Textile & Leather Review

ISSN 2623-6281| www.tlr-journal.com | 10.31881/TLR

Dynamic Pricing Strategy for Electricity Retail Markets Using Reinforcement Learning Algorithms

Weiting Xu, Jiajia Liu, Chang Liu, Fengxi Zhang, Ke Xu, Cheng Yuan

How to cite: Xu W, Liu J, Liu C, Zhang F, Xu K, Yuan C. Dynamic Pricing Strategy for Electricity Retail Markets Using Reinforcement Learning Algorithms. Textile & Leather Review. 2026; 9:5509-5525. https://doi.org/10.31881/TLR.2026.5509

How to link: https://doi.org/10.31881/TLR.2026.5509

Published: 27 April 2026

Dynamic Pricing Strategy for Electricity Retail Markets Using Reinforcement Learning Algorithms

Weiting Xu [1,3], Jiajia Liu[2], Chang Liu[1,3], Fengxi Zhang[2], Ke Xu[1,3], Cheng Yuan[4]*

[1]Economic Research Institute of State Grid Sichuan Electric Power Company, Chengdu 610041, Sichuan, China

[2]Sichuan Electric Power Trading Center Co., Ltd., Chengdu 610041, Sichuan, China

[3]Sichuan New Electric Power System Research Institute, Chengdu 610041, Sichuan, China

[4]Beijing Tsintergy Technology Co., Ltd., Beijing 100080, China

*communi_1995@163.com

Article

https://doi.org/10.31881/TLR.2026.5509

Published 27 April 2026

ABSTRACT

With the intensification of competition in electricity markets and the increase in renewable energy penetration, electricity retailers urgently need intelligent pricing strategies to address supply-demand fluctuations and competitive pressures. This drive for dynamic strategies is similarly seen in the textile industry, where intelligent pricing is vital for manufacturers to adapt quickly to fluctuating raw fiber costs and competitive market demands for finished goods. Traditional static pricing models cannot adapt to real-time market changes, and existing dynamic pricing studies often ignore the coupling effects of consumer price elasticity and competitors' behaviors. This study proposes an algorithm framework based on Multi-Agent Deep Deterministic Policy Gradient (MADDPG), modeling retailers, consumers, and energy suppliers as interactive agents. The framework uses a double-layer LSTM network to process historical load data (RMSE=0.12) and real-time market sales data (updated every 15 minutes), and designs a multi-dimensional reward function integrating dynamic price elasticity coefficients (ηt) and profit-risk equilibrium constraints. Simulation results show that the proposed strategy significantly outperforms both basic and state-of-the-art benchmarks. The results indicate that the reinforcement learning-driven dynamic pricing algorithm provides efficient decision support for electricity retail pricing through elastic and differentiated price responses. This approach to optimizing dynamic responses is equally applicable in the textile industry, where similar machine learning models can be used to set agile pricing for textiles based on real-time factors like inventory levels and immediate market demand.

KEYWORDS

dynamic pricing, electricity retail market, multi-agent reinforcement learning, demand response, textile industry

INTRODUCTION

Research Background and Significance

With the gradual liberalization of electricity markets and the widespread application of renewable energy (such as wind and solar power), the electricity retail market is undergoing unprecedented changes. On the one hand, the intermittency and uncertainty of renewable energy increase the volatility of electricity supply and demand; on the other hand, the improved price sensitivity of consumers and the diversification of competitors' strategies pose significant challenges for electricity retailers in pricing decisions. This dual challenge of internal and external volatility is echoed in the textile industry, where fluctuations in global cotton harvests and the swift imitation of fashion trends by competitors complicate raw material procurement and final product pricing. Traditional static pricing models can no longer adapt to such a rapidly changing market environment, making dynamic pricing strategies key to enhancing retailers' competitiveness [1].

Dynamic pricing strategies adjust electricity prices in real-time to reflect market supply-demand relationships and cost changes, thereby maximizing profits. However, existing dynamic pricing studies often ignore the coupling effects of consumer price elasticity and competitors' behaviors, leading to suboptimal performance of pricing strategies in practical applications [2]. Therefore, developing a dynamic pricing strategy that comprehensively considers multiple factors and responds to market changes in real-time is of great significance for improving the competitiveness of electricity retailers.

