Research on Collaborative Control Strategy of Virtual Power Plant Based on Deep Reinforcement Learning Framework

Longfei Yue[1,2,*], Xianxia Liang[1], Litao Sun[1], Yupeng Li[1] and Longxue Cheng[1]

[1]*Department of Electrical Engineering, Hebei Institute of Mechanical and Electrical Technology, Xingtai 054000, China*
[2]*Xingtai Technology Innovation Centre for Intelligent Production Line and Equipment, Xingtai 054000, China*
E-mail: yuelongfei2025@163.com
[*]*Corresponding Author*

Received 09 April 2025; Accepted 26 April 2025

Abstract

To address the limitations of the traditional Deep Q-Network (DQN) in VPPs collaborative dispatch, such as large state space, low computational efficiency, and poor generalization, an improved DQN (IDQN) is proposed. Firstly, consider various factors of VPPs and optimize its dispatch process to enhance its prediction accuracy. Secondly, a new reward mechanism is designed to guide agents to consider both long-term stability and rational allocation of resources to improve energy utilization. Then, IDQN is employed to optimize the VPPs cooperative scheduling, which can improve the scheduling efficiency and control its cost. Finally, based on the improved DQN, a multi-objective decision-making framework is proposed for collaborative

Distributed Generation & Alternative Energy Journal, Vol. 40_3, 533–558.
doi: 10.13052/dgaej2156-3306.4034

optimization scheduling of VPP to improve system stability. Compared with the classical Q-learning algorithm, the introduced IDQN has lower rejection rate, lower cost and higher scheduling efficiency when used in VPP collaborative scheduling.

Keywords: Deep reinforcement learning, DQN method, multi-objective decision-making framework, virtual power plant, collaborative scheduling optimization.

1 Introduction

The traditional power system faces unprecedented challenges, particularly in maintaining the stability and security of the power grid [1]. Under this background, VPPs, as a new dispatching and management mode, has been widely studied and applied. Virtual power plants integrate and coordinate various energy sources, such as wind and solar power, to enable intelligent management of energy production and consumption. Essentially, they utilize information and communication technologies to collaboratively schedule multiple small energy systems, ensuring optimal resource allocation and efficient utilization within a large-scale power grid [2]. However, the issue of coordinating scheduling and control in virtual power plants continues to present numerous challenges, including the variability and spatial distribution of resources across these plants, which significantly increases the complexity of the control and scheduling processes [3]. There are significant differences in performance, response time and control strategy among different resources such as battery energy storage, distributed power generation and demand response, which makes it difficult for traditional centralized control methods to adapt to their dynamic and complex operating environment. In addition, the uncontrollable load fluctuation and the uncertainty of renewable energy adds complexity to dispatching [4].

In recent years, increasing research has focused on intelligent algorithm-based solutions for collaborative control in VPPs. As an adaptive decision-making method based on trial-and-error learning, Deep Reinforcement Learning (DRL) leverages the capabilities of DL for intelligent optimization [5]. Especially in collaborative control of virtual power plants, DRL can continuously optimize the strategy through interaction with the environment, so as to achieve efficient scheduling of energy resources. The collaborative problem of virtual power plant can be modeled and solved from a brand-new perspective [6]. In traditional control methods, scheduling strategies often

rely on expert experience or rule making, which is difficult to cope with environmental changes and system uncertainty. DRL can learn independently and adjust decision continuously to adapt to the changing power demand [7]. Although deep reinforcement learning has great potential in collaborative control of virtual power plants, it still faces some technical challenges in practical application. For example, how to conduct efficient training and rapid decision-making in actual power systems is still a difficult problem. In addition, the algorithm design of deep reinforcement learning needs to be considered [8]. How to dispatching and design an appropriate reward mechanism under the premise of safety, so as to coordinate and ensure that each resource can effectively cooperate under the global optimal goal is an important research topic [9].

To address the challenge of cooperative optimal scheduling, the structure of virtual power plants (VPPs) is analyzed, considering factors such as grid load, energy production, and energy storage. The complex and dynamic states involved in the scheduling process are optimized and modeled to enhance the state space structure, aligning it with practical scheduling requirements and improving prediction accuracy. A novel reward mechanism is introduced to handle grid load complexities, guiding agents to balance short-term objectives with long-term stability and rational resource allocation. The IDQN approach integrates wind, photovoltaic, energy storage, and other energy sources to improve new energy scheduling efficiency while controlling virtual power plant costs, thereby maximizing resource utilization. Finally, a multi-objective decision-making framework is proposed, incorporating factors such as cost, curtailment rates of wind and solar, and new energy utilization rates. This framework facilitates the collaborative optimal scheduling of multiple VPPs, enabling the model to balance diverse demands, enhance overall scheduling efficiency, and improve system stability.

The main contributions of this study are as follows:

1. By considering factors like grid load, energy production and energy storage state, the state space of DQN is improved, and the accuracy of collaborative optimization scheduling model is improved.
2. To address the fluctuation complexity of grid load, a new incentive mechanism is introduced to guide agents in balancing short-term goals with efficient resource allocation, thereby improving energy utilization.
3. An enhanced DQN is proposed to increase scheduling efficiency of renewable energy while controlling virtual power plant costs, maximizing resource utilization.

4. A multi-objective decision-making framework is proposed, incorporating VPP costs, renewable energy utilization, and other factors, to enable collaborative optimal scheduling of multiple VPPs, thereby enhancing overall scheduling efficiency and system stability.

2 Related Work

As a new energy management and scheduling mode, VPPs involves the coordination and optimal scheduling of various distributed energy resources [10]. Designing effective control strategies has become a key focus of scholarly research. In the traditional research of virtual power plant control, Zhu et al. [11] proposed a two-level real-time economic model, where system operators schedule VPPs based on a price incentive mechanism, while VPPs respond with optimal control strategies. In addition, mapping method and bilevel optimization method are proposed as the solution methods of the model. Aiming at the uncoordinated and uncertain operation, Arman et al. [12] proposed a novel virtual power plant model incorporating the uncertainty of electric vehicles (EVs), which integrates both the power market and existing energy systems. The model incorporates all uncertain parameters related to EVs within a comprehensive stochastic optimization framework. This approach is applicable to VPPs that include wind power generation and parking lots hosting electric vehicles, facilitating the optimization of asset scheduling prior to participation in the electricity market. Numerical simulations validate the theoretical approach and demonstrate the practicality of establishing a self-scheduling model. Li et al. [13] proposed a two-stage scheduling framework for virtual power plants, namely, day-ahead planning and day-to-day adjustment, to solve the uncertainty of wind power plants and photovoltaic. To address the operational challenge, Tomasz et al. [14] analyzed data, developed a PV model, optimized the coordinated control system, and used the Prophet model for short-term power generation forecasting to minimize reliance on external power supply.

