,a,r,s'} \left\{ \underbrace{Q(s, a; \theta)}_{\text{prediction}} - \underbrace{\left[r + \gamma \max_{a' \in A} Q(s', a'; \theta^-) \right]}_{\text{target}} \right\}^2.$$

Traditional reinforcement learning algorithms are classified into value-based approach and policy-based approach. Both two categories of approaches have significant drawbacks.

Asynchronous advantage Actor-Critic(A3C) algorithm

Actor-Critic, as a mixed approach of value-based approach and policy gradient-based approach, usually performs better than each of them. There are two parts in Actor-Critic algorithm. One part is Actor, which selects an action using a neural network. The corresponding neural network approximating the policy is called a policy network. The other part is Critic judging whether good or bad the actions selected by Actor are, where the network estimating the value of actions is called the value network. We define θ_t as the weights of the policy network. Besides, the learning rate α and the policy π_θ are defined. Then, we use the parameter θ to update the policy network:

$$\theta_{t+1} \approx \theta_t + \alpha [\nabla \log \pi_\theta(a|s) Q_\pi(s, a)],$$

where $Q_\pi(s, a)$ is the total value by following the policy π after the selected action a in the current state s .

Since the training process involves multiple neural networks, the Actor-Critic algorithm has the disadvantage of

slow convergence. A3C is an Actor-Critic algorithm proposed to solve the non-convergence problem. In some classical reinforcement learning algorithms, such as deep Q-network (DQN), they use experience pool to improve convergence by reducing the correlation between data. Instead, in order to reduce the memory usage, A3C algorithm uses multiple workers to perform their own training on multiple environment instances asynchronously, and updates the global network asynchronously. In this way, A3C improves the speed of convergence. Compared with actor-critic algorithm, A3C algorithm mainly makes three optimizations: First, asynchronous training framework makes the network model interact with the environment better, which helps the model to converge quickly; Second, network structure optimization puts Actor and Critic together, so that the input state can output the state value and strategy. The third is critic assessment.

In the equation above, the Q value is not normalized. If Q is too large, the parameter θ changes too much. On the contrary, θ won't change a lot while the predicted value is small. Thus, A3C uses the difference value of the Q value and the value of the previous state, instead of the predicted Q value. The difference is called the advantage function, which represents the increase of value obtained with action a . If the value function at time-step t is $V(s_t) = \mathbb{E}[R_t|s_t = s]$, the advantage function can be expressed as:

$$\begin{aligned} A(s_t, a_t) &= Q(s_t, a_t) - V(s_t) = \mathbb{E}[R_t|s_t, a_t] - V(s_t) \\ &\approx r_t + \gamma V(s_{t+1}|s_t, a_t) - V(s_t) = \delta(s_t). \end{aligned}$$

The gradient of the actor is $\nabla \log \pi_\theta(a|s) \delta(s_t)$, then

$$\theta_{t+1} \approx \theta_t + \alpha [\nabla \log \pi_\theta(a|s) \delta(s_t)].$$

In addition, when updating the value network, the loss function is given as $\delta(s_t)^2$.

The knowledge and experience of power flow convergence *Knowledge of generator active power output on the convergence*

In the actual power grid, the unreasonable arrangement of generator sets may result in excessive active power transmission, exceeding the transmission capacity of the network [29]. In response to this situation, adding reactive power compensators or changing the transformer ratio on the line can improve the transmission capacity of the network to a certain extent. However, when faced with extremely unreasonable arrangements, these methods are difficult to achieve a satisfactory decision result. Therefore, to ensure that the active power transmitted by the power line does not exceed the upper limit of its transmission capacity, the output of each generator in the generator set needs to be adjusted [30].

Knowledge of power line transmission limit on the convergence

The capacity of the transmission line reaching the limit is the main factor for the static stability of the system. Under general conditions, the transmission power of the lines in the grid changes with the changes in the generator output and the active and reactive power of the load.

There are two situations when the active power of a transmission line reaches its transmission power limit: (i) With the continuous increase of the injected power, the active power of the transmission line reaching the limit will not continue to increase (or increase very little), and the increase in injected power is transmitted through other transmission channels; (ii) The active power of the transmission line increases with the increase of injected power, but the reactive power transmission of the line reaches the limit.

The line reaching its transmission power limit is a necessary condition for the power system to lose static stability. In this case, the system power flow has no solution, and the adjustment does not converge. By finding the line that reaches the transmission limit as knowledge and experience, it is possible to add the corresponding reactive power compensation and adjust the method of power injection. In this way, we can realize the purpose of non-convergence adjustment for power flow management in a given power network.

Experience in manual adjustment of non-convergent power flow

a) Adjustment of generator output

For small-scale distribution networks with power supply path compensation and direct power supply without boosting, adjusting generator output is a relatively

economical power flow adjustment method. At this time, changing the generator terminal voltage can achieve good results, and it is no need to add additional electrical equipment for adjustment. For power supply systems with long lines and multiple voltage levels, the adjustment of generators alone cannot meet the requirements of power flow convergence.

b) Adjustment of transformer ratio

Changing the transformer ratio can increase or decrease the voltage of the secondary winding. There are several taps for selection on the high-voltage side winding of the double-winding transformer and the high-voltage side and medium-voltage side winding of the three-winding transformer. The one corresponding to the rated voltage is called the main connector.

