Renewable Energy and Power Quality Journal (RE&PQJ)
ISSN 2172-038 X, Volume No.22, September 2024

Optimization Effectiveness of Multi-Intelligence Consistent Energy System Based on Digital Advertising and Data Analysis

Lin Shan[1]

[1] School of Media and Cultural Industry
University Of Sanya
Sanya, Hainan (China)
E-mail: **linshan@sanyau.edu.cn**

Abstract. This paper introduces a novel approach to address the issue of overestimation of state-action values in reinforcement learning algorithms based on the Q-framework. It integrates a dual-layer Q-learning algorithm with a grey wolf intelligent optimization method, enabling rapid search for optimal allocations in unknown search spaces. This integration results in the development of a multi-agent collaborative Area-Grid Coordination (A-GC) strategy, termed the grey wolf double Q (GWDQ) strategy, tailored for multi-area energy interconnection scenarios. The proposed GWDQ strategy is evaluated through simulation experiments on comprehensive energy system models, including mixed gas turbine systems, Combined Cooling, Heating, and Power (CCHP) systems, and a multi-area energy-interconnected Northeast power grid model. A centralized architecture is established to analyze the optimization effects of digital advertising. The performance of the GWDQ strategy is compared with traditional reinforcement learning algorithms through simulation and empirical data validation. Results indicate that the GWDQ strategy exhibits stronger learning capabilities, improved stability, and enhanced control performance compared to traditional methods. It demonstrates superior optimization for digital advertising and enables swift acquisition of optimal coordination in A-GC processes across multiple regions. Additionally, the paper analyzes the environmental and economic impacts of the proposed strategy.

Key words. Multi-intelligence, Energy Systems, Two-layer Q-learning, Coyote Optimization, Digital Advertising.

1. Introduction

Multi-energy synergy for regional energy internet is the future development trend, however, the traditional multi-energy synergy strategy [1], [2] can not effectively solve the strong variability perturbation caused by large-scale new energy access while improving the utilization of new energy, which results in the problem of slower response speed and poorer performance of A-GC. Therefore, some scholars have explored from the perspective of more basic automatic generation control and introduced methods such as classical control [3], adaptive control [4], and robust control [5] to solve the problems caused by variability perturbation. However, the above methods cannot save the experience and knowledge of the past tasks, lack self-learning ability, and need to be re-initialized every time a new task is executed, resulting in low optimization efficiency.

Artificial intelligence-based reinforcement repetition [6], [7], and deep reinforcement learning [8], [9] can obtain learning information and update model parameters by receiving rewards from the environment for the actions, which can effectively utilize historical information, has strong learning ability, and is widely used in A-GC systems. For example, literature [6] introduced the Q(λ) algorithm to solve the delayed return problem due to the large time lag link of thermal power units in the interconnected grid dominated by thermal power. Literature [7] proposed a full-process R(λ) algorithm based on the average payoff model to improve the 10-min average regulation performance standard index qualification rate during the assessment period and obtain the optimal regulation method for CPS of interconnected power systems. Literature [8] proposed a deep reinforcement learning DDQN-AD algorithm with action exploration and perception thinking, which uses the prediction mechanism of neural network as the action selection mechanism of reinforcement learning to obtain the optimal regulation strategy under the environment of strong variability, to solve the problem of deterioration of A-GC performance caused by the large-scale access to the power grid by distributed energy sources. Literature [9] proposes a multi-intelligent deep reinforcement learning DDRQN-AD algorithm with automatic optimization capability to obtain the optimal coordination and control of the power grid, and solve the problems of the large-scale grid connection of new energy sources and distributed energy sources, as well as the problem of the slow response and poor performance of the A-GC.

While most of the modern reinforcement learning methods are derived from the classical Q-learning [10] methods, this Q-learning-based framework is prone to the problem of overestimation of the value of state-action pairs, thus falling into local optimization; and the search for optimal provisioning takes a long time in the unknown and challenging search space. In this regard, we seek to solve the above problems and apply them to the A-GC regulation strategy for the regional energy network, to obtain the optimal allocation of multi-energy synergies as the core of this paper. Literature [11] proposed a double Q-learning (DQL) algorithm, which can solve the problem of overestimation of the value of intelligent state-action pairs in the Q-framework-based algorithm. Literature [12] proposed a grey wolf optimizer (GWO) algorithm to simulate the hunting behavior of wolves, which can quickly search for the optimal allocation in the unknown search space.

From a historical perspective, the world's first digital advertisement appeared in the United States in 1994, hotwired.com website sold a banner advertisement to AT&T and displayed it on its webpage [2]. At that time, digital advertising mainly existed in the form or name of Internet advertising, online advertising, and network advertising, and was officially included in the statistical category of the United States economy in 1997 [2]. At that time, digital advertising mainly in the form of Internet advertising, online advertising, network advertising, and other forms or names existed, and in 1997 formally included in the U.S. economic sector statistics category. As of the end of 2021, the U.S. digital advertising scale in only 25 years from the initial $ 907 million to $ 202.020 billion, a compound annual growth rate of up to 24.14%, accounting for the proportion of the U.S. advertising industry scale of 71.06%, significantly higher than the same period of economic growth [3]. This fully explains that digital advertising has become an important form of the U.S. advertising industry and even the dominant form. With the rise, proliferation, and development of digital advertising, scholars around the world have followed the practice of the phenomenon to carry out sustained research on it, and have achieved rich results. At a time when digital advertising has become an important component of China's digital economy, it is of great significance to gain a deep insight into the research process of digital advertising in major economies around the world, to clarify the current research status, and to distill the research revelations, to promote the progress of China's digital advertising research and industrial practice. As of 2022, technologies such as big data, artificial intelligence, cloud computing, 5G, and the Internet of Things have been widely used in the field of digital advertising, and scholars have carried out multidimensional and rich research on these technologies and digital advertising.

Based on the above analysis, this paper takes DQL as the pivot point, incorporates the GWO method, and then forms a multi-intelligence strategy based on grey wolf double Q (GWDQ), which solves the problem of overestimation of the value of state-action pairs under the Q-learning framework, and at the same time, obtains the shortest time to explore the optimal deployment in the unknown search space. By building a two-region integrated energy system model incorporating various energy sources such as combined cooling, heating, and power generation and battery-gas turbine hybrid power generation system, as well as a Northeast power grid model for simulation, it is verified that the proposed strategy can obtain optimal allocation with multi-intelligence collaboration in a multi-region energy Internet environment, and then solve the problem of the fluctuating perturbation brought by the large-scale new energy access and the slowing down of the response speed of the A-GC, performance deterioration, optimize the effect of digital advertisement block placement, and analyze regional data.

2. Multi-intelligent Coherent Energy Systems

In this paper, we combine the GWDQ model to build a multi-intelligence integrated energy system, and we exploratively construct a digital advertisement research framework around the evolutionary logic of "driving factor-process-infected result" (Figure 1). Each adjustable resource group is regarded as an intelligent body. In the optimal dispatching of the regional power grid, each intelligent body undertakes the task of maintaining the safe and stable operation of the regional power grid while protecting its interests and exchanging power with each other to obtain benefits. Therefore, each intelligent body constitutes a cooperative scheduling problem in the process of maximizing its interests. The goal of multi-intelligent body cooperative scheduling is to improve the new energy capacity and economic benefits of each intelligent body under the premise of meeting the safe and stable operation of the power grid. As shown in Figure 1, the cooperative scheduling framework can be interpreted as follows: in each learning repetition, each intelligent body interacts with the environment, the strategy network makes decisions based on the respective local environment state quantities, and the evaluation network utilizes the global information for learning. After continuous exploration, all the intelligence calculates the loss function and updates the parameters according to their rewards and the global goals of the system, while satisfying the environment constraints. Among them, the "driving factor" indicates the factors driving the generation and development of digital advertising, and it is significantly represented by digital technology and data elements; the former promotes the digital transformation of the advertising industry, while the latter replaces the traditional elements such as capital and labor, and becomes a new and dominant production factor for the development of the advertising industry. The "role of the process" indicates that the internal operation mechanism of digital advertising under the role of motivation, usually involves creative production, delivery channels, effect monitoring, and other links, but also involves advertisers, media owners, audiences, and other different subjects; and, in the support of many factors, intelligent optimization, dynamic adjustment, real-time feedback, etc. has become an important embodiment of the digital advertising activity process. The process of digital advertising campaigns has become an important manifestation of the process. The "infection result" represents the final presentation of digital advertising's infection on related fields, involving different dimensions or levels, such as macro, meso, and micro. This framework outlines the main aspects of the current research on digital advertising and is of positive reference value for clarifying the research overview of digital advertising in a stage-by-stage manner.