In view of the problems that running virtual power plants only based on economic feasibility, Park et al. [15] considered the cooperation between VPP and distribution system operators, put forward an operation method and an optimized operation model considering DSC uncertainty, and established an operation strategy and an operation plan considering punishment provided by DSO. To improve power market efficiency, Wang et al. [16] proposed an operational method that accounts for the benefits of all participants and

developed a corresponding model. Ghulam et al. [17] put forward a model that integrates energy hubs and considers market operation balance, leading to a joint optimization strategy for the day-ahead market. By relaxing conditions and binary expansion to bilinear terms, MPEC is approximated as a mixed integer linear programming, and stochastic schemes are used to model the uncertainty of renewable energy, load and price, and its profits are improved through strategic cooperative electric energy trading.

With the development of artificial intelligence technology, Deep Reinforcement Learning (DRL) has become an important research direction of collaborative control of virtual power plants because of its self-learning and optimization in dynamic environment. Many researchers began to explore the scheduling method based on DRL. To address the challenges of DRL methods being disrupted and impacting system performance when data is private in VPPs, Feng et al. [18] proposed an enhanced DRL strategy that accounts for interference, and random control was utilized to improve training speed, and a gradient filter was introduced to maintain the balance of internal VPP power. For the purpose to solve the collaboration problem, Li et al. [19] proposed a collaborative framework, which designed several DRL agents, and conducted confrontational training to solve the internal pricing problem, improving the profits of VPP operators and MGs. Xue et al. [20] designed TD3 to achieve economic scheduling of VPPs. By incorporating network security constraints and using a differential evolution algorithm, they reduced operational costs. In an effort to address the challenges of distributed power sources and achieve reliable economic scheduling in VPPs, Lin et al. [21] proposed an optimized DRL approach that analyzes the random characteristics of distributed generation, reduces computational complexity, and, through continuous state space, learns both the characteristics of distributed power sources and industrial users' needs online to minimize VPP costs. Liu et al. [22] proposed an enhanced DRL method to coordinate VPPs, using a double-delay depth deterministic strategy gradient approach that allows all participating entities to share equal decision-making rights simultaneously. In addition, in order to further tap the operating potential of distributed energy, an external market based on price quota curve is proposed, which makes VPPs with market power become a price maker, which can simultaneously increase operating costs and maintain network operation security. To address the challenge of managing large-scale distributed energy in virtual power plants due to inaccurate parameters, Yi et al. [23] presented an enhanced DRL method with adjustment service and decomposition, using an off-line

simulator to learn the dynamic characteristics of the DER aggregator and train the control strategy with SAC, and prior knowledge is used to enable more accurate and cost-effective management of the DER aggregator in tracking adjustment requests. In order to realize more effective energy management of virtual power plant, Seyyed et al. [24] proposed a stochastic VPPs power dispatching formula. By introducing DRL program, operators can manage VPP online in the dispatching program, which effectively reduced the VPP profit.

3 Multi-VPPs Collaborative Optimization Model

3.1 Dispatching Architecture

Multi-VPPs usually requires each VPP to provide detailed physical models of internal adjustable loads and adjustable units, which is not conducive to the demands of participants for privacy data protection. In view of the privacy data protection problems faced by multi-VPPs participating in global optimal dispatching, as well as the complex, time-series coupling and non-linear characteristics of adjustable resources of virtual power plants, this paper proposes a global optimal dispatching framework of multi-VPPs based on deep reinforcement learning, taking into account the power constraints of power system frequency modulation demand, aiming at minimizing the reserve cost of power frequency modulation and protecting the privacy data of participants.

In the mode of deep reinforcement learning, the upper dispatching center first sends the public key to each VPP. Then, each VPP obtains the initial parameters of the training model through its own internal data, and uploads the DQN model gradient and the local model training accuracy. Moreover, as a semi-trusted third party, the power system dispatching center calculates the global loss function gradient of each VPP. At the same time, each VPP updates the model according to the global loss function gradient and global update interval provided by the upper dispatching center, and dynamically adjusts the adjustable unit output and adjustable load management strategy in the virtual power plant to respond to the frequency modulation reserve demand of the power system. Finally, each VPP analyzes and calculates the frequency modulation standby cost corresponding to different dispatching instructions, and transmits the cost data to the dispatching center, and the upper dispatching center participates in the frequency modulation service by checking each virtual power plant.

3.2 Dispatching Cost Model

Optimal dispatching of frequency modulation backup based on multi-VPPs is proposed when a single traditional power supply cannot meet the frequency modulation demand of power system. It is defined as the mapping of any dispatching instruction with the total response cost of traditional units and virtual power plants at any time. The mapping rules are based on the changes of active power of traditional units and virtual power plants. The global optimization dispatching model takes the minimum frequency modulation standby dispatching cost as the objective function. When adjustable resources participate in FM service, traditional power generation takes the unit as an independent unit to calculate the cost, and VPP participates in FM auxiliary service as an independent auxiliary service provider. Therefore, multi-VPPs submits the adjustment cost as an independent whole to the superior dispatching center, and the formula is as follows:

$$\{t^{\mathrm{S}}, T^{\mathrm{PD}}, \Delta p^{\mathrm{Tot}}(\tau)\} \rightarrow \{C^{\mathrm{Tot}}[t^{\mathrm{S}}, T^{\mathrm{PD}}, \Delta p^{\mathrm{Tot}}(\tau)]\} \tag{1}$$

Where, t^{S} is the dispatching instruction issuing time. T^{PD} is dispatching period. $\Delta p^{\mathrm{Tot}}(\tau)$ is the power of the given dispatching instruction in the τ period of the power system. C^{Tot} schedules the total cost to meet the demand of FM standby. Multi-VPPs dispatching cost model includes the stable traditional unit and the flexible internal dispatching cost, which are expressed as follows:

$$\begin{cases} \min C^{\mathrm{Tot}} = C^{\mathrm{TPP}} + \sum\limits_{M}^{m} C_m^{\mathrm{VPP}} \\ C^{\mathrm{TPP}} = \sum\limits_{E}^{e} \sum\limits_{T^{\mathrm{PD}}}^{\tau} [\Delta p_e^{\mathrm{TPP}}(\tau) C_e^{\mathrm{Gen}}] \\ C_m^{\mathrm{VPP}} = C_m^{\mathrm{DG}} + C_m^{\mathrm{AL}} + \sum\limits_{T^{\mathrm{PD}}}^{t} U_{m,\tau}^{\mathrm{VPP}} \end{cases} \tag{2}$$

Where, C^{TPP} is the traditional unit dispatching cost. C_m^{VPP} is *m*-th dispatching cost. $\Delta p_e^{\mathrm{TPP}}(\tau)$ is the effective power of traditional unit e in τ period. C_e^{Gen} is the unit power dispatching cost of traditional unit e. C_m^{DG} is the distributed cost. C_m^{AL} indicates the compensation cost of adjustable load participating in the demand response of frequency modulation reserve in power system. M and E represent the total number of VPP and traditional

units respectively. $U_{m,\tau}^{\mathrm{VPP}}$ is the punishment when the internal constraints of VPP cannot be met.