c) Reactive power compensation

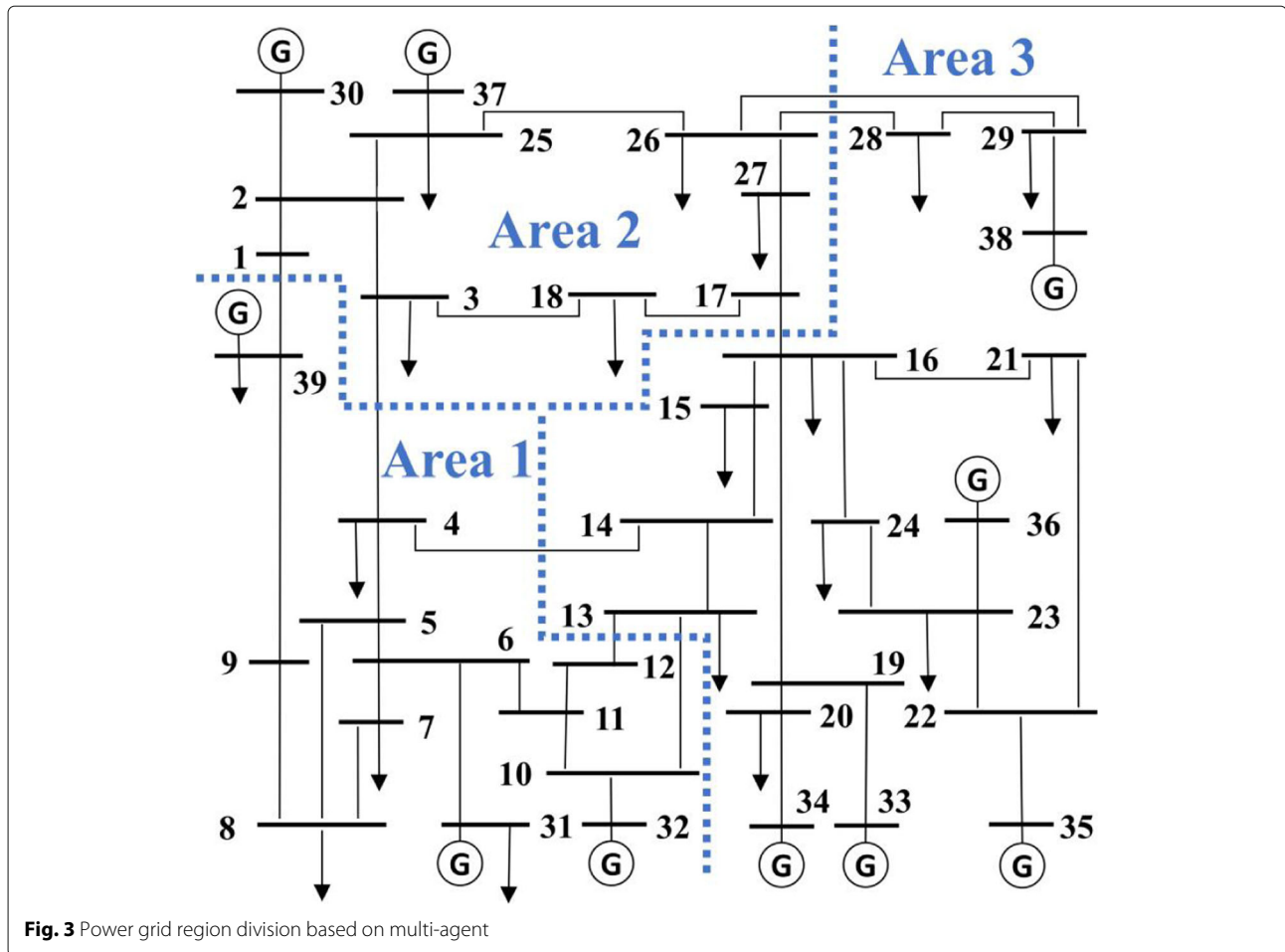
The generation of reactive power does not consume energy, and the transmission of reactive power along the power grid will cause active power loss and voltage loss. Suitable configuration of reactive power compensation and changing the reactive power flow distribution of the network can reduce the active power loss and voltage loss in the power system.

Automatic adjustment of power flow convergence based on DRL

Deep reinforcement learning has been used to adjust the non-convergence of power flow automatically. However, it is challenging to realize the real-time information sharing of each microgrid, and it is also difficult to dispatch and control each microgrid through a centralized organization. Therefore, we proposed to solve this problem by using multi-agent deep reinforcement learning. In some similar research work, in addition to the observation information of the environment, each decision unit also needs the observation information, such as the strategies and rewards of other agents. Considering the balance of active power and reactive power simultaneously, we propose the solution of automatic power flow non-convergence adjustment based on the knowledge and experience of power flow adjustment and multi-agent deep reinforcement learning.

Sub-grid partition

As shown in Fig. 3, according to the actual geographical location and electrical equipment distribution of the IEEE 39 bus system, the power grid is divided into three sub-grids. An agent is responsible for dispatching and controlling each sub-grid. Each agent can only observe the grid information of its sub-grid and maintain the electrical equipment of the sub-grid. In addition, the grid allows different agents to communicate with each other to achieve more efficient scheduling and control.



State design

For an agent, its state refers to the variables observed from the environment, which will affect the agent’s exploration efficiency. In the selection of state variables, we mainly consider the output of each generator, the voltage on each bus and the load of each transformer. Therefore, for the data of m samples, the total size of state space is: $m(g + p + q)$ where g is the total number of generators, p is the total number of buses, and Q is the total number of transformers. However, each agent can only observe the state information of its sub-grid, so the number of its observation space is: $m(g_i + p_i + q_i)$ where, g_i , p_i and q_i are the number of generators, buses and transformers in the sub-grid of each agent respectively. In addition, it can be seen from Table 1 that for different types of electrical equipment in the power system, the observation range settings of each observation equipment point are also different. This is mainly due to the combination of the characteristics of various electrical equipment.