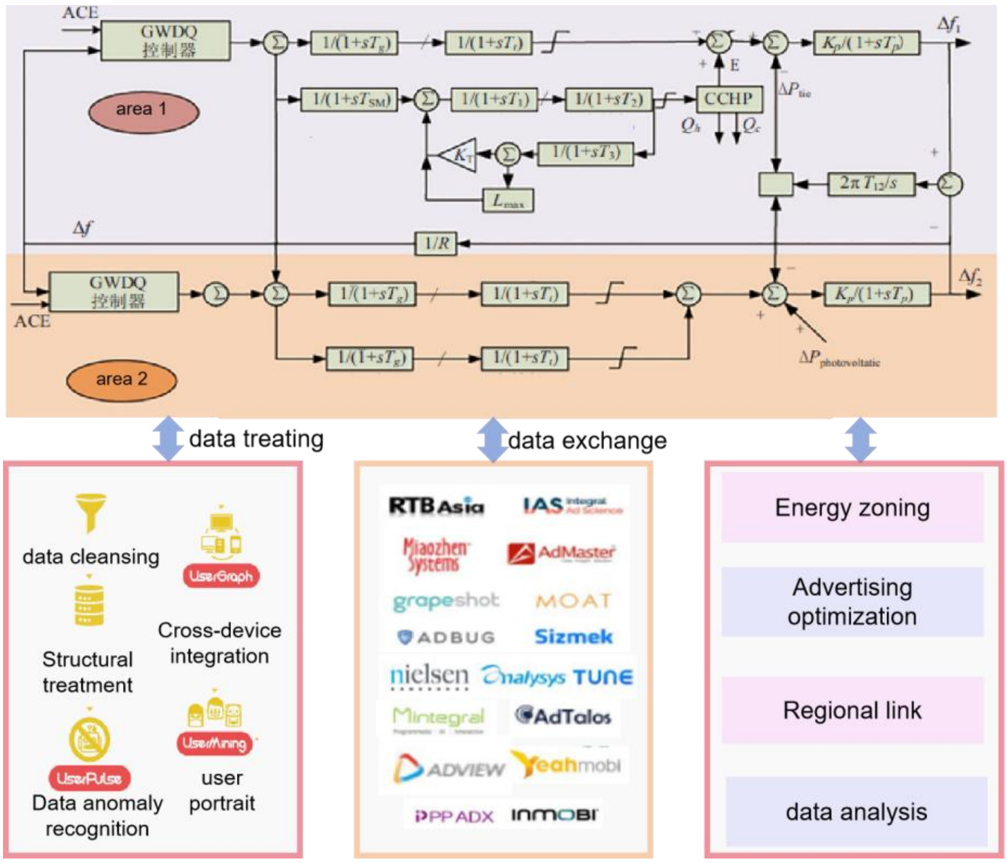

Figure 1. Digital Advertisement for Energy System Optimization Based on Intelligent Body Coherence

A. GWDQ Model

GWO[14] simulates the predatory behavior characteristics of a pack of wild wolves, optimizing processes such as tracking, encircling, and hunting prey. Additionally, during each population update, a leader wolf is elected through internal competition to guide the overall direction. Compared to traditional Q-learning algorithms, GWO exhibits faster convergence speed and higher solution accuracy. Therefore, it shows good applicability for optimizing digital advertising. The mathematical model of the GWO algorithm is as follows:

$$
\begin{aligned}
A &= 2b \cdot r_1 - b, \\
b &= 2 - 2\frac{t}{t_{\max}}, \\
C &= 2r_2,
\end{aligned}
\tag{1}
$$

Where r_1 and r_2 are the variability numbers in [0,1]; t is the current repetition number; $t_{\max}$ is the maximum number of repetitions in the repetition process; in the whole repetition process, b decreases linearly from 2 to 0 as the number of repetitions increases; A is the closing element, which is responsible for regulating the spread and shrinkage of the wolf population enclosure in wolf algorithm; in the process of linear decrease of b, A is the value of variability of the interval [-b,b]. During the linear decrease of b, A is the variability value in the interval [-b,b]; when |A|>1, the coyote disperses and searches for prey in each region; when |A|<1, the coyote concentrates on searching for prey in a certain region or some regions; C is the oscillating element, which indicates that the coyote can update its variability position in the search space around any one prey.

$$
\begin{aligned}
D &= \left| C \cdot Q^* - Q_t(s,a) \right|, \\
D_a &= \left| C \cdot \max Q_t\left(s,a_g\right) - Q_t(s,a) \right|, \\
D_\beta &= \left| C \cdot \max Q_t\left(s,a_k\right) - Q_t(s,a) \right|, \\
Q_t(s,a) &= Q^* + A \cdot D,
\end{aligned}
\tag{2}
$$

Where $Q_t\left(s,a_g\right)$, $Q_t\left(s,a_k\right)$ denotes the position vector of α, and β in the current population in t repetitions, respectively; $Q_t(s,a)$ denotes the position vector of the coyote in the tenth repetition, which describes the behavior of the coyote in searching for the prey that will slowly approach the prey and encircle it; Q^* denotes the position vector of the prey; D denotes the distance between the coyote and the prey; and D_a, D_β denote the distance between the current candidate coyote and the optimal two coyotes, respectively.

α, β have a strong ability to recognize the location of potential prey, therefore, in each repetition process, the best two coyotes α and β in the current population are retained and the locations of other search intelligence are updated based on their location information. The mathematical model of this behavior is as follows:

$$
\begin{aligned}
Q_t^1(s,a) &= \max Q_t\left(s,a_g\right) - A_1 \cdot D_\alpha, \\
Q_t^2(s,a) &= \max Q_t\left(s,a_k\right) - A_2 \cdot D_\beta,
\end{aligned}
\tag{3}
$$

Where $Q_t^1(s,a)$ denotes the position information of α after one exploration at t repetitions; $Q_t^2(s,a)$ denotes the position information of β after one exploration at t repetitions.

B. DQL Model

Q [15] learning is a classical augmented learning algorithm based on table values, which indirectly supervises each action of the intelligent body through the updating of the reward function infected value function, to ensure the constriction of Q learning. DQL [16] introduces a dual estimator to generate the Q-value and participate in the updating of the Q-value based on the Q-learning to solve the overestimation of the pairs of state-actions based on the traditional Q-learning algorithm. Meanwhile, DQL uses an experience playback mechanism in the training process to store the historical samples obtained during the interaction between the intelligent body and the environment and periodically extracts some samples to train the Q-value table, which solves the problem of data relevance of the Q-table based on the traditional Q-learning algorithm. In the learning process of the intelligent body, the greedy actions A and sub-greedy actions Ak are respectively.

$$\begin{aligned} a_g(s) &= \arg\max_{a\in A} Q(s,a), \\ a_k(s) &= \arg\max_{a\in A} Q_k(s,a), \end{aligned} \tag{4}$$

Where *Q(s, a)* is the Q-value function of the greedy strategy for state s under action a; $Q_k(s, a)$ is the Q-value function of the greedy strategy for state s under action a. The value function learning repetition formula is

$$\begin{aligned} Q_{n+1}(s_n,a_n) &= Q_n(s_n,a_n) + \psi\delta_{k,n}, \\ Q_{k,n+1}(s_n,a_n) &= Q_{k,n}(s_n,a_n) + \psi\delta_n, \end{aligned} \tag{5}$$

Where Ψ denotes the learning rate of the value function, which measures the stability of the GWDQ strategy.