For VPP operation, whether effective dispatching power of traditional units and VPP meets the power demand of power system for frequency modulation reserve $\Delta p^{\mathrm{Req}}(\tau)$ is an important constraint condition of dispatching system, which is expressed by the formula as follows:

$$\text{s.t.} \sum_{E}^{e} \Delta p_e^{\mathrm{TPP}}(\tau) + \sum_{M}^{m} \Delta p_m^{\mathrm{VPP}}(\tau) \sim \Delta p^{\mathrm{Req}}(\tau) \tag{3}$$

Where, $\Delta p_m^{\mathrm{VPP}}(\tau)$ represents the effective regulated power of virtual power plant M. $\Delta Req(\tau)$ is the reserve power demand for frequency modulation in power system.

3.3 VPP Adjustable Resource Model

VPPs can control energy output and controllable load, and its internal flexibility is more conducive to the development. Under the constraint of power system frequency modulation, the global optimization goal of dispatching based on virtual power plant is to optimize industrial load, electric vehicle (EV), and photovoltaic. VPPs involved in this paper include adjustable units represented by photovoltaic and energy storage units, as well as adjustable loads represented by industrial users and electric vehicle aggregators. In order to simplify the model, this paper assumes that there is no energy coupling relationship among virtual power plants, and selects typical adjustable units and adjustable loads to model the virtual power plants.

In order to meet the demand of frequency modulation reserve in power system, multi-VPPs can moderately reduce and increase the EV aggregator and industrial load for adjustable load. However, this will change the use habits of industrial users and electric vehicle users, so multi-VPPs operators will provide corresponding compensation to various users based on power. Industrial users need to consider the proportion of interruptible load when participating in frequency modulation assistance, and it should not exceed the maximum interruptible load. As a FM resource, electric vehicles need to meet the basic service needs of EV aggregators. For the adjustable load compensation cost C_m^{AL}, the formula is as follows:

$$C_m^{\mathrm{AL}} = C_m^{\mathrm{IU}} + C_m^{\mathrm{EV}} \tag{4}$$

Where, C_m^{IU} is the regulation cost given by multi-VPPs to industrial users. C_m^{EV} is to control the cost for EV aggregators.

The adjustment cost of industrial users refers to the economic expenditure generated by industrial users' participation in demand response and dispatching optimization, including the extra operating cost of power demand regulation and the equipment maintenance cost needed to achieve dispatching objectives, which is expressed as follows:

$$\begin{cases} C_m^{\mathrm{IU}} = \sum\limits_{\tau} \alpha_m^{\mathrm{IU}}(\tau) p_m^{\mathrm{IU}}(\tau) z_m^{\mathrm{IU}}(\tau) \Delta\tau \\ 0 \le z_m^{\mathrm{IU}}(\tau) \le Z_m^{\mathrm{IU}}(\tau) \end{cases} \tag{5}$$

Where, $\alpha_m^{\mathrm{IU}}(\tau)$ represents the adjustable load of the *m*-th industrial VPP in τ period. $p_m^{\mathrm{IU}}(\tau)$ represents the regulation compensation coefficient of the *m*-th industrial VPP in τ period. $z_m^{\mathrm{IU}}(\tau)$ represents the percentage of adjustable load interruption of the *m*-th industrial VPP in τ period. $Z_m^{\mathrm{IU}}(\tau)$ represents the maximum interruptible percentage of industrial load during τ period. $\Delta\tau$ indicates the dispatching time interval.

The regulation costs of EV aggregators include power procurement, dispatching management, equipment maintenance and battery dispatching. By optimizing charging dispatching, balancing the demand of electric vehicles, the cost of regulation can be effectively reduced, and the economic benefits can be improved while ensuring the grid stability, which is represented as follows:

$$\begin{cases} C_m^{\mathrm{EV}} = \sum\limits_{T^{\mathrm{PD}}}^{\tau} \sum\limits_{I^{\mathrm{EV}}}^{i} \alpha_m^{\mathrm{EV}}(\tau) \eta_i^{\mathrm{C}} p_{m,i}^{\mathrm{EV}}(\tau) \Delta\tau \\ p_{m,i}^{\mathrm{TE}}(\tau) \le p_{m,i}^{\mathrm{EV}}(\tau) \le p_{m,i}^{\mathrm{EM}}(\tau) \end{cases} \tag{6}$$

Where, $p_{m,i}^{\mathrm{TE}}(\tau)$ represents the charging dispatching power of EV aggregator of the *m*-th VPP in τ period. C_m^{EV} is the number of EV aggregators. $\alpha_m^{\mathrm{EV}}(\tau)$ indicates the compensation coefficient of EV polymerization quotient of the *m*-th VPP in τ period. η_i^{C} represents the charging efficiency of EV polymerization quotient. $p_{m,i}^{\mathrm{TE}}(\tau)$ represents the minimum charging power requirement of EV aggregator.

Distributed energy, as a flexible and adjustable unit in multi-VPPs, is a powerful support for FM standby resources. The cost composition of multi-VPPs distributed energy dispatching is expressed as follows:

$$C_m^{\mathrm{DG}} = C_m^{\mathrm{PV}} + C_m^{\mathrm{MT}} + C_m^{\mathrm{ES}} \tag{7}$$

Where, C_m^{PV} stands for photovoltaic cost. C_m^{ES} stands for energy storage cost. C_m^{MT} stands for micro gas turbine operation and maintenance cost.

Power generation fluctuations and load demand changes require the dispatching system to have efficient dynamic responsiveness. Through intelligent dispatching and optimization of photovoltaic power generation in complex power grid environment, the operation and maintenance costs are reduced. The operation and maintenance cost of photovoltaics is given by:

$$C_m^{\mathrm{PV}} = \sum_{T^{\mathrm{PD}}}^{\tau} \alpha_m^{\mathrm{PV}} p_m^{\mathrm{PV}}(\tau)\Delta\tau 0 \le p_m^{\mathrm{PV}}(\tau) \le p_m^{\mathrm{PVM}}(\tau) \tag{8}$$

Where, $p_m^{\mathrm{PV}}(\tau)$ represents the photovoltaic output provided by the *m*-th VPP in τ period. α_m^{PV} is the unit operation and maintenance cost coefficient of M photovoltaic in virtual power plant. $p_m^{\mathrm{PVM}}(\tau)$ represents the upper limit of photovoltaic maximum output power.