Action design

Action is the actual strategy taken by the agent in the process of exploration. It is the key to the real-time flow convergence. We consider the regulation of both active power and reactive power, including the output multiple of each generator, the number of reactive power compensators on each heavy-duty bus and the ratio of each transformer. Therefore, for the data of m samples, the number of action Spaces constructed is: $m(g + p + q)$. Similarly, each agent can only control the electrical equipment in its sub-grid, and the number of its action space is: $m(g_i + p_i + q_i)$. Similar to the state design, in addition to the different number of electrical equipment in each sub-grid, for different types of electrical equipment, we combined their own characteristics and set different action ranges for each type of equipment to reduce the action space to achieve differentiation. As shown in Table 2, in order to reduce the difficulty of the agent’s decision-making, we discretize all the variables in the power grid, so that the whole action space is transformed into a discrete action

Table 1 The state space of the agent in each sub-grid

No.	Active power of generators	Buses voltage magnitude	Transformers load rate
agent 1	2 * (0 – 10000)	14 * (0 – 3)	4 * (0 – 500)
agent 2	2 * (0 – 10000)	10 * (0 – 3)	2 * (0 – 500)
agent 3	5 * (0 – 10000)	14 * (0 – 3)	5 * (0 – 500)

space, thus accelerating the whole process of multi-agent deep reinforcement learning.

In addition, we will also select the region with heavy line load in each flow adjustment process, which is helpful for agents to make better decisions to adjust the movement.

Reward design

To make full use of the relevant knowledge and experience of flow adjustment and improve the exploration efficiency of agents, we set up a variety of reward mechanisms. First of all, if the power flow adjustment of the sample converges, the highest positive return value r_1 can be obtained; if the power flow adjustment does not converge, the negative return value r_2 is finally added. Next, consider the upper limit of the generator output. According to whether the output active power of the generator is greater than its maximum active power limit, the reward value r_3 is set. Similarly, depending on whether the reactive power output of the generator is greater than its maximum reactive power limit, increase the reward value r_4 . Line load rate is also an important part of power flow adjustment. If the line load rate exceeds its maximum line load rate limit, the agent receives a negative reward of r_5 . In addition, we also consider the voltage level across the bus. If the voltage on the bus is within the specified maximum and minimum voltage range, the plus value r_6 is increased. Finally, the maximum load limit on the transformer constitutes the bonus value r_7 . The reward value R for each step of the agent is equal to the sum of the above 7 types of rewards:

$$R = r_1 + r_2 + r_3 + r_4 + r_5 + r_6 + r_7.$$

In particular, since power flow convergence is the common goal of all agents, the benefits brought by flow convergence can make every sub-grid gain benefits. Therefore, the whole process of flow adjustment convergence adjustment can be regarded as a cooperative game among multiple agents. Furthermore, we set the reward of each agent to be the same.

Multi-agent asynchronous advantage actor critic algorithm

We design multi-agent asynchronous Advantage Actor Critic (MAA3C) as our multi-agent deep reinforcement learning algorithm. Each agent maintains an A3C structure, which is used to select and evaluate strategies for the local states observed by the agent. Different agents maintain their own sub-grid and can communicate with each other to jointly pursue the power flow convergence goal of the whole grid. However, each A3C of the next layer has multiple workers composed of actor-critic to receive parameter updates of the global network, undergo reinforcement learning training, and update the global network asynchronously. Each actor-critic consists of 2 deep neural networks, namely the strategy network and the value network. Policy networks are used to explore policies, and value networks evaluate actions and provide critic values, which help actors learn the gradients of policies and tune the parameters of their networks to make updates work in a better direction.

Numerical results

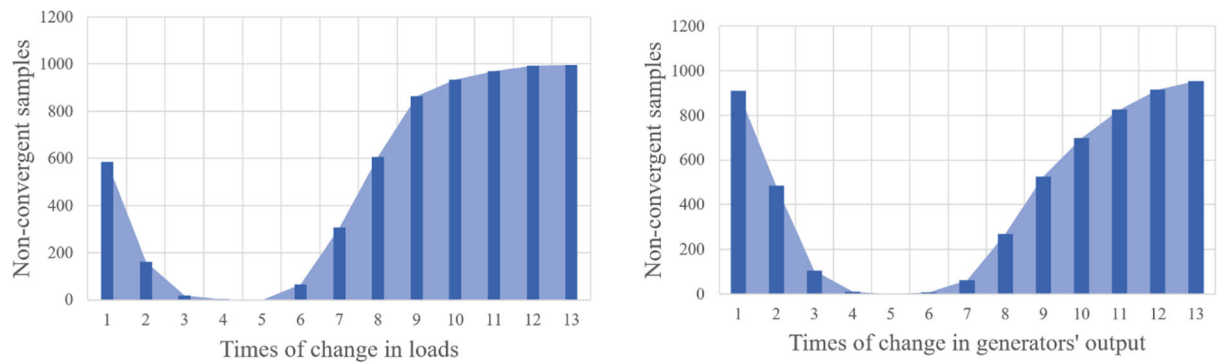
Simulation setting

In the experimental part, based on the Python 3.7 environment, we adopted Pandpower, an open-source third party simulator for power flow adjustment and analysis. By modifying some parts of the source code in the simulator, we obtained the intermediate data of power flow calculation as our knowledge experience of multi-agent deep reinforcement learning.