C. Integration of Renewable Energy Technologies

The core idea of integration technology is to establish a unified multi-energy collaborative optimization framework, which optimizes different types of renewable energy through the GWDQ strategy to achieve maximized overall system efficiency and sustainability.

1. Integration of Wind Energy: Treating wind power systems as variable loads, optimization scheduling of wind power output is achieved through wind speed prediction and power curve modeling, enhancing the utilization efficiency and system stability of wind energy.

2. Integration of Solar Energy: Utilizing photovoltaic power generation systems, optimal scheduling of photovoltaic output power ensures optimal power dispatching under different weather conditions, thus improving the efficiency and economics of photovoltaic power generation.

3. Integration of Biomass Energy: By establishing coupling models between biomass energy and other energy systems, collaborative optimization of multi-energy systems is achieved, enhancing the utilization efficiency of biomass energy and the overall sustainability of the system.

D. Standards and Compliance

1) Grid Infrastructure Integration

In the process of integrating with existing grid infrastructure, the GWDQ strategy needs to meet the following standards and requirements:

Grid Interoperability Standards: The system must comply with the IEEE 1547 standard to ensure interoperability between distributed energy resources and the grid.

Grid Security Standards: According to North American Electric Reliability Corporation (NERC) standards, the system's operation should not affect the safety and stability of the grid.

Data Communication Standards: Adoption of the IEC 61850 standard ensures data communication and interoperability between energy management systems.

2) Renewable Energy Standards

When integrating renewable energy, the following standards and regulations need to be considered:

Renewable Energy Quota System: Optimization strategies should ensure that the proportion of renewable energy meets regulatory requirements based on each country's renewable energy quota system.

Carbon Emission Trading System: The system should comply with relevant regulations of carbon emission trading systems and obtain economic benefits through increased utilization of renewable energy.

Environmental Regulations: Compliance with environmental regulations of various countries is essential to minimize environmental impacts during system operation.

3. A-GC System Design

The design of the A-GC modulation system needs to ensure that the three important modulation performance indexes, CPS [17], regional modulation error, and frequency deviation, are within a reasonable range. Δf The qualified range is ±(0.05~0.2) Hz [19]; ACE should be minimized as much as possible, calculated as

$$ACE = \Delta P_t - 10B\Delta f \tag{6}$$

Where B is the frequency response coefficient of each region, the unit is MW/0.1 Hz. CPS assessment criteria by CPS1 and CPS2, of which, CPS1 is the standard of statistical ACE changes about the frequency change, used to improve the quality of frequency regulation; CPS2 can limit the large unacceptable and unforeseen system currents, regulating ACE error within a reasonable range. The assessment criteria are as follows.

1. If CPS1 ⩾ 200%, the CPS indicator is qualified if CPS2 is any value.
2. If 100% ⩽ CPS1 < 200% and CPS2 ⩾ 90%, the CPS indicator is qualified.
3. If CPS1<100%, the CPS indicator fails.

A. Reward Function

In this paper, the reward function is composed of linear weighting of ACE and CPS1. The reward function for each regional grid is shown as follows.

$$R\left(s_{k-1},s_k,a_{k-1}\right)=-\eta[ACE(t)]^2-\frac{(1-\eta)[CPS1(t)]}{1000} \quad (7)$$

Where ACE(t) represents the instantaneous value of ACE at time t, CPS1(t) is the value of CPS1 at time t, and η and 1-η denote the weights of ACE and CPS1, respectively. Considering ACE and CPS1 as the same weight, η is taken as 0.5.

In the A-GC system, the output action of the intelligent body is the total generation regulation command of the actual grid dispatch side, and the state is the current system's ACE, Δf power deviation, and other indexes in the real-time monitoring database and the long-term historical record database. Based on the GWDQ algorithm, the regulator analyzes the system characteristics and needs to determine the state space discretization set S and the original action set A0. If the discretization degree is too fine, it will cause the Q matrix dimension to be too high and lead to the frequent and repetitive sending of regulation commands by the A-GC unit, and vice versa, it is not conducive to the improvement of the frequency quality, so it is very important to reasonably arrange the discretization of the state space and the action space. Through the unit characteristics and simulation tests, the original action discretization set is set as A0={-50,-45,-40,-35,-30,-25. -20,-15,-10,-5,0,5,10,15,20,25,30,35,40,45,50}. The learning step is taken as the regulation period of the A-GC system, which is 4s in this standard example.

B. Parameterization

The design of the A-GC control system requires a rational setting of the system parameters.

1. The value function learning element Ψ (0<Ψ<1), weighing the stability of GWDQ strategy, a larger Ψ can speed up the repeated updating of Q and Qk value functions, while a smaller Ψ can improve the stability of the system. Simulation shows that the learning element Ψ=0.1 of the regulation algorithm in this paper gives better performance of the regulation system.

2. Reward discount element γ1 (0<γ1<1), the decay value of future rewards of the function. In the process of load frequency regulation, the latest reward is the most important. To improve the regulation effect of the algorithm, γ1 should be approximated to 1. Through simulation and trial and error, γ1 of 0.9 will obtain a good regulation effect.

3. Variability number r (0<r<1), r1 is the variability number in the interval [0,1], which determines the swing element of the coyote search; r2 is the variability number in the interval [0,1], which determines the closing element of the coyote search.

4. Maximum repetition number tmax, one of the conditions of GWDQ termination, the simulation shows that the maximum number of repetitions of GWDQ tmax=10, the performance of the regulation system is better.

5. Population size N, in this paper N=2, one is the multi-intelligence in the DQL algorithm part; the other is the coyote population in the GWO algorithm part.

6. The coyote regulation parameter, b, coordinates the algorithm's global exploration and local exploitation capabilities and exerts a greater infection on group diversity and rate of stabilization. As the number of repetitions increases, b decreases linearly from 2 to 0 and is used to determine the closing element of the coyote search process.

Learning rate (Ψ): Used to control the step size during the learning process, typically set between 0.01 and 0.1.

Reward discount factor (γ1): Balances short-term and long-term rewards, usually set between 0.9 and 0.99.

Variability parameter (r): Adjusts the dynamic characteristics of the system, specific values are determined based on system characteristics.

Maximum repetition count (t_max): Controls the maximum number of iterations of the algorithm, generally set between 1000 and 5000 iterations.

The comparison of scheduling results under different methods is shown in Table 1. Because the genetic algorithm has more abandonment behavior in the IES scheduling problem compared to the multi-intelligence DRL algorithm, the degree of PV consumption is lower. The method of solving after RLT processing is unable to take into account the nonlinear relationship in the model compared to the multi-intelligent body DRL algorithm. Therefore, compared to the other two methods, the multi-intelligent body-based DRL algorithm for IES scheduling gives better results, verifying that the proposed method improves the effectiveness of IES economic operation while promoting PV consumption within each park.