Energy storage (ES) balances load, adjusts grid frequency, and mitigates renewable energy fluctuations. By dynamically adjusting strategies, equipment utilization is maximized, reducing operating costs and achieving optimal dispatching. The operating cost mainly includes the investment cost of equipment, maintenance cost, battery life and energy loss during charging and discharging, etc., which is expressed by the following formula:

$$\begin{cases} C^{\mathrm{ES}} = \sum_{T^{\mathrm{PD}}}^{\tau} \sum_{k} \alpha_{m,k}^{\mathrm{ES}} p_{m,k}^{\mathrm{ES}-}(\tau)\Delta\tau \\ p_{m,k}^{\mathrm{ES}}(\tau) = p_{m,k}^{\mathrm{ES}-}(\tau) - p_{m,k}^{\mathrm{ES}+}(\tau) \end{cases} \tag{9}$$

Where, $p_{m,k}^{\mathrm{ES}}(\tau)$ is the net power. ES is the energy units. $\alpha_{m,k}^{\mathrm{ES}}$ is the kilowatt-hour cost. $p_{m,k}^{\mathrm{ES}-}(\tau)$ and $p_{m,k}^{\mathrm{ES}+}(\tau)$ are the discharge and charging power of the *m*-th VPP and *k*-th energy storage unit in τ period, respectively.

The operating parameters of the energy storage unit, including minimum and maximum discharge power and state of charge range, are crucial for optimal dispatching, ensuring the proper operation and effective regulation in the grid, as expressed below:

$$\begin{cases} p_{m,k}^{\mathrm{EDI}}(\tau) \le p_{m,k}^{\mathrm{ES}-}(\tau) \le p_{m,k}^{\mathrm{EDX}}(\tau) \\ p_{m,k}^{\mathrm{ECI}}(\tau) \le p_{m,k}^{\mathrm{ES}+}(\tau) \le p_{m,k}^{\mathrm{ECX}}(\tau) \end{cases} \tag{10}$$

Where, $p_{m,k}^{\mathrm{ECI}}(\tau)$ and $p_{m,k}^{\mathrm{ECX}}(\tau)$ indicate the minimum and maximum charging unit respectively. $p_{m,k}^{\mathrm{EDI}}(\tau)$ and $p_{m,k}^{\mathrm{EDX}}(\tau)$ are the minimum and maximum discharge power rates of the energy storage unit, respectively.

The energy storage unit's operating parameters, including discharge power limits and state of charge range, are essential for optimal dispatching and effective grid regulation, as shown below:

$$p_m^{\mathrm{Grid}}(\tau) + p_m^{\mathrm{PV}}(\tau) + \sum_{J^{\mathrm{MT}}}^{j} p_{m,j}^{\mathrm{MT}}(\tau) + \sum_{K^{\mathrm{ES}}}^{k} p_{m,k}^{\mathrm{ES}}(\tau)$$

$$= p_m^{\mathrm{IU}}(\tau) + \sum_{l^{\mathrm{EV}}}^{i} p_{m,i}^{\mathrm{EV}}(\tau) + p_m^{\mathrm{OL}}(\tau) \tag{11}$$

Where, $p_m^{\mathrm{Grid}}(\tau)$ is the power purchased by VPP and external power grid during τ period. $p_m^{\mathrm{OL}}(\tau)$ is other load in τ period. $p_{m,j}^{\mathrm{MT}}(\tau)$ is the number of micro gas turbines in virtual power plant m.

4 Collaborative Control Framework Based On DRL

4.1 Problem Description

The power dispatching mode integrates various equipment and electric vehicles to enable flexible energy dispatching, efficient utilization, and support for renewable energy absorption, ensuring grid stability. However, there are still many technical problems in the actual dispatching process of virtual power plants, and innovative control methods are needed to optimize resource allocation and dispatching strategies. VPPs integrates the characteristics and constraints of various resources in the dispatching process, facing optimization challenges in uncertain environments.

Because of its advantages in dealing with complex dynamic systems, deep reinforcement learning has become a potential solution, which can continuously optimize decision-making strategies and adapt to the dynamically changing system requirements. In virtual power plant dispatching, DRL can automatically adjust the dispatching strategy by learning the laws in historical data to cope with various complex environmental changes. However, DRL may have convergence problems in the training process, and may face long training time and high calculation cost in practical application. Therefore, this study aims to explore a collaborative control strategy to enhance resource integration and dispatching efficiency in virtual power plants by optimizing dispatching algorithms, addressing multi-resource collaboration, time-varying loads, renewable energy fluctuations, and optimization under constraints, thereby supporting grid stability and economic operation.

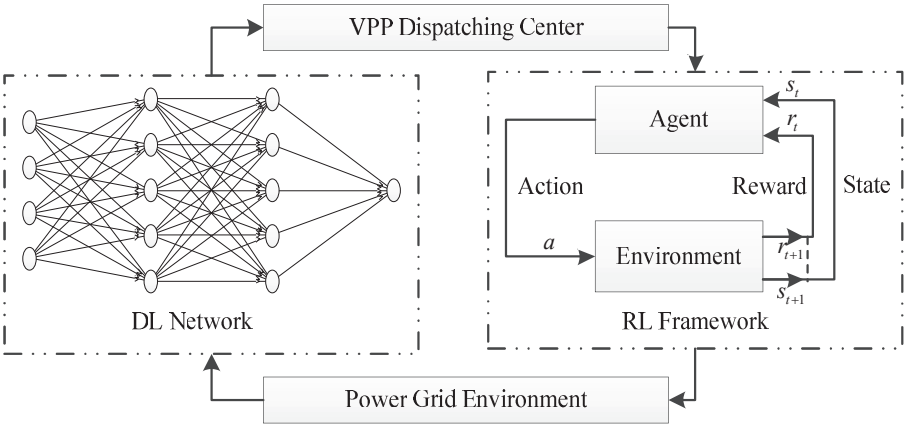

Figure 1 VPP optimal dispatching framework.

4.2 DRL Framework

The essence of DRL is defined by a quadruple (S, A, R, π). Then, S represents the states. A represents the actions. R represents the reward function. π represents the strategy function of the agent. The VPP interactive dispatching framework based on DRL is depicted in Figure 1.

In the figure, s_{t+1} and r_{t+1} are the state and reward of the environment at the next time. The agent is the VPP dispatching center, and the environment is the distribution network in the area where VPP is located.