As for the method of power flow calculation, Newton-Raphson power flow algorithm with optimal multiplier is adopted. The correction vector obtained in each iteration of conventional Newton-Raphson algorithm is used as the search direction, and the objective function is regarded as one variable function of the step factor, with the scalar multiplier introduced to adjust the correction step size of the variable. In this way, better robustness than the Newton-Raphson algorithm can be obtained.

Table 2 The action space of the agent in each sub-grid

No.	multiples of generators' output	Number of compensators	Transformer ratio value
agent 1	2 * (0 – 5)	14 * (0 – 10)	4 * (0 – 5)
agent 2	2 * (0 – 5)	10 * (0 – 10)	2 * (0 – 5)
agent 3	5 * (0 – 5)	14 * (0 – 10)	5 * (0 – 5)



(a) The number of non-convergent samples varies with the loads' change (b) The number of non-convergent samples varies with the generators' change

Fig. 4 The impact of loads and generators' output on power flow adjustment

Data preprocessing

We select the IEEE 39 bus system in New England as the target of our experiment. The 345kV network consists of 10 generators, 12 double-winding transformers and 34 transmission lines, with a base power of 100MVA. According to the convergent data in the initial system, we randomly adjust the load and output of the generator in the range of 0-4 times. Then the Newton-Raphson method with the optimal multiplier is used to carry out the power flow calculation one by one. Consequently, we get 996 non-convergent samples, which are used as the data for adjustment. As shown in Figs.4(a) and (b), it can be found that within the random adjustment range of 0-1 times, with the decrease of load and generator active power, the number of non-convergent samples in power flow adjustment gradually increases. However, within the

range of 1-3 times of random adjustment, when the load and power generation output are farther away from the rated value, the number of samples that do not converge in power flow calculation also increases gradually. Especially after the proportion exceeds 200%, the number of non-convergent samples gradually occupies most of the samples.

Simulation results

To comprehensively present the advantages of our algorithm, we firstly compare the algorithm with centralized learning algorithm in one agent, such as A2C and A3C. Furthermore, the comparison with other multi-agent reinforcement learning algorithms are also considered. As can be seen from the total reward of the grid in Fig. 5, MAA3C algorithm can reach a convergence value faster than other