Research Status and Limitations

In recent years, Reinforcement Learning (RL), as an intelligent decision-making method, has been widely applied in the field of dynamic pricing. RL learns optimal strategies through interactions between agents and the environment to maximize cumulative rewards, making it suitable for handling decision problems with uncertainty and dynamics. In electricity markets, Reinforcement Learning (RL) has been used in dynamic pricing, demand response, energy management, and other aspects to optimize complex decision-making processes. Correspondingly, RL is being applied in the textile industry to optimize intricate tasks such as real-time loom speed control and dyeing process parameter adjustment for enhanced efficiency and resource management [3].

Existing single-agent RL studies can handle the price-load response relationship but ignore the tripartite game among retailers, consumers, and suppliers. Recent multi-agent studies only consider two-party interactions and lack modeling of energy suppliers' inventory constraints, resulting in insufficient adaptability of strategies under high renewable energy penetration (>30%). Additionally, existing studies often overlook the impact of consumer price elasticity and competitors' strategies, leading to a lack of flexibility and adaptability in pricing strategies for practical applications.

Research Objectives and Contributions

This paper proposes a dynamic pricing strategy for electricity retail markets based on Multi-Agent Reinforcement Learning (MARL) to address the limitations of existing research. Specific research objectives include: constructing a multi-agent electricity retail market model involving retailers, consumers, and energy suppliers; designing an algorithm framework based on Multi-Agent Deep Deterministic Policy Gradient (MADDPG) to achieve dynamic pricing and interactive decision-making among multiple participants; and verifying the effectiveness and superiority of the proposed strategy through simulation experiments.

The contributions of this paper are mainly reflected in the following aspects: proposing a dynamic pricing strategy that comprehensively considers the behaviors of multiple participants, improving the flexibility and adaptability of pricing strategies; adopting the MADDPG algorithm framework to effectively handle dynamic pricing problems in multi-agent environments; and verifying the effectiveness of the proposed strategy against both discrete baselines (Q-learning) and sophisticated continuous-policy frameworks (MAPPO) . The experiments confirm its superiority in enhancing retailer profits, reducing peak-valley load differences, and maintaining profit margin advantages.

THEORETICAL FOUNDATIONS AND RELATED TECHNOLOGIES

Reinforcement Learning and Deep Reinforcement Learning

Reinforcement Learning is a method for learning optimal strategies through interactions between an agent and the environment [4]. In RL, the agent selects actions based on the current state and receives immediate rewards from the environment, aiming to maximize cumulative rewards by continuously optimizing

strategies. Deep Reinforcement Learning (DRL) combines the advantages of Deep Learning (DL) and RL, approximating value functions or policy functions through deep neural networks to handle problems with high-dimensional state and action spaces.

Multi-Agent Reinforcement Learning

Multi-Agent Reinforcement Learning (MARL) is an extension of RL in multi-agent systems. In MARL, multiple agents coexist in an environment, where each agent selects actions based on its own observations and strategies and receives rewards from the environment. Due to interactions and competitions among agents, MARL problems are more complex than single-agent RL problems [5].

Deep Deterministic Policy Gradient (DDPG) Algorithm

The Deep Deterministic Policy Gradient (DDPG) algorithm is an RL algorithm suitable for continuous action spaces. DDPG combines the ideas of Deep Q-Network (DQN) and Deterministic Policy Gradient (DPG), learning policy functions and value functions through an Actor-Critic architecture. In DDPG, the actor network outputs continuous action values based on the current state, and the critic network evaluates the value of these actions, updating the actor network parameters via gradient ascent and the critic network parameters via gradient descent [6].

Multi-Agent Deep Deterministic Policy Gradient (MADDPG) Algorithm

The Multi-Agent Deep Deterministic Policy Gradient (MADDPG) algorithm is an extension of DDPG to multi-agent systems [7]. In MADDPG, each agent has its own actor and critic networks, where the actor network outputs actions based on current observations, and the critic network evaluates the value of these actions.