4.3 Reward Function

The states from the VPP regional environment to the dispatching center include available output and residential load, defining the state space of DRL as follows:

$$S = [P_{\mathrm{WR}}(t), P_{\mathrm{PR}}(t), P_{\mathrm{load}}(t), \Gamma_{\mathrm{B}}(t), \Gamma_{\mathrm{S}}(t), S_{\mathrm{OC}}(t)] \tag{12}$$

Where, $P_{\mathrm{WR}}(t)$ indicates the available output of wind power. $P_{\mathrm{PR}}(t)$ represents the available output of photovoltaic.

The actions taken by the VPP dispatching center include DES charging and discharging power, defining the action space of DRL as follows:

$$A = [P_{\mathrm{S}}(t), P_{\mathrm{grid}}(t)] \tag{13}$$

Where, $P_{\mathrm{S}}(t)$ stands for charging and discharging power. $P_{\mathrm{grid}}(t)$ stands for alternating power.

The reward function should reflect the total benefit, considering both the optimal dispatching target and penalty terms for underutilizing renewable energy output or failing to meet the distribution network's operational constraints.

The penalty terms specifically include the penalty term D_q that DG output decision is less than the penalty term D_q, the penalty term D_p that DG output decision is greater than the penalty term D_p that can be used for wind power and photovoltaic power generation, and the penalty term D_b that DES charge and discharge power decision exceeds the limit, so the penalty function of DRL expressed as:

$$D = D_p + D_q + D_b = c_p \sum_{T}^{t=1} d_{p,t} + c_q \sum_{T}^{t=1} d_{q,t} + c_b \sum_{T}^{t=1} d_{b,t} \tag{14}$$

Where, c_p is the unit difference power penalty for insufficient scenery power. C_q is the unit difference power penalty for abandoning wind and light. C_b is the penalty of unit difference of energy storage overcharge or overdischarge. $d_{p,t}$ is the insufficient power of new energy in t period. $d_{q,t}$ is the amount of abandoned wind and photovoltaic power in t period. $d_{b,t}$ is the amount of energy storage overcharge or overdischarge in t period, which is expressed by the formula as follows:

$$d_{\mathrm{b},t} = \begin{cases} (S_{\mathrm{OC,min}} - S_{\mathrm{OC}}(t)) \cdot E_{\mathrm{b}}, S_{\mathrm{OC}}(t) < S_{\mathrm{OC,min}} \\ (S_{\mathrm{OC}}(t) - S_{\mathrm{OC,max}}) \cdot E_{\mathrm{b}}, S_{\mathrm{OC}}(t) > S_{\mathrm{OC,max}} \end{cases} \tag{15}$$

Where, $S_{\mathrm{OC}}(t)$ represents the DES polymerizer state. Then the reward function in the multi-VPPs dispatching framework expressed as:

$$R = -(C + D) \tag{16}$$

Where, R indicates the reward function of multi-VPPs dispatching framework. C represents the cost of multi-VPPs operation. D represents the penalty term of dispatching framework.

4.4 DQN Collaborative Dispatching Framework

Unlike existing data encapsulation mechanisms for VPPs in optimal dispatching, the proposed model aims to enable cooperative participation of multiple VPPs while fully protecting the physical models and operational data of all participants.

For the adjustable resources with multiple VPPs, the adjustable load and the response behavior of the units belong to a time series markov decision processes. After the distributed energy and adjustable load perform specific power regulation actions, the virtual power plant will move to a new state of regulation and energy demand. In any dispatching period, the upper-level dispatching center only needs to issue an instruction package containing the dispatching period and the dispatching power, and use the traditional unit power regulation ability and the virtual power plant energy management strategy to meet the frequency regulation demand in this period.

For any virtual power plant, the DQN optimal scheduling model takes the obtained scheduling instruction and adjustable load state as the learning environment, and empowers the virtual power plant as an agent to perform adjustable power scheduling in this environment. The dqn optimal scheduling model quantitatively analyzes the return value function of the VPP scheduling system based on whether the scheduling instruction requirements are met, whether the various adjustable load energy requirement. That is, the power scheduling cost and the corresponding punishment. DQN optimal scheduling model obtains better execution actions by constantly learning trial and error, and determines the scheduling power with the minimum cost.

Since the dispatching period setting of multi-VPPs needs to give consideration to both operating cost and response speed, the state space can be divided into multiple subspaces. By dividing 24 hours a day into 96 periods with an interval of 15 min, the optimal dispatching model is based on DQN, and the optimal power dispatching scheme is made according to the current state in each period, thus ensuring the stability of the system and minimizing the dispatching cost. For the multi-VPPs to participate in the whole network frequency modulation, the state space is further improved, which is composed of dispatching instructions from the upper dispatching center, VPP load, real-time compensation coefficients of industrial and EV users, and is expressed as follows by the formula:

$$\begin{cases} s_{1,\tau} = \{t_1^{\mathrm{S}}, T_1^{\mathrm{PD}}, \Delta p_1(\tau), P_1^{\mathrm{IU}}(\tau), P_1^{\mathrm{TE}}, \alpha_1^{\mathrm{IU}}(\tau), \alpha_1^{\mathrm{EV}}(\tau)\} \\ \vdots \\ s_{m,\tau} = \{t_m^{\mathrm{S}}, T_m^{\mathrm{PD}}, \Delta p_m(\tau), P_m^{\mathrm{IU}}(\tau), P_m^{\mathrm{TE}}, \alpha_m^{\mathrm{IU}}(\tau), \alpha_m^{\mathrm{EV}}(\tau)\} \end{cases} \tag{17}$$

Where, $s_{m,\tau}$ represent the state space of multi-VPPs in τ period. $\Delta p_1(\tau)$ represents the charging power demand set of EV aggregator of multi-VPPs. t_m^{S}, T_m^{PD} and $\Delta p_m(\tau)$ respectively represent the start time, duration and scheduling power of the multi-VPPs under the scheduling instruction.

The dispatching power of traditional units can be directly determined by the dispatching center according to the global cost model, so only the action space is considered for the DQN model. By improving the action space, the action space of each virtual power plant includes four actions: industrial load adjustable coefficient, photovoltaic output, EV aggregator scheduling power output and energy storage unit discharge power, which are expressed as follows:

$$\begin{cases} a_{1,\tau} = \{z_{1,\tau}^{\mathrm{IU}}, P_1^{\mathrm{PV}}(\tau), P_1^{\mathrm{EV}}(\tau), P_1^{\mathrm{MT}}(\tau), P_1^{\mathrm{ES-}}(\tau)\} \\ \vdots \\ a_{m,\tau} = \{z_{m,\tau}^{\mathrm{IU}}, P_m^{\mathrm{PV}}(\tau), P_m^{\mathrm{EV}}(\tau), P_m^{\mathrm{MT}}(\tau), P_m^{\mathrm{ES-}}(\tau)\} \end{cases} \tag{18}$$

Where, $a_{m,\tau}$ is the action space of VPP in τ period. $P_1^{\mathrm{PV}}(\tau)$ is the adjustable power set of EV aggregator in VPP.