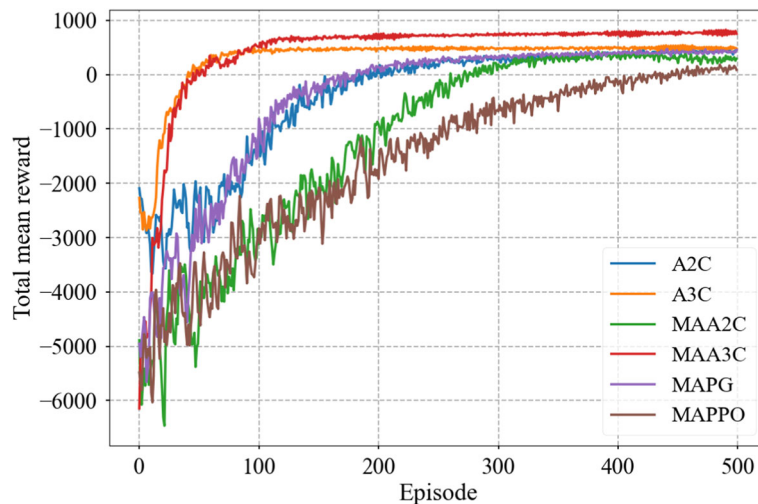


Fig. 5 The convergence performance under different algorithms

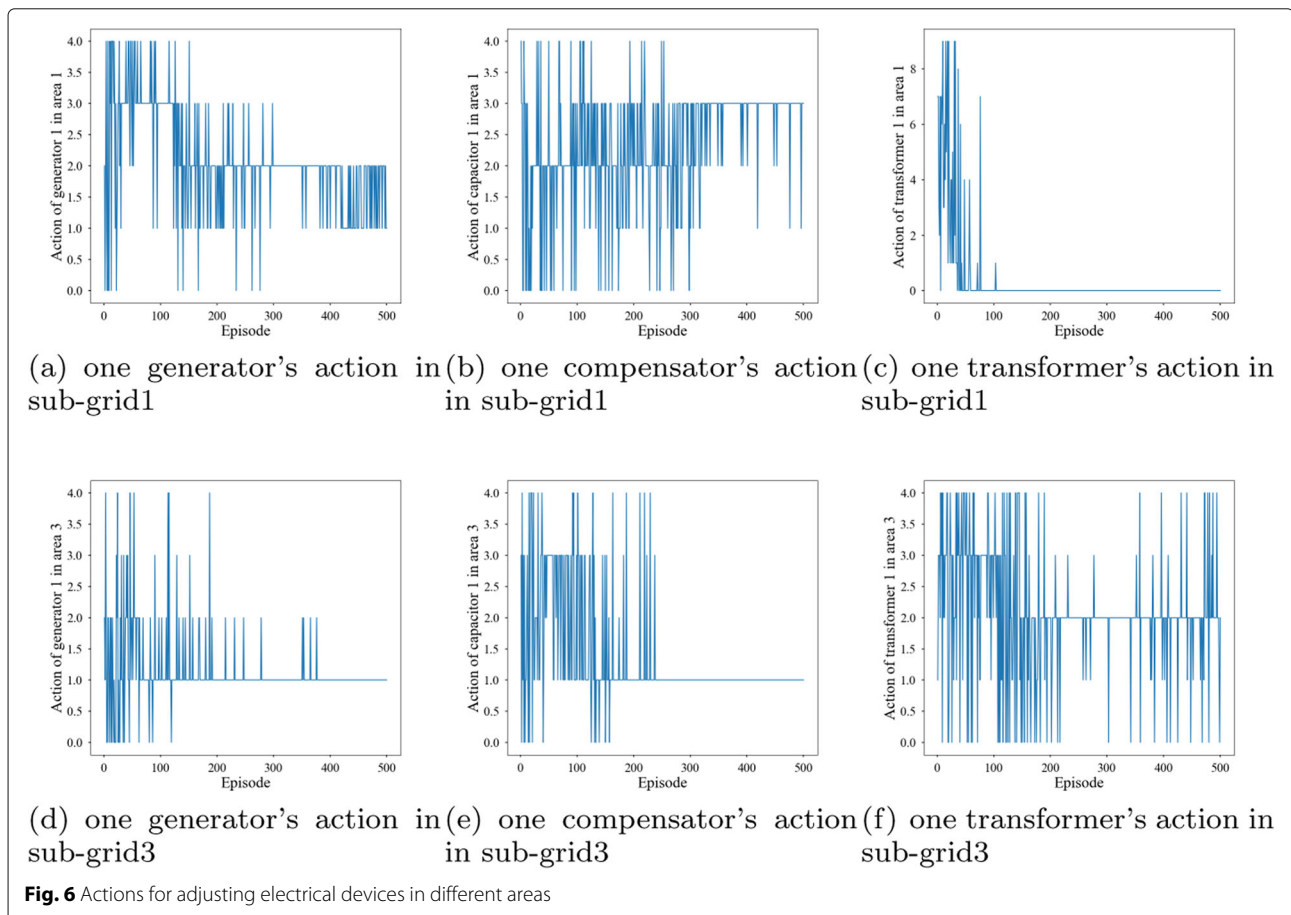
multi-agent reinforcement learning algorithms, and the stability in the process of convergence is much better than other algorithms. This relies heavily on the asynchronous updating method in the A3C architecture, which reduces the correlation between data, achieving faster convergence. In addition, our algorithm can finally obtain the maximum reward value among all the algorithms, which will also be reflected in the subsequent experiments. From the comparison of curve between MAA3C and A3C, under the condition of incomplete information, the convergence speed of multi-agent learning is almost the same as that of centralized learning. In face of such a large environment as the power grid, the multi-agent system may be more robust than centralized control.

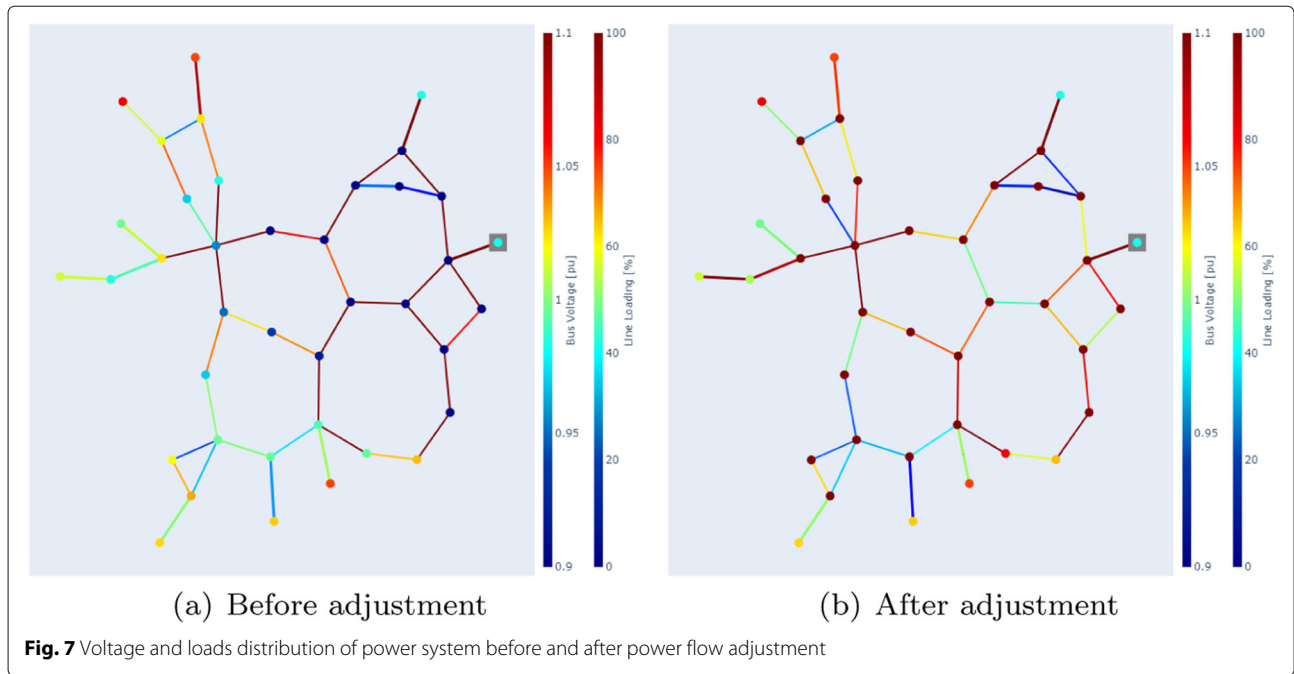
The actions of electrical devices controlled by different agents under different sub-grids reflect the actual changes of power grid decided by each agent under the MAA3C algorithm. As shown in Fig. 6, we randomly select generators, reactive power compensators and transformers from sub-grid 1 and sub-grid 3 to check their output multiples, increase number of compensators and percentage change of transformer ratio, respectively. It can be seen that after 300 iterations, each electrical device converges to a

specific value. It fluctuates a little due to the exploration of each agent.