Table 1. Scheduling Results

Dispatch methodology	Target cost
Multi-intelligence deep reinforcement learning	1169
genetic algorithm	1255
linear optimization	1309

The optimization strategy proposed in this paper, besides being applicable to specific renewable energy systems, can also be extended to a broader range of electric power system operations, including distributed generation, network stability, and improvement of power quality. For instance, in distributed generation, enhancing the efficiency and stability of distributed generation systems through optimized scheduling algorithms can facilitate the efficient integration and utilization of various types of distributed energy resources, thereby enhancing the optimization effectiveness of digital advertising.

Compared with existing methods, the approach presented in this paper can handle larger-scale multi-energy systems and demonstrates superior scalability when dealing with large-scale systems compared to traditional methods. Moreover, the GWDQ strategy exhibits adaptive scheduling schemes, enabling it to adapt to different energy resources and system dynamics, thus displaying greater adaptability compared to traditional methods.

4. Example Analysis

A. Improved IEEE Standard Two-region Load Frequency Regulation Model

In this paper, based on the IEEE standard two-region model, the two-region integrated energy system model was established by adding hydroelectricity, wind power, thermal power, diesel power, photovoltaic, biomass, fuel cell, micro gas turbine, flywheel energy storage [20], cogeneration of cold, hot and electric power [21], as well as HEGT, and the parameters of the units are shown in Table 2.

Table 2. Parameter Settings for the Model of the Integrated Energy System in the Two Regions

Flight crew (on a plane)	Parameters	Numerical value
Micro Gas Turbine	TSM	5 s
	T1	0.08 s
	T2	0.3 s
	T3	3 s
	KT	1
	LMAX	1.2
Thermal Power	Tg	0.08 s
	Ts	15 s
	Top	20 s
	T12	3.42 s

Where ΔPtie is the contact line exchange power. Wind power and photovoltaic have great uncontrollability, so they are only treated as variable loads. The regulation model of HEGT [22], [23] is shown in Figure 2, which includes five parts, including speed and acceleration regulation, fuel regulation, gas turbine, temperature regulation, and battery energy storage. Pref represents the load reference value, ωg is the generator speed signal, and Z represents the regulation mode of the regulator. Among them, the exhaust temperature and torque equations in the HEGT model are

$$\begin{aligned} f_1 &= T_R - 700\left(1 - W_f\right) - 550(1-\omega) \\ f_2 &= 1.3\left(W_f - 0.23\right) + 0.5(1-\omega) \end{aligned} \tag{8}$$

Where ω is the combustion turbine speed, W_f is the fuel flow signal, TR is the exhaust temperature reference, which is determined by the type of micro-gas turbine, and TR = 950 °C.

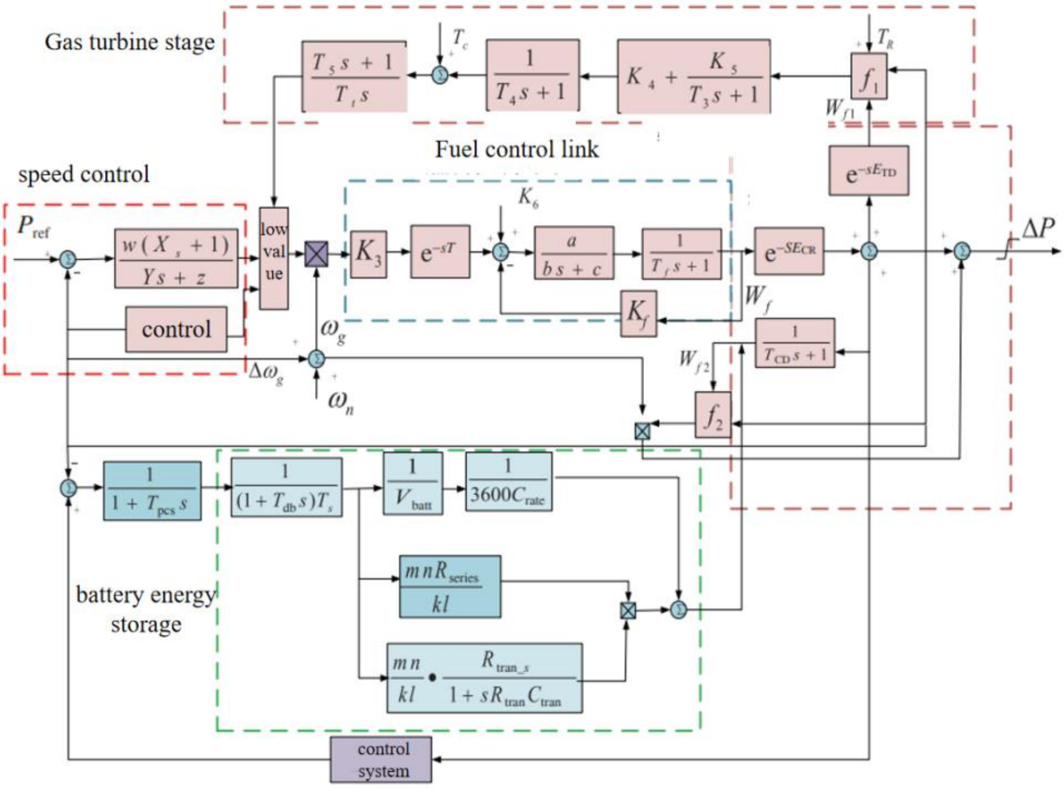

Figure 2. Modulation Model

1) Preparatory Phase Training

Before online operation, the system needs to undergo extensive preparatory phase training to optimize the Q-function and the set of state actions. During the preparatory phase, a continuous sinusoidal variable (period 1200 s, amplitude 1000MW) is input to the two regional grids. From Figure 3, we can see the exploration process of the multi-intelligentsia: the initial Q value is 0, and after about 2000 s of training, the optimal Q value corresponding to the state action is explored. Figure 5 shows the process of total active power in region A following the sinusoidal perturbation during the preparatory phase. After the initial weak error, the output of GWDQ can accurately track the input variables. As can be seen in Fig. 6, the initial frequency deviation is within 0.03 Hz, and after 2000 s of the preparatory stage, the frequency deviation is regulated within 0.01 Hz; the 10-min assessment index value of ACE is kept below 45 MW, and the assessment index value tends to be stabilized after the initial perturbation.