Reward function is an important part of state updating and guiding learning. Because of the time-varying energy demand of industrial users and EV aggregators in virtual power plants, when VPP can't respond to the frequency modulation demand of the whole network and the energy demand of users can't be met, VPP will be punished accordingly. When the supply and demand of VPP's internal demand side, supply side and energy storage side are unbalanced, VPP needs to buy electricity from external power grid, which increases the operating cost of VPP. For each subject virtual power plant, when performing actions $a_{m,\tau}$ in states $s_{m,\tau}$, the sum of adjustable resource cost and various penalties is set as the return value function $r_{m,\tau}$ by improving the reward function, which is defined as:

$$\begin{cases} r_{1,\tau} = -C_{1,\tau}^{\mathrm{AL}} - C_{1,\tau}^{\mathrm{DG}} - U_{1,\tau}^{\mathrm{VPP}} \\ \vdots \\ r_{m,\tau} = -C_{m,\tau}^{\mathrm{AL}} - C_{m,\tau}^{\mathrm{DG}} - U_{m,\tau}^{\mathrm{VPP}} \end{cases} \tag{19}$$

Where, $r_{m,\tau}$ represent the reward function. Then, the power demand of VPP is defined as:

$$U_{m,\tau}^{\mathrm{VPP}} = U_{m,\tau}^{\mathrm{UI}} + U_{m,\tau}^{\mathrm{EV}} + U_{m,\tau}^{\mathrm{Grid}} \tag{20}$$

Where, $U_{m,\tau}^{\mathrm{UI}}$ and $U_{m,\tau}^{\mathrm{EV}}$ respectively represent the economic penalties for VPP when industrial users and EV aggregators in virtual power plants can't meet the energy demand during τ period. When the percentage of industrial

users' adjustable load terminals is greater than the maximum interruptible percentage coefficient, industrial users can't meet the basic production activities. At this time, the virtual power plant should be subject to economic penalties for industrial users and electric vehicles, which are expressed as follows:

$$\begin{cases} U_{m,\tau}^{\mathrm{UI}} = \delta_{m,\tau}^{\mathrm{UI}} p_m^{\mathrm{IU}}(\tau)[z_m^{\mathrm{IU}}(\tau) - Z_m^{\mathrm{IU}}(\tau)]\Delta\tau \\ U_{m,\tau}^{\mathrm{EV}} = \delta_{m,\tau}^{\mathrm{EV}}[p_{m,i}^{\mathrm{TE}}(\tau) - \eta_i^{\mathrm{C}} p_{m,i}^{\mathrm{EV}}(\tau)]\Delta\tau \\ U_{m,\tau}^{\mathrm{Grid}} = \delta_{m,\tau}^{\mathrm{Grid}} p_m^{\mathrm{Grid}}(\tau)\Delta\tau \end{cases} \tag{21}$$

Where, $\delta_{m,\tau}^{\mathrm{UI}}$ represents the power purchase cost consumption. η_i^{C} represents the adjustable load terminal coefficient.

5 Simulation Studies And Results

5.1 Parameter Configuration

In order to validate the effectiveness of multi-VPPs collaborative optimization scheduling model, the VPP data in July 2024 was selected in the test set, and the VPP was optimized at 96 time points with a time scale of $T = 24$ h a day and an interval of $\Delta t = 15$ min. Then, the total capacity of distributed wind power aggregator is 200 kW, and the total capacity of distributed photovoltaic aggregator is 100 kW. The total capacity of DES polymerizer is 200 kWh, the initial state of charge $SOC = 0.6$, the range of state of charge $SOC_{min} = 0.15$ and $SOC_{max} = 0.8$, and the upper limit of charge and discharge power $\eta = 0.25$. The upper limit of interactive power between VPP and power grid is 180 kW. In addition, the learning rate of DQN algorithm is 0.0003, the discount factor is 0.95, the training batch size is 256, the exploration function is set to random sampling, and the scheduling model after training is directly applied to the test set.

5.2 Contrastive Analysis

To validate its effectiveness, this paper evaluates the performance of the multi-VPPs optimal scheduling model, comparing the results of Q-Learning and IDQN. Q-Learning is famous for its strong self-learning ability, which can balance short-term and long-term goals through reward mechanism and gradually optimize decision-making, and is widely used in power dispatching tasks. Therefore, Q-Learning is selected as the benchmark algorithm and compared with the proposed IDQN method in this study. The average reward

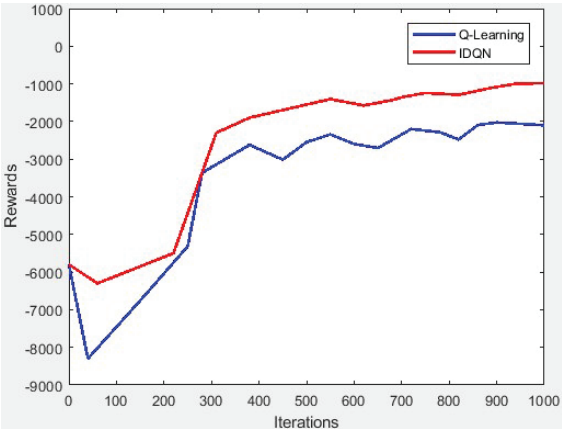

Figure 2 Two models reward in collaborative scheduling of VPPs.

value of the two models in scheduling optimization of virtual power plants is shown in Figure 2.

IDQN converges around 300 iterations with the highest reward value, while the traditional DQN stabilizes after 500 iterations but with a lower reward. The IDQN algorithm shows better convergence and a higher average reward compared to DQN. In addition, because the improved DQN algorithm can dynamically adjust the weight of state space, constantly adapt to the difference of feature weight distribution brought by different data and make a better choice of model, the improved DQN algorithm has better generalization performance and adaptability when dealing with test sets different from training sets.

By setting the same virtual power plant scheduling scenarios and tasks, the fairness of experimental conditions is ensured. By using the improved IDQN algorithm and the traditional DQN algorithm proposed in this paper, the collaborative optimization scheduling of virtual power plants is solved respectively, and the performance of the two models under the same scheduling tasks is compared. The experimental results are shown in Figure 3.