We randomly select a sample that has completed the power flow adjustment, and plot the load rates of the bus and transmission lines in the grid system before and after the power flow adjustment. Figure 7 shows that the power grid before adjustment on the left has the situation that the load rate of local transmission lines is too high, and the bus voltage is too low, which is probably the main reason for the non-convergence of power flow adjustment. From the adjusted power grid on the right, it can be seen that the overload situation of local transmission lines has been well improved, and the bus voltage has also been reduced from too low to a relatively high and controllable level, so the power flow can be converged again.

To intuitively reflect the adjustment effect of MAA3C algorithm on power flow calculation of grid, we randomly selected 160 samples from 996 non-convergent samples as the test set, with the rest as the training set. Then we compare the successful adjustment numbers of non-convergent samples under different algorithms. To minimize the impact of accidental factors on the results, we calculate 10 times and average the results of 10 times. As

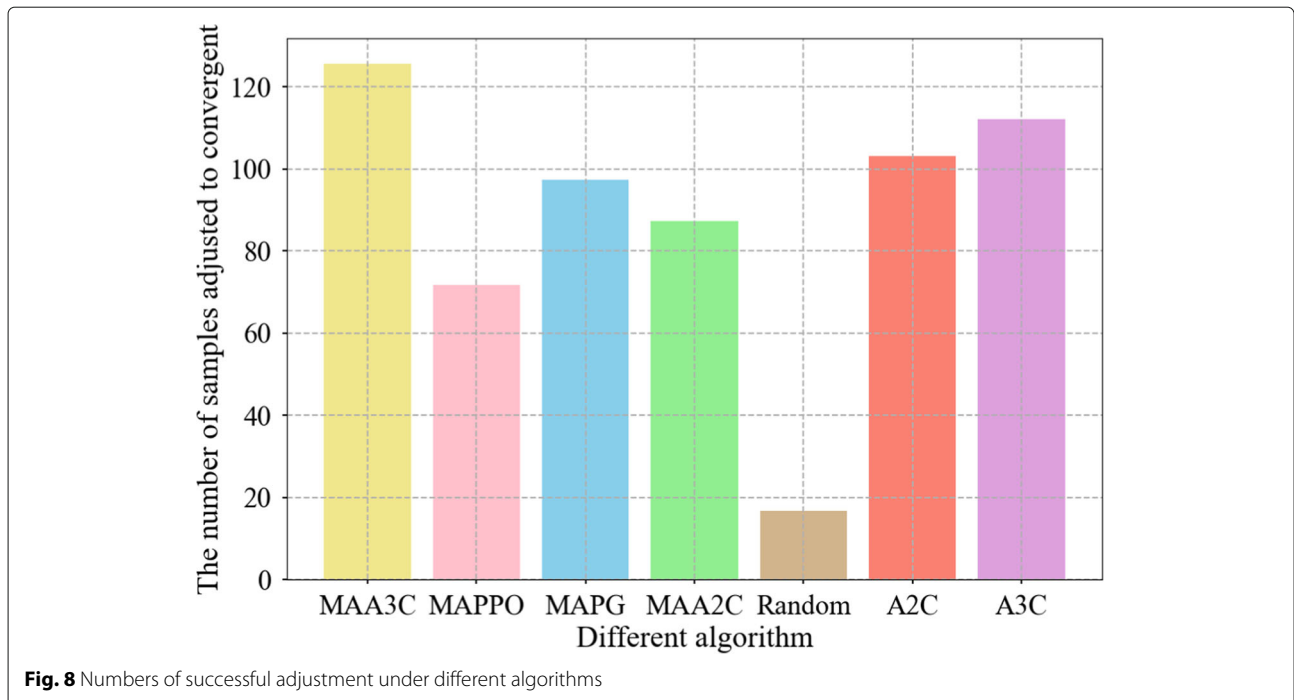




shown in Fig. 8, MAA3C algorithm has obvious advantages, whether compared with the single-agent deep reinforcement learning algorithm or with other multi-agent deep reinforcement learning algorithm. It can be observed that if the random strategy is used, the success rate of adjustment is less than 10%.

Conclusion

In this article, we proposed an edge computing-assisted comprehensive framework for smart grid management and control. Consequently, it assists microgrids in realizing real-time demand response and local autonomy in data sensing, processing and controlling. Primarily, we



proposed a power flow adjustment algorithm based on multi-agent deep reinforcement learning considering the grid knowledge and requirement in microgrids, which improves the efficiency and flexibility compared with the traditional methods. Finally, we adopt the IEEE 39 bus system with the Pandapower simulator to verify the effectiveness of our proposed algorithm under various grid conditions.

In future work, we will further discuss the following two points. Deployment and application of computing power near perception and control devices are emerging trends in smart grids. Edge-cloud collaboration can realize intelligent collaboration and efficient decision-making of IoT devices, which will gradually be widely adopted. How to realize the dynamic adaptation and flexible scheduling of the system is an open question. On the other hand, there will be more supply units, storage units, and load units in the power grid. How to model and analyze the characteristics of these new units becomes another problem worthy of further study.

Acknowledgements

The authors would like to thank to anonymous reviewers for their valuable comments on the manuscript.

Authors' contributions

All authors took part in the discussion and analysis of the work described in the paper. Tianjiao Pu initiated the research and led the entire work. Xinying Wang and Ji Qiao contributed to the methodology and design of this paper, Yifan Cao and Zhicheng Liu carried out the experimental work and drafted the manuscript, Chao Qiu took part in writing of the manuscript, and Shuhua Zhang contributed data analysis. All authors read and approved the final manuscript.