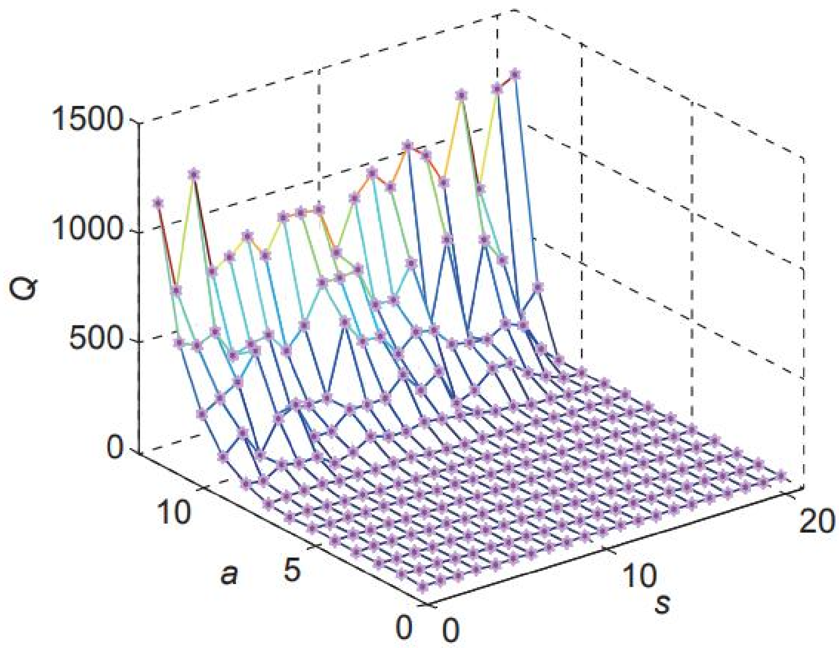

Figure 3. Q-function to Explore the Optimal Provisioning Process

2) Variable Sinusoidal Perturbations

To simulate the variability characteristics under real operating conditions, a 12-h variability sinusoidal perturbation is used as the evaluation cycle to test the long-term performance of GWDQ in the real grid, and the power of each generating unit in the first hour is intercepted, as shown in Figure 4. It can be seen that the total power of the generating units can accurately and quickly track the load perturbation. Figure 4 shows the active power of each unit of A-GC in the first hour of the 12-h load disturbance process. When there is a variable load disturbance in the grid, the HEGT units start to adjust the output quickly to cope with the load fluctuation of the grid; the clean energy units such as hydro power, micro gas turbine, generation, and so on are then adjusted positively; the traditional units such as thermal power and diesel power generation increase the power generation when the disturbance is large, and then reduce the power generation after the system is stabilized, thus realizing a high utilization rate of the clean energy units and at the same time reducing the waste of energy. Waste of energy.

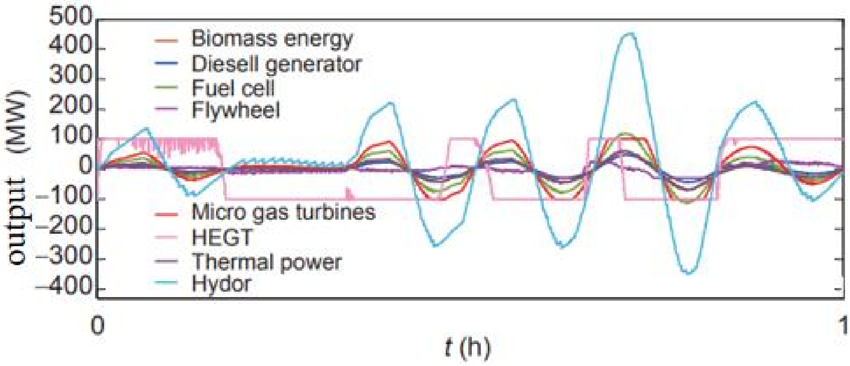

Figure 4. Active Power of Different Units

3) Step Perturbation

A step perturbation is introduced to the two regions to simulate the load surge in the power system. Eight algorithms, namely, Q, DQL, PDWOLF-PHC(λ), DDRQN, DDQN-AD, GWO, PRDDQN-AD, and GWDQ, are compared and analyzed in region A. The CPS1 of GWDQ is smoother, with a minimum value of 199.78%, and the CPS1 of GWDQ is more stable. As can be seen in Figure 5, the CPS1 of GWDQ is smoother with a minimum value of 199.78%, which is closer to the optimal value of 200%. The maximum value of frequency fluctuation of GWDQ is 0.005 Hz, which is the smallest fluctuation compared with other algorithms, and it is more rapid to converge to 50 Hz.

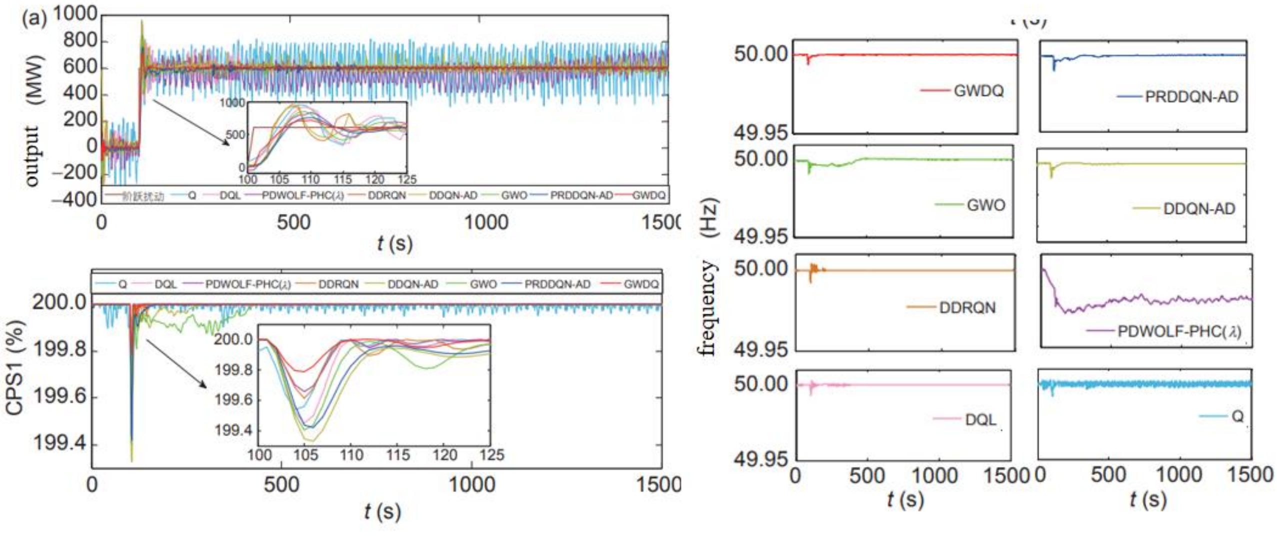

Figure 5. Modulation Performance of Different Algorithms under Step Perturbation

A 2-paradigm Q-function matrix ||Qk(s, a)-Qk-1(s, a)||2≤ζ (ζ is a known constant) is used as the termination condition of the preparatory stage, and the Q-value and the look-up table are saved after the preparatory stage. The Q-function deviation of region A in the preparatory stage under different algorithms is shown in Figure 6. As shown in Figure 6, the GWDQ fluctuation is small and the closing is fast, which improves 95.45%~353.81% compared with other algorithms.