The IDQN model closely matches the real production power values in the VPPs collaborative optimization scheduling, and the deviation is small, which reflects that the model has high accuracy in the scheduling process and can predict and make decisions more accurately in the scheduling tasks of virtual power plants, which is superior to the traditional DQN model. Simulation results demonstrate that the IDQN model adapts well to the dynamic changes in virtual power plants within complex scheduling environments, effectively

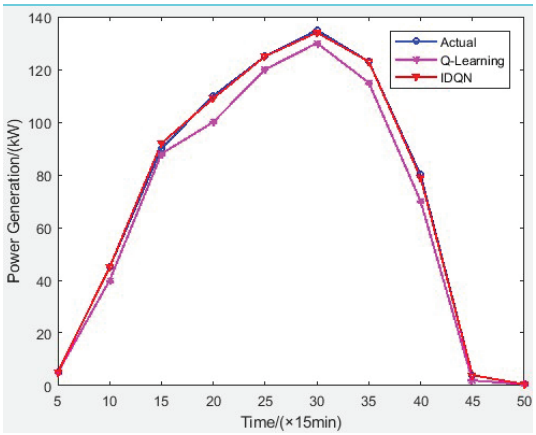

Figure 3 Collaborative scheduling results of VPPs under different models.

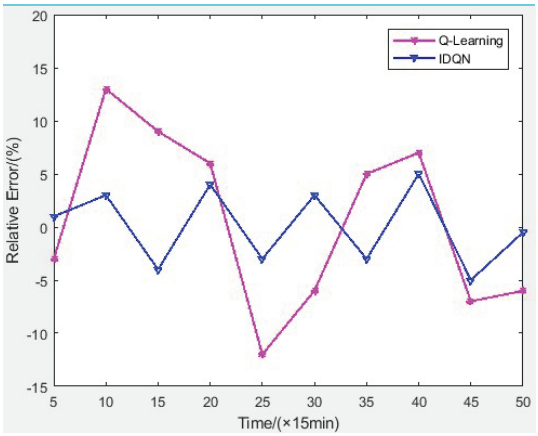

Figure 4 Collaborative scheduling relative error rate of VPPs under different models.

capturing the correlations and dynamics of various factors, thus enhancing scheduling efficiency, accuracy, stability, and reliability.

To provide a more intuitive validation of the IDQN model's performance, the relative error rate between the IDQN and traditional DQN model results is presented in Figure 4.

As can be seen from the figure, the relative error rate of the improved DKN model proposed in this paper always fluctuates within 5% in the process of collaborative optimal scheduling of complex virtual power plants. This result shows that the model has high accuracy and stability in forecasting

Table 1 Comparison results of VPPs collaborative optimal scheduling under different models

Model	VPP Cost (RMB)	Abandon Wind (kWh)	Abandon Photovoltaic (kWh)	Insufficient Power (kWh)	Overcharge Power (kWh)	Total Punishment (RMB)
Q-Learning	989.56	83.45	118.64	24.37	14.25	213.25
IDQN	736.45	58.69	86.81	16.92	6.21	120.95

power grid production. In contrast, the relative error rate of the traditional DQN model fluctuates greatly, even exceeding 12% in some cases, which further proves the superior performance of the improved DQN model in dealing with power grid dispatching tasks. By optimizing the construction of state space, adjusting the reward mechanism and introducing the improved structure of deep neural network, the improved DQN model can capture the dynamic changes in the power grid more accurately, effectively improve the accuracy of dispatching decision, and show stronger adaptability, especially in the situation of large power grid load fluctuation. Therefore, the improved DQN model can significantly reduce the errors in the scheduling process, and provide more reliable technical support for efficient collaborative scheduling of virtual power plants.

To facilitate a better comparison of the scheduling performance of the IDQN model, performance indicators such as VPP cost, waste air volume, waste light volume, insufficient new energy power, overcharge volume, and total punishment are used. The total punishment is an index employed to measure the power dispatching quality optimized by power dispatching algorithms in multi-virtual power plant environment. The smaller the total punishment, the better the optimal scheduling quality and the higher the generalization ability of the algorithm. These multi-dimensional indicators comprehensively evaluate the performance of both models in virtual power plant scheduling. The summarized performance results are shown in Table 1.

As shown in the table, the IDQN model outperforms the traditional DQN model across all performance metrics. In the aspect of VPP cost, the improved model successfully reduced 43.24%, showing its remarkable advantages in resource optimization and cost control. In terms of new energy utilization, the rate of abandoning wind and light decreased by 68.62% on average, showing higher energy utilization efficiency. The improved DQN model also performs well in the shortage rate of new energy and overcharge rate of new energy, which are reduced by 29.36% and 87.54% respectively, effectively improving the stability of power grid dispatching and system security. In

addition, the improved DQN model has also achieved remarkable results in the control of the total penalty, with an average reduction of 62.36%, further reducing the negative impact in the scheduling process. These results indicate that the IDQN model better handles complex load variations and new energy fluctuations in virtual power plants, optimizing resource allocation and improving scheduling efficiency.

6 Conclusion

In this paper, an improved DQN algorithm is innovatively proposed for collaborative optimal scheduling of VPP. Firstly, considering many factors such as power grid load, energy production and energy storage state, the complex and changeable state in the scheduling process of virtual power plant is optimized to improve its state space structure and make it more in line with the scheduling requirements. Then, a new incentive mechanism is designed to address fluctuations, and the IDQN model is proposed to integrate wind, photovoltaic, energy storage, and other sources, enhancing scheduling efficiency and controlling virtual power plant costs to maximize resource utilization. Finally, based on the improved DKN algorithm, a multi-objective decision-making framework is proposed, which takes into account the cost, abandoned wind and light rate, new energy utilization rate and other factors to improve the overall scheduling efficiency and system stability. Simulation results demonstrate that the IDQN outperforms the classical DQN in cost efficiency, abandoned wind and abandoned light rate and total penalty function for multi-VPPs collaborative optimization scheduling. The introduced IDQN model is only evaluated in simulation, and may face challenges such as large dimension of action space and real-time requirements in practical application. Future research should focus on applying the IDQN method to the actual power grid dispatching to verify its effectiveness.

Potential Conflicts of Interest

The author declares that there is no potential conflict of interest.

Funding Information

This paper was supported by Key Research and Development Plan Self Funded Project of Xingtai, Hebei Province in 2023 “Research on Fuzzy

Control of Coke Oven Temperature Based on Configuration Software" (No:2023ZC019).

Acknowledgments

The author thanks the anonymous reviewers for their valuable comments.

Research Involving Human Participants and/or Animals

Not Applicable.

Informed Consent

The author unanimously agreed to the revision and publication of the manuscript.

Data Availability

Not Applicable.