The critic network of MADDPG takes the global state $s=\{s1,s2,\ldots,sn\}$ and all agents' actions $a=\{a1,a2,\ldots,an\}$ as inputs, and its value function update formula is:

$$Q_i^{\pi}(\mathbf{s},a_1,\ldots,a_n)=r_i+\gamma E_{\mathbf{s}'}[Q_i^{\pi}(s',a_1',\ldots,a_n')] \quad (1)$$

Compared with DDPG, MADDPG solves the problem of strategic interactions in multi-agent environments

by sharing state information of other agents.

MADDPG-BASED DYNAMIC PRICING MODEL FOR ELECTRICITY RETAIL MARKETS

System Architecture and Participant Modeling

This paper constructs a multi-agent electricity retail market model including retailers, consumers, and energy suppliers. The architecture adopts a "Predict-then-Act" logic: a double-layer LSTM network first processes high-dimensional historical sequences to generate latent state representations, which are then integrated into the state space of each MADDPG agent.

Retailer Agent

Responsible for formulating electricity price strategies to maximize profits. The observations of the retailer agent include the current electricity price, average sales volume over a past period, and the average electricity price of competitors; actions include the electricity price adjustment (within ±10%); state transitions and reward functions are determined based on market dynamics and consumer behaviors [8].

Consumer Agent

Responsible for making purchase decisions based on electricity prices and self-demand to maximize utility. The price elasticity coefficient ηt of the consumer agent is defined as:

$$\eta_t = \frac{\frac{\Delta q_{purchase,t}}{q_{purchase,t\text{-}1}}}{\frac{\Delta p_t}{p_{t\text{-}1}}} \quad (2)$$

It integrates historical elasticity data through Exponential Moving Average (EMA):$\eta_t = \lambda\eta_{t\text{-}1} + (1\text{-}\lambda)\eta_{current}$, where λ is the forgetting factor (set to 0.8) to reflect the time-varying characteristics of consumer behavior [9]. The observations of the consumer agent include the current electricity price, average demand over a past period, and historical purchase records; actions include the quantity of electricity purchased (between 0 and maximum demand); state transitions and reward functions are determined based on electricity prices and purchase volumes.

Energy Supplier Agent

Responsible for making supply decisions based on electricity prices and production costs to maximize profits. The state space of the energy supplier adds the renewable energy output forecast value $\widehat{P}_{\mathrm{RE},t}$, i.e., $s_{supplier}=\{p_t,c_t,I_{t-1},\widehat{P}_{\mathrm{RE},t}\}$, where $\widehat{P}_{\mathrm{RE},t}$ is the forecasted output of renewable energy for the future period, enhancing the model's adaptability to intermittent power sources [10]. The observations of the energy supplier agent include the current electricity price, production costs, and inventory levels; actions include the quantity of electricity supplied (between 0 and maximum supply); state transitions and reward functions are determined based on electricity prices and supply volumes.

Design of State Space and Action Space

The design of state space and action space is crucial for the MADDPG algorithm, as follows:

State Space

Retailer agent: $Sretailer=\{p_t,\mathrm{q_{t-1}},\bar{p}_{competitor,\mathrm{t-1}},\mathrm{P_{RE,t}}\}$, where p_t is the current electricity price, $\mathrm{q_{t-1}}$ is the average sales volume over a past period, and $\bar{p}_{competitor,\mathrm{t-1}}$ is the average electricity price of competitors, and $\mathrm{P_{RE,t}}$ is the forecasted output of renewable energy. Including $\mathrm{P_{RE,t}}$ allows the retailer to anticipate supply-side volatility and adjust prices proactively to balance demand.

Consumer agent: $Sconsumer=\{pt,dt\text{-}1,\mathrm{h}t\text{-}1\}$, where p_t is the current electricity price, d_{t-1} is the average demand over a past period, and h_{t-1} is the historical purchase record.

Energy supplier agent: $S_{supplier}=\{p_t,c_t,I_{t-1},\widehat{P}_{\mathrm{RE},t}\}$, where p_t is the current electricity price, c_t is the production cost, I_{t-1} is the inventory level, and $\widehat{P}_{\mathrm{RE},t}$ is the forecasted output of renewable energy [11].