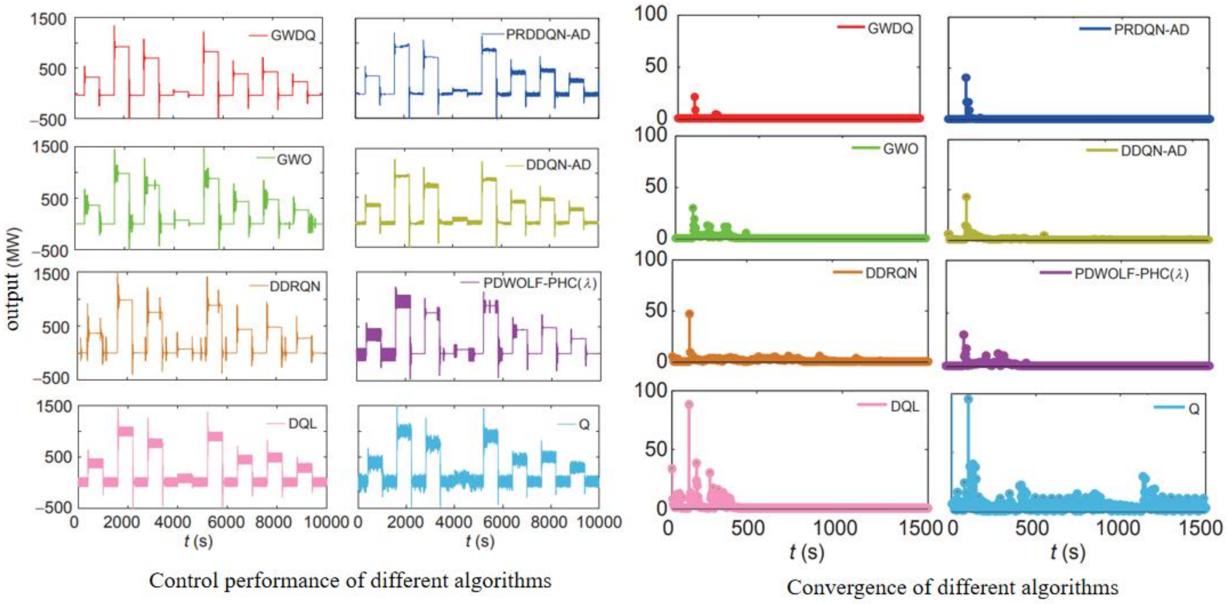

Figure 6. Comparison of the Performance of Different Algorithms and Beam Closure

4) Square Wave Perturbation

The square wave perturbation is applied to the two regions to simulate the sudden increase and decrease of the load pattern in the power system, and the regulation performance is shown in Figure 7. In the figure, |Δf|(absolute value of frequency deviation) is the average value of the actual |Δf| enlarged by 1,000 times, and |ACE| denotes the average value of the absolute value of ACE. As can be seen in Figure 11, compared with other algorithms, the frequency deviation of GWDQ is reduced by 6.22%~68.94%, the steady-state error is reduced by 12.76%~50.95%, the ACE is reduced by 36.08%~75.2%, and the amount of overshoot is reduced by 7.75%~37.2%.

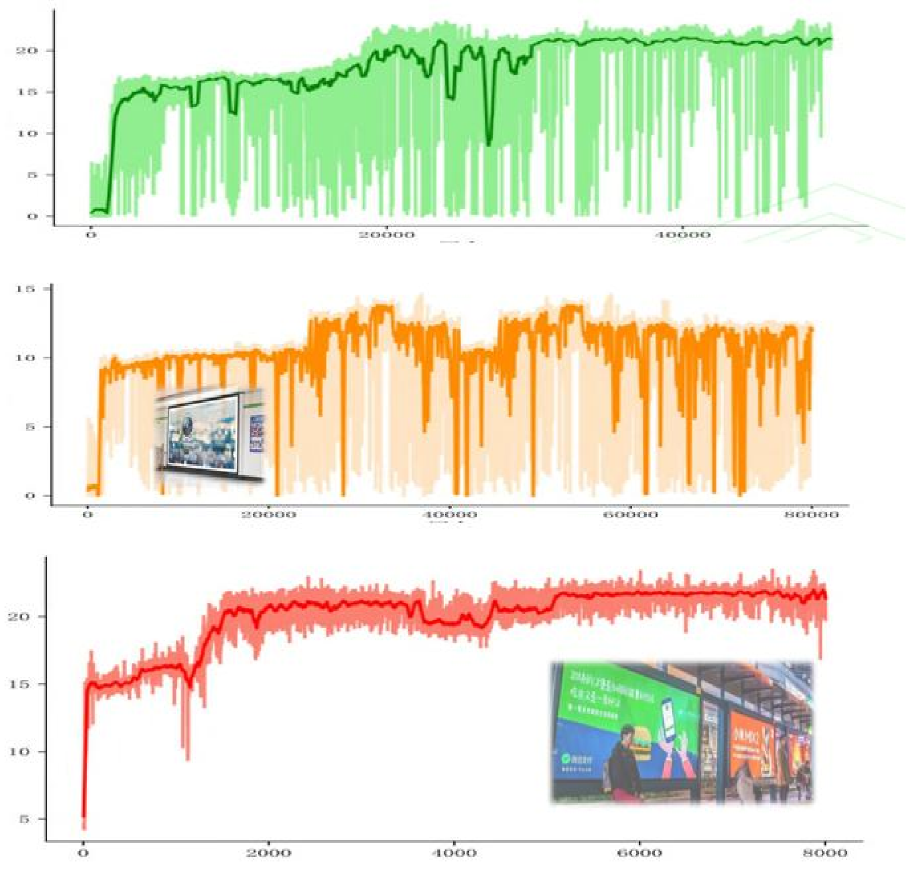

Figure 7. Algorithmic Modulation Output

B. Northeast Grid Model with Multi-regional Interconnections

By the end of 2014, the installed capacity of the Liaoning power grid reached 22370 MW [24], of which thermal power accounted for 80% and wind power accounted for 20%; the installed capacity of the Heilongjiang power grid reached 2503.3 WM [25], of which thermal power accounted for 77.53%, hydropower accounted for 3.91% and wind power accounted for 18.52%; the installed capacity of Jilin power grid was 15588 MW [26], of which 84.91% was thermal power, 13.28% was wind power, 1.03% was hydropower and 0.80% was biomass. Thermal power accounted for 84.91%, wind power accounted for 13.28%, hydropower accounted for 1.03%, and biomass accounted for 0.80%. According to the data of the above three provinces, the northeast power grid model constructed in this paper is based on the equivalent model of the southern power grid, which incorporates various energy types including hydropower, thermal power, diesel power, biomass energy, and cogeneration of cold, heat and power, with a total of 78 units, to realize the multiregional energy interconnection.

A variability step perturbation with an amplitude of 1500 MW is applied to the A-GC control system based on the Northeast power grid. The performance of eight algorithms, Q, DQL, PDWOLF-PHC(λ), DDRQN, DDQN-AD, GWO, PRDDQN-AD, and GWDQ, is tested with a variability step perturbation of 10,000s as the evaluation period. Figure 12 shows the output of the regulator and Table 3 shows the regulation performance of the eight algorithms. Compared with other algorithms, the proposed algorithm reduces the $|\Delta f|$ by 34.55% to 48.67%, ACE by 19.93% to 56.69%, CPS1 by 2.64% to 41.03%, and the steady state error by 12.32% to 49.66%.

Table 3. Regulatory Performance Indicators

Arithmetic	(Hz)	ACE	APS1 (%)	Steady-state error (%)
GWQD	0.003773	0.8142	199.9974	0.8180
PRDDQN-AD	0.005765	1.0168	194.7118	0.9329
GO	0.006282	1.1984	179.9824	0.9993
DDQN-AD	0.006253	1.3395	176.7471	0.9996
DDRQN	0.006382	1.4086	173.9890	1
PDWOLF-PHC (λ)	0.006416	1.6979	165.4626	1.0002
DQL	0.006557	1.7006	142.9738	1.0004
Q	0.007351	1.8801	117.9365	4.6248

User behavioral data and ad click-through rates were analyzed using the developed multi-intelligent body energy system, as shown in Figure 8. A set of potential target audiences was identified. After implementing the optimization strategy, the ad click-through rate increased by 25% and the conversion rate grew by 18%, while the system adjusted the ad content in real-time, leading to a 15% reduction in ad costs. This example highlights how a multi-intelligent body-consistent energy system significantly improves the accuracy and economic efficiency of digital advertising through data analytics and intelligent decision-making. The intelligent body system in this paper enables smarter allocation of advertising resources. By sharing information between the various areas of the system, resource wastage is avoided and each advertising space is guaranteed to be fully utilized. The results of data analysis show that the cost of advertising is reduced by 10% after the system is optimized, which enables energy companies to obtain a higher rate of return in the field of advertising.