References

[1] H. M. Rouzbahani, H. Karimipour, and L. Lei, "A review on virtual power plant for energy management," *Sustainable energy technologies and assessments*, vol. 47, p. 101370, 2021.

[2] D. Qiu, Y. Wang, W. Hua, and G. Strbac, "Reinforcement learning for electric vehicle applications in power systems: A critical review," *Renewable and Sustainable Energy Reviews*, vol. 173, p. 113052, 2023.

[3] H. Gao, T. Jin, C. Feng, C. Li, Q. Chen, and C. Kang, "Review of virtual power plant operations: Resource coordination and multidimensional interaction," *Applied Energy*, vol. 357, p. 122284, 2024.

[4] Q. Zhang, J. Yan, H. O. Gao, and F. You, "A systematic review on power systems planning and operations management with grid integration of transportation electrification at scale," *Advances in Applied Energy*, vol. 11, p. 100147, 2023.

[5] Y. Kuang et al., "Model-free demand response scheduling strategy for virtual power plants considering risk attitude of consumers," *CSEE Journal of Power and Energy Systems*, vol. 9, no. 2, pp. 516–528, 2021.

[6] X. Liu, "Bi-layer game method for scheduling of virtual power plant with multiple regional integrated energy systems," *International Journal of Electrical Power & Energy Systems*, vol. 149, p. 109063, 2023.
[7] A. K. Podder et al., "Systematic categorization of optimization strategies for virtual power plants," *Energies*, vol. 13, no. 23, p. 6251, 2020.
[8] X. Li, F. Luo, and C. Li, "Multi-agent deep reinforcement learning-based autonomous decision-making framework for community virtual power plants," *Applied Energy*, vol. 360, p. 122813, 2024.
[9] H. Wu, D. Qiu, L. Zhang, and M. Sun, "Adaptive multi-agent reinforcement learning for flexible resource management in a virtual power plant with dynamic participating multi-energy buildings," *Applied Energy*, vol. 374, p. 123998, 2024.
[10] S. Wang, W. Sheng, Y. Shang, and K. Liu, "Distribution network voltage control considering virtual power plants cooperative optimization with transactive energy," *Applied Energy*, vol. 371, p. 123680, 2024.
[11] J. Zhu, P. Duan, M. Liu, Y. Xia, Y. Guo, and X. Mo, "Bi-Level real-time economic dispatch of VPP considering uncertainty," *IEEE Access*, vol. 7, pp. 15282–15291, 2019.
[12] A. Alahyari, M. Ehsan, and M. Mousavizadeh, "A hybrid storage-wind virtual power plant (VPP) participation in the electricity markets: A self-scheduling optimization considering price, renewable generation, and electric vehicles uncertainties," *Journal of Energy Storage*, vol. 25, p. 100812, 2019.
[13] Q. Li et al., "A scheduling framework for VPP considering multiple uncertainties and flexible resources," *Energy*, vol. 282, p. 128385, 2023.
[14] T. Popławski, S. Dudzik, P. Szeląg, and J. Baran, "A case study of a virtual power plant (VPP) as a data acquisition tool for PV energy forecasting," *Energies*, vol. 14, no. 19, p. 6200, 2021.
[15] S.-Y. Park, S.-W. Park, and S.-Y. Son, "Optimal VPP Operation Considering Network Constraint Uncertainty of DSO," *IEEE Access*, vol. 11, pp. 8523–8530, 2023.
[16] X. Wang, H. Lu, Y. Zhang, Y. Wang, and J. Wang, "Decentralized coordinated operation model of VPP and P2H systems based on stochastic-bargaining game considering multiple uncertainties and carbon cost," *Applied Energy*, vol. 312, p. 118750, 2022.
[17] K. M. Muttaqi and D. Sutanto, "A cooperative energy transaction model for VPP integrated renewable energy hubs in deregulated electricity markets," *IEEE Transactions on Industry Applications*, vol. 58, no. 6, pp. 7776–7791, 2022.

[18] B. Feng, Z. Liu, G. Huang, and C. Guo, "Robust federated deep reinforcement learning for optimal control in multiple virtual power plants with electric vehicles," *Applied Energy*, vol. 349, p. 121615, 2023.
[19] Y. Li, W. Chang, and Q. Yang, "Deep reinforcement learning based hierarchical energy management for virtual power plant with aggregated multiple heterogeneous microgrids," *Applied Energy*, vol. 382, p. 125333, 2025.
[20] L. Xue, Y. Zhang, J. Wang, H. Li, and F. Li, "Privacy-preserving multi-level co-regulation of VPPs via hierarchical safe deep reinforcement learning," *Applied Energy*, vol. 371, p. 123654, 2024.
[21] L. Lin, X. Guan, Y. Peng, N. Wang, S. Maharjan, and T. Ohtsuki, "Deep reinforcement learning for economic dispatch of virtual power plant in internet of energy," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6288–6301, 2020.
[22] X. Liu, S. Li, and J. Zhu, "Optimal coordination for multiple network-constrained VPPs via multi-agent deep reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 14, no. 4, pp. 3016–3031, 2022.
[23] Z. Yi et al., "An improved two-stage deep reinforcement learning approach for regulation service disaggregation in a virtual power plant," *IEEE Transactions on Smart Grid*, vol. 13, no. 4, pp. 2844–2858, 2022.
[24] S. M. Nosratabadi, R.-A. Hooshmand, and E. Gholipour, "Stochastic profit-based scheduling of industrial virtual power plant using the best demand response strategy," *Applied energy*, vol. 164, pp. 590–606, 2016.

Biographies

Longfei Yue, Master's degree, Master of Engineering, is a lecturer at the Department of Mechanical and Electrical Technology. His work centers on control theory analysis and control-related engineering applications.

Xianxia Liang, Master's degree, Master of Engineering, is a lecturer at the Department of Electrical Engineering, Hebei Institute of Mechanical and Electrical Technology. Her research interests include mechatronics, industrial robotics, machine vision, and image processing.

Litao Sun, Master's degree, Master of Engineering, is a lecturer at the Department of Electrical Engineering, Hebei Institute of Mechanical and Electrical Technology. His research focuses on intelligent control systems, pattern recognition, and AI-driven defect detection.

Yupeng Li, Master's degree, Master of Engineering, is a lecturer at the Department of Electrical Engineering, Hebei Institute of Mechanical and

Electrical Technology. His research focuses on mobile robotics, embedded systems, applied electronics, and IoT technologies.

Longxue Cheng, Master's degree, Master of Engineering, is a lecturer at the Department of Electrical Engineering, Hebei Institute of Mechanical and Electrical Technology. Her research covers applied electronics, information engineering, mobile robot localization algorithms, path planning algorithms, multi-sensor fusion frameworks, and embedded system development.