Authors' information

Tianjiao Pu, professorate senior engineer. He is the director of the Artificial Intelligence Application Research Department of China Electric Power Research Institute (CEPRI), a core member of the "Science and Technology Tackling Team of Energy-saving Economic Dispatching" in the State Grid Corporation of China, a Fellow of IET, a senior member of IEEE, a senior member of the Chinese Society for Electrical Engineering, a CIGRE member, Secretary General of the AI Committee of CSEE, and a deputy editor-in-chief of IET Smart Grid. He graduated from the Department of Electric Power Systems and Automation of Tianjin University in 1997 and has worked ever since at the China Electric Power Research Institute. He has been engaged in the research and management of power dispatching automation, smart grid simulation, active distribution network, artificial intelligence and other fields.

Xinying Wang, Senior Engineer. He is the Deputy Director of the Artificial Intelligence Application Research Section of the China Electric Power Research Institute (CEPRI), the member of CSEE. He received his Ph.D. degree from Dalian University of Technology in 2015 and has worked ever since at the CEPRI. His main fields of interest are electric artificial intelligence.

Yifan Cao is received the B.S. degree from Tianjin University, Tianjin, China, in 2020. He is currently pursuing the M.S. degree in the College of Intelligence and Computing, Tianjin University. His research interests include energy trading, blockchain and deep reinforcement learning.

Zhicheng Liu received the B.S. degree in information security from Guizhou University, China, in 2015, and the M.S. degree in computer science and technology from Inner Mongolia University, China, in 2019. Currently, he is working toward the Ph.D. degree in computer science with the college of intelligence and computing, Tianjin University, China. His research interests include edge computing, multi-agent learning, and game theory.

Chao Qiu is currently a Lecturer in the School of Computer Science and Technology, College of Intelligence and Computing, Tianjin University. She received the B.S. degree from China Agricultural University in 2013 in

Communication Engineering and the Ph.D. degree from Beijing University of Posts and Telecommunications in 2019 in Information and Communication Engineering. From September 2017 to September 2018, she visited Carleton University, Ottawa, ON, Canada, as a Visiting Scholar. Her current research interests include machine learning, computing power networking and blockchain.

Ji Qiao received his Ph.D. degree in electrical engineering from Tsinghua University, Beijing, China, in 2018. He is currently working at the China Electric Power Research Institute, Beijing, China. His research interests include big data and AI applications in power system analysis and operations.

Shuhua Zhang received B.S. and the M.S. degrees in measurement and control technology and instrument from the Tianjin University, Tianjin, in 2002 and 2005. He is currently pursuing the Ph.D. degree with North China Electric Power University, China. He is also a professor of engineering with China Electric Power Research Institute Company Ltd. His research interests include integrated circuit, artificial intelligence, power edge computing.

Funding

This work is supported by the Science and Technology Project of SGCC (5400-201955369A-0-0-00): Power and Load Forecasting and Optimal Dispatch of Active Distribution Network Based on Machine Learning.

Availability of data and materials

The implementation of the IEEE 39 bus system used in the experiment is included in the simulator Pandapower, we have quoted it in the main content and can see more details in <http://pandapower.org>.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

¹China Electric Power Research Institute, Haidian District, 100192 Beijing, China. ²College of Intelligence and Computing, Tianjin University, Jinnan District, 300350 Tianjin, China.

Received: 22 February 2021 Accepted: 23 July 2021

Published online: 09 September 2021

References

1. Wang X, Han Y, Leung V, Niyato D, Chen X (2020) Convergence of edge computing and deep learning: A comprehensive survey. *IEEE Commun Surv Tutor* 22(99):869–904
2. Thurner L, Scheidler A, Schäfer F, Menke J, Dollichon J, Meier F, Meinecke S, Braun M (2018) pandapower—an open-source python tool for convenient modeling, analysis, and optimization of electric power systems. *IEEE Trans Power Syst* 33(6):6510–6521. <https://doi.org/10.1109/TPWRS.2018.2829021>
3. Zhang Z, Zhang D, Qiu RC (2020) Deep reinforcement learning for power system applications: An overview. *CSEE J Power Energy Syst* 6(1):213–225
4. Li X, Wang X, Li K, Han Z, Leung VCM (2017) Collaborative multi-tier caching in heterogeneous networks: Modeling, analysis, and design. *IEEE Trans Wirel Commun* 16(10):6926–6939. <https://doi.org/10.1109/TWC.2017.2734646>
5. Tran TX, Hajisami A, Pandey P, Pompili D (2017) Collaborative mobile edge computing in 5g networks: New paradigms, scenarios, and challenges. *IEEE Commun Mag* 55(4):54–61. <https://doi.org/10.1109/MCOM.2017.1600863>
6. Wang X, Han Y, Wang C, Zhao Q, Chen M (2019) In-edge ai: Intelligentizing mobile edge computing, caching and communication by federated learning. *IEEE Network* 33(5):156–165
7. Trajano AFR, de Sousa AAM, Rodrigues EB, de Souza JN, de Castro Callado A, Coutinho EF (2019) Leveraging mobile edge computing on smart grids using LTE cellular networks. In: 2019 IEEE Symposium on Computers and Communications (ISCC). pp 1–7. <https://doi.org/10.1109/ISCC47284.2019.8969784>
8. Barik RK, Gudey SK, Reddy GG, Pant M, Dubey H, Mankodiya K, Kumar V (2017) FogGrid: Leveraging fog computing for enhanced smart grid network. In: 2017 14th IEEE India Council International Conference (INDICON). pp 1–6. <https://doi.org/10.1109/INDICON.2017.8487997>