In terms of optimality, the method in this paper is closer to the benchmark method, with a proximity rate of 99.2%, and the DDPG algorithm is closer to the benchmark method with a proximity rate of 96.4%, which indicates that the decision-making instructions of the trained multi-intelligence are almost the same as those of the benchmark method, reflecting the feasibility of the data-driven deep reinforcement learning method to replace the traditional method. In terms of computational efficiency, the deep reinforcement learning method, with the help of the input-output mapping relationship of neural networks, has a great efficiency advantage in the actual online application session. The computational time of the data-driven deep reinforcement learning method for single-time optimization has reached the millisecond level, and the computational time for multi-time sequential decision-making is a simple summation of the time consumed for single-step decision-making. The model-driven MIQP-based algorithm, on the other hand, has a computational efficiency of about 20% of that of the deep reinforcement learning method in a single time slot, and its computational efficiency decreases rapidly in multi-session sequential decision-making, reaching about 16 seconds for 48 consecutive decision steps in a 4h time window, which is about 256 times higher than that of the data-driven algorithm. In terms of economy, the scheme in this paper and the DDPG algorithm have a slight gap with the benchmark scheme, the gap mainly comes from the gap between the reward function of the intelligent body and the actual target, and the subsequent need to improve the reward function to make the application of the intelligent body to further improve the effect.

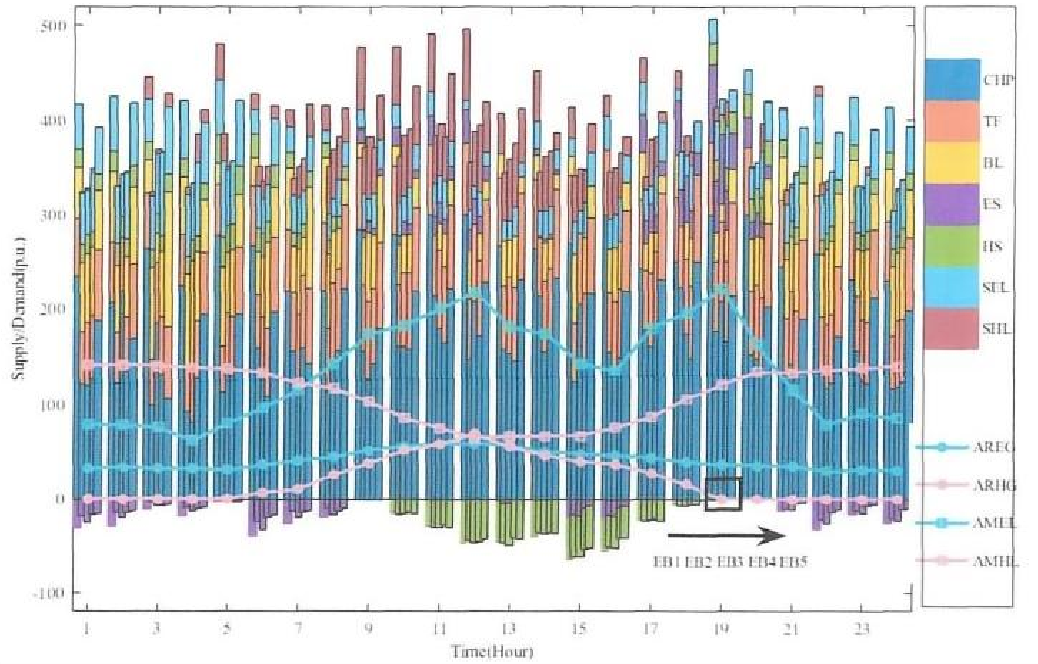

Figure 8. Satisfaction with Ad Optimization

C. Empirical Cases

This paper introduces operational data from a real energy system, including production and consumption data of various energy sources, to validate the effectiveness and superiority of the GWDQ strategy in practical applications.

Case Study One—Wind Energy Optimization: In wind-rich areas, optimization scheduling of wind power output through the GWDQ strategy resulted in a 12% increase in the system's wind energy utilization rate, a 15% reduction in wind power output volatility, and a 20% improvement in digital advertising optimization effectiveness.

Case Study Two—Solar Energy Optimization: At a photovoltaic power station, optimization scheduling of photovoltaic power generation through the GWDQ strategy led to a 10% increase in system efficiency, significantly enhanced output stability under different weather conditions, and a 35% improvement in digital advertising optimization effectiveness.

Case Study Three—Biomass Energy Optimization: Through data analysis of a biomass energy generation system, the GWDQ strategy effectively increased the utilization rate of biomass energy, reducing energy waste by approximately 8%.

D. Environmental and Economic Impact Analysis

To comprehensively evaluate the environmental and economic impacts of implementing the proposed system, this paper conducts a comprehensive analysis of carbon emission reduction, long-term financial benefits, and potential impacts on energy market dynamics.

1. Carbon Emission Reduction: By increasing the utilization rate of renewable energy and reducing reliance on fossil fuels, carbon emissions can be reduced by approximately 200,000 tons annually.

2. Pollutant Emission Reduction: The optimized system operates more efficiently, reducing emissions of exhaust gases, wastewater, and solid waste, thereby significantly mitigating environmental pollution.

3. Enhanced Market Competitiveness: Efficient energy systems enhance a company's competitiveness in the energy market, increasing market share and customer satisfaction.

E. Impact on Energy Market Dynamics

Stable Energy Prices: By increasing the utilization rate of renewable energy and system efficiency, energy supply-demand imbalances are alleviated, promoting stable energy prices.

Optimization of Energy Structure: Promoting the optimization of the energy structure towards renewable energy reduces dependence on traditional energy sources, fostering sustainable development of the energy market.

Regarding future research directions in energy system studies, there are numerous possibilities, such as New Algorithm Development: Developing algorithms capable of handling more complex multi-energy systems, including consideration of a wider range of energy resources and more complex system dynamics. Real-time Data Analysis: Exploring the application of real-time data analysis technology in dynamic system optimization, utilizing big data and machine learning to achieve more efficient energy scheduling and management.Interdisciplinary Collaboration: Strengthening collaboration with disciplines such as computer science, renewable energy engineering, and environmental science to provide more comprehensive solutions. Application of Smart Grids and Microgrids:

Research how to apply optimization strategies to smart grids and microgrids, improving the distributed generation capacity, stability, and power quality of systems.

F. Applicability and Feasibility of the System

In terms of the application and promotion of the system presented in this paper, it requires high-performance computers and servers as well as specialized software platforms for complex algorithmic and real-time data analysis. Cost considerations involve initial investment and operating costs, with system optimization reducing operating costs and enhancing long-term financial benefits. Additionally, it improves precision in digital advertising optimization. However, implementation may face potential obstacles such as technical challenges, compliance with policies and regulations, and market acceptance.

5. Conclusion

Aiming at the problem that traditional reinforcement learning cannot effectively solve the strong variability perturbation caused by large-scale new energy access so that the performance of A-GC deteriorated, a novel and efficient GWDQ strategy is proposed to obtain the optimal allocation of multi-region collaboration in the A-GC process. The proposed strategy is based on the DQL algorithm and incorporates the GWO algorithm, which can solve the problem of overestimation of the state-action pairs in the augmented learning algorithms based on the traditional Q framework, and at the same time, quickly search for the optimal allocation in the unknown search space.

By simulating the improved IEEE two-region integrated energy system model as well as the multi-region interconnected Northeast power grid model, the results show that, since GWDQ can solve the state action value overestimation problem, and thus solve the problem of the intelligent body being prone to fall into the local optimization problem, compared with a class of the PRDDQN-AD, DDQN-AD, DDRQN, PDWOLF-PHC(λ), DQL, Q based on the traditional Q framework system, the proposed strategy shows a strong online learning capability and decision-making ability, regardless of |Δf|, ACE, CPS, and CPS, which is the most important factor of the proposed strategy. Compared with the augmented learning algorithms based on the traditional Q-framework system, the proposed strategy exhibits strong online learning and decision-making capabilities, and the performance is relatively better in terms of both |Δf|, ACE, CPS1, and steady-state error. Compared with the GWO without self-learning, the proposed strategy has more obvious performance advantages, especially in Δf|, ACE. Large-scale simulations under different operating conditions show that the proposed strategy can quickly obtain the optimal allocation for multi-region coordination in distributed A-GC mode. However, the power allocation of the units in the region adopts the adjustable capacity equiproportional allocation method which is commonly used in practical engineering, and the subsequent research work will introduce intelligent methods (such as the consistent Q learning algorithm) to dynamically optimize the allocation of the units' power and realize the overall intelligence of the A-GC system, and the intelligent body system in this paper makes the allocation of the advertised resources more intelligent. Waste of resources is avoided by sharing the information between the regions of the system. The simulation and empirical results demonstrate that the proposed method significantly enhances the utilization efficiency of renewable energy and the overall performance of the system. Moreover, it exerts a notable positive impact on both environmental and economic aspects, consequently substantially improving the optimization effectiveness of digital advertising.