9. Huang Y, Lu Y, Wang F, Fan X, Liu J, Leung VCM (2018) An edge computing framework for real-time monitoring in smart grid. In: 2018 IEEE International Conference on Industrial Internet (ICII). pp 99–108. <https://doi.org/10.1109/ICII.2018.00019>
10. El-Awadi R, Fernández-Vilas A, Díaz Redondo RP (2019) Fog computing solution for distributed anomaly detection in smart grids. In: 2019 International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob). pp 348–353. <https://doi.org/10.1109/WiMOB.2019.8923222>
11. Chen S, Wen H, Wu J, Lei W, Hou W, Liu W, Xu A, Jiang Y (2019) Internet of things based smart grids supported by intelligent edge computing. IEEE Access 7:74089–74102. <https://doi.org/10.1109/ACCESS.2019.2920488>
12. Albataineh H, Nijim M, Bollampall D (2020) The design of a novel smart home control system using smart grid based on edge and cloud computing. In: 2020 IEEE 8th International Conference on Smart Energy Grid Engineering (SEGE). pp 88–91. <https://doi.org/10.1109/SEGE49949.2020.9181961>
13. Han Y, Guo D, Cai W, Wang X, Leung V (2020) Virtual machine placement optimization in mobile cloud gaming through QoE-oriented resource competition. IEEE Trans Cloud Comput. Early Access
14. Liu X, Yu J, Feng Z, Gao Y (2020) Multi-agent reinforcement learning for resource allocation in IoT networks with edge computing. China Commun 17(9):220–236. <https://doi.org/10.23919/JCC.2020.09.017>
15. Wang X, Wang C, Li X, Leung VCM, Taleb T (2020) Federated deep reinforcement learning for internet of things with decentralized cooperative edge caching. IEEE Internet Things J 7(10):9441–9455. <https://doi.org/10.1109/JIOT.2020.2986803>
16. Yang T, Zhao L, Li W, Zomaya AY (2021) Dynamic energy dispatch strategy for integrated energy system based on improved deep reinforcement learning. Energy 235:121377
17. Fang X, Wang J, Yin C, Han Y, Zhao Q (2020) Multiagent reinforcement learning with learning automata for microgrid energy management and decision optimization. In: 2020 Chinese Control And Decision Conference (CCDC). pp 779–784. <https://doi.org/10.1109/CCDC49329.2020.9164742>
18. Ji Y, Wang J, Xu J, Fang X, Zhang H (2019) Real-time energy management of a microgrid using deep reinforcement learning. Energies 12(12):2291
19. Etemad S, Mozayani N (2015) Using reinforcement learning to make smart energy storage sources in microgrid. In: 2015 30th International Power System Conference (PSC). pp 345–350. <https://doi.org/10.1109/IPSC.2015.7827771>
20. Liu W, Zhuang P, Liang H, Peng J, Huang Z (2018) Distributed economic dispatch in microgrids based on cooperative reinforcement learning. IEEE Trans Neural Netw Learn Syst 29(6):2192–2203. <https://doi.org/10.1109/TNNLS.2018.2801880>
21. Brida M, Frederik R, Fred S, Geert D (2017) Battery energy management in a microgrid using batch reinforcement learning. Energies 10(11):1846
22. Dabbaghjamesh M, Zhang J (2020) Deep learning-based real-time switching of reconfigurable microgrids. In: 2020 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT). pp 1–5. <https://doi.org/10.1109/ISGT45199.2020.9087729>
23. Ma D, Shen C, Chen Y, Li D (2020) Feasibility study on the convergence criterion of power flow calculation based on deep learning. South Power Syst Technol (Chinese) 014(002):46–54
24. Wang T, Tang Y, Guo Q, Huang Y, Chen X, Huang H (2020) Automatic adjustment method of power flow calculation convergence for large-scale power grid based on knowledge experience and deep reinforcement learning. Proc Chin Soc Electr Eng (Chinese) 40(8):2396–2405
25. Zhu X, Liu C, Su C, Liu J Learning-based probabilistic power flow calculation considering the correlation among multiple wind farms. IEEE Access 8(99):136782–136793. <https://doi.org/10.1109/ACCESS.2020.3011511>
26. Yang Y, Yang Z, Yu J, Zhang B, Zhang Y, Yu H (2020) Fast calculation of probabilistic power flow: A model-based deep learning approach. IEEE Trans Smart Grid 11(3):2235–2244. <https://doi.org/10.1109/TSG.2019.2950115>
27. Su T, Liu Y, Shen X, Liu T, Qiu G, Liu J (2020) Deep learning-driven evolutionary algorithm for preventive control of power system transient stability. Proc Chin Soc Electr Eng (Chinese) 40(12):103–114
28. Huang Q, Huang R, Hao W, Tan J, Fan R, Huang Z (2020) Adaptive power system emergency control using deep reinforcement learning. IEEE Trans Smart Grid 11(2):1171–1182
29. Sinan M, Sivakumar WM, Anguraja R (2021) Power system voltage stability analysis with renewable power integration. Int J Innov Technol Exploring Eng 10(6):114–117
30. Leveringhaus T, Hofmann L (2011) Optimized voltage and reactive power adjustment in power grids using the least-squares-method: Optimization of highly utilized power grids with stochastic renewable energy-sources. In: 2011 International Conference on Power and Energy Systems. pp 1–6

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)