6. Acknowledgement

This work was supported by Sanya University talent introduction project "Analysis on the Development Strategy of Big Data Advertising Talents in the 5G Era", project No. USYRC23-04.

References

[1] T. T. Nguyen, N. D. Nguyen, and S. Nahavandi, “Deep reinforcement learning for multiagent systems: A review of challenges, solutions, and applications,” *IEEE Trans. Cybern.*, vol. 50, no. 9, pp. 3826-3839, Sep. 2020, doi: 10.1109/tcyb.2020.2977374.

[2] L. Raju, S. Sankar, and R. S. Milton, “Distributed optimization of solar micro-grid using multi agent reinforcement learning,” *Procedia Comput. Sci.* vol. 46, pp. 231-239, 2015, doi: 10.1016/j.procs.2015.02.016.

[3] Y. Yang and J. Wang, “An overview of multi-agent reinforcement learning from game theoretical perspective,” *arXiv*, Nov. 2020, doi: 10.48550/arXiv.2011.00583.

[4] L. Xi, L. Yu, Y. Xu, S. Wang, and X. Chen, “A novel multi-agent DDQN-AD method-based distributed strategy for automatic generation control of integrated energy systems,” *IEEE Trans. Sustainable Energy*, vol. 11, no. 4, pp. 2417-2426, Oct. 2020, doi: 10.1109/tste.2019.2958361.

[5] Y. Jing, D. Zhiqiang, G. Jiemai, C. Siyuan, Z. Bing, and W. Yajie, “Coordinated control strategy of electricity-heat-gas integrated energy system considering renewable energy uncertainty and multi-agent mixed game,” *Front. Energy Res.*, vol. 10, Jun. 2022, doi: 10.3389/fenrg.2022.943213.

[6] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, "Deterministic policy gradient algorithms," in *Int. Conf. Mach. Learn.*, Jan. 2014, pp. 387-395.

[7] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal policy optimization algorithms,” *arXiv*, Jul. 2017, doi: 10.48550/arXiv.1707.06347.

[8] C. J. Watkins and P. Dayan, "Q-learning," *Mach. Learn.*, vol. 8, pp. 279-292, 1992.

[9] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nat.*, vol. 518, no. 7540, pp. 529-533, 2015.

[10] T. P. Lillicrap *et al.*, "Continuous control with deep reinforcement learning," *arXiv*, Sep. 2015, doi: 10.48550/arXiv.1509.02971.

[11] S. Fujimoto, H. Hoof, and D. Meger, "Addressing function approximation error in actor-critic methods," in *Int. Conf. Mach. Learn.*, Jul. 2018, pp. 1587-1596.

[12] R. Lowe, Y. I. Wu, A. Tamar, J. Harb, O. Pieter Abbeel, and I. Mordatch, "Multi-agent actor-critic for mixed cooperative-competitive environments," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.

[13] L. Raju, S. Sankar, and R. S. Milton, “Distributed optimization of solar micro-grid using multi agent reinforcement learning,” *Procedia Comput. Sci.*, vol. 46, pp. 231-239, 2015, doi: 10.1016/j.procs.2015.02.016.

[14] Y. Yang, Wang J. An overview of multi-agent reinforcement learning from game theoretical perspective[J]. arXiv, 2011.00583, 2020

[15] L. Xi, L. Yu, Y. Xu, S. Wang, and X. Chen, “A novel multi-agent DDQN-AD method-based distributed strategy for automatic generation control of integrated energy systems,” *IEEE Trans. Sustainable Energy*, vol. 11, no. 4, pp. 2417-2426, Oct. 2020, doi: 10.1109/tste.2019.2958361.

[16] R. Wang, X. Wen, X. Wang, Y. Fu, and Y. Zhang, “Low carbon optimal operation of integrated energy system based on carbon capture technology, LCA carbon emissions and ladder-type carbon trading,” *Appl. Energy*, vol. 311, p. 118664, Apr. 2022, doi: 10.1016/j.apenergy.2022.118664.

[17] X. Sun, Y. Zhang, Y. Zhang, J. Xie, and B. Sun, “Operation optimization of integrated energy system considering power-to-gas technology and carbon trading,” *Int. Trans. Electr. Energy Syst.*, vol. 2022, pp. 1-12, Mar. 2022, doi: 10.1155/2022/5026809.

[18] Y. Wang, J. Qiu, Y. Tao, and J. Zhao, “Carbon-oriented operational planning in coupled electricity and emission trading markets,” *IEEE Trans. Power Syst.*, vol. 35, no. 4, pp. 3145-3157, Jul. 2020, doi: 10.1109/tpwrs.2020.2966663.

[19] H. R. Sadeghian and M. M. Ardehali, “A novel approach for optimal economic dispatch scheduling of integrated combined heat and power systems for maximum economic profit and minimum environmental emissions based on Benders decomposition,” *Energy*, vol. 102, pp. 10-23, May 2016, doi: 10.1016/j.energy.2016.02.044.

[20] P. Stanislav, K. Bryan, and M. Tihomir, “Smart Grids better with integrated energy system,” in *2009 IEEE Electr. Power Energy Conf. (EPEC)*, Oct. 2009, doi: 10.1109/epec.2009.5420982.

[21] Z. Lu, X. Xu, Z.Yan, J. Wu, D. Sang, and S. Wang, “Overview on data-driven optimal scheduling methods of power system in uncertain environment,” *Autom. Electr. Power Syst.*, vol. 44, no. 21, pp. 172-184, 2020.

[22] C. Yang, W. Sun, D. Han, and K. Tan, “Distributionally-robust economic dispatch considering uncertain wind power output,” *Power Syst. Technol.*, vol. 44, no. 10, pp. 3649-3655, 2020.

[23] H. Jia, D. Wang, X. Xu, and X. Yu, “Research on some issues of regional integrated energy system,” *Autom. Electr. Power Syst.*, vol. 39, no. 7, pp. 198-207, 2015.

[24] Q. Ai and R. Hao, “Key technologies and challenges for multi-energy complementarity and optimization of integrated energy system,” *Autom. Electr. Power Syst.*, vol. 42, no. 4, pp. 2-10, 2018.

[25] Z. Li, F. Zhang, J. Liang, Z. Yun, and J. Zhang, “Optimization on microgrid with combined heat and power system,” *Proc. CSEE*, vol. 35, no. 14, pp. 3569-3576, 2015.

[26] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conf. Comput. Vision Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[27] M. Koziarski and B. Cyganek, “Image recognition with deep neural networks in presence of noise—Dealing with and taking advantage of distortions,” *Integr. Comput.-Aided Eng.*, vol. 24, no. 4, pp. 337-349, Sep. 2017, doi: 10.3233/ica-170551.

[28] S. Shubathra, P. Kalaivaani, and S. Santhoshkumar, “Clothing image recognition based on multiple features using deep neural networks,” in *2020 Int. Conf. Electron. Sustainable Commun. Syst. (ICESC)*, Jul. 2020, doi: 10.1109/icesc48915.2020.9155959