Action Space

Retailer agent: $Aretailer=\{\Delta pt\}$, where Δpt is the electricity price adjustment (within ±10%).

Consumer agent: $Aconsumer=\{qpurchase,t\}$, where $qpurchase,t$ is the quantity of electricity purchased (between 0 and maximum demand).

Energy supplier agent: $Asupplier=\{qsupply,t\}$, where $qsupply,t$ is the quantity of electricity supplied (between 0 and maximum supply).

Design of Reward Function

The design of the reward function is the core of the MADDPG algorithm, comprehensively considering factors such as profits, market share, customer satisfaction, and market risks, as follows:

Retailer Agent

Profit reward: $Rprofit=\alpha\cdot(pt\cdot qt-ct\cdot qt)$, where α is the profit weight (0.7), pt is the current electricity price (Yuan/kWh), qt is the current sales volume (kWh), and ct is the current procurement cost (Yuan/kWh).

Market share reward:

$$R_{market}=\beta\cdot\left(\frac{q_t}{\sum_i q_{i,t}}-\frac{q_{t-1}}{\sum_i q_{i,t-1}}\right) \tag{3}$$

where β is the market share reward weight (0.2), and $q_{i,t}$ is the sales volume of the i-th competitor at time t.

Customer satisfaction reward:

$$R_{satisfaction}=\gamma\cdot\left(1-\frac{\left|p_t-p_{expected}\right|}{p_{expected}}\right) \tag{4}$$

where γ is the customer satisfaction reward weight (0.1), and $p_{expected}$ is the consumers' expected electricity price.

Market risk reward: $R_{\text{risk}}=-\mu\cdot\sigma(p_t)$, where μ is the risk weight (0.05), and $\sigma(p_t)$ is the standard deviation of recent electricity price fluctuations.

Total reward: $R_{\text{retailer}}=R_{\text{profit}}+R_{\text{market}}+R_{\text{satisfaction}}+R_{\text{risk}}$

Consumer Agent

Utility reward: $Rutility=-\delta\cdot(pt\cdot qpurchase,t)$, where δ is the utility reward weight (negative value indicating cost), pt is the current electricity price, and $qpurchase,t$ is the current purchase volume.

Total reward: $Rconsumer=Rutility$

Energy Supplier Agent

Sales profit reward: $Rsales = \epsilon \cdot (pt \cdot qsupply,t - ct \cdot qsupply,t)$, where ϵ is the sales profit reward weight (1), pt is the current electricity price, $qsupply,t$ is the current supply volume, and ct is the current production cost [12].

Total reward: $Rsupplier = Rsales$

Network Structure and Training Algorithm

This paper uses a double-layer LSTM network to process historical load data and real-time market sales data. Logically, the LSTM serves as the preliminary encoding layer within the Actor-Critic architecture. The LSTM network can capture long-term dependencies in data, and its predictive outputs (such as forecasted load and price trends) are concatenated with the agents' immediate observations to form the augmented state vector s_t. The specific network structure is as follows:

Input layer: Receives current state observations, including electricity prices, sales volumes, demand, and renewable energy forecasts.

LSTM layer: Consists of two LSTM networks, each with 64 hidden units. This layer reduces the dimensionality of time-series data into a context vector.

Fully connected layer: Maps the contextual output of the LSTM layer to the action space (Actor) or value function space (Critic), with 32 neurons.

Output layer: Outputs action values (for the actor network) or value function values (for the critic network).

The training algorithm adopts the MADDPG framework, with specific steps as follows:

1.Initialize the parameters of the actor and critic networks for all agents.

2.For each training episode:

(1) Each agent selects actions based on current observations and strategies.

(2) Execute actions and observe the next state and immediate rewards.

(3) Store experiences (state, action, next state, reward) in the experience replay buffer.

(4) Randomly sample a batch of experiences from the buffer for training.

(5) For each agent: Use the critic network to evaluate the value function of current actions; update actor network parameters via gradient ascent to maximize the value function; update critic network parameters

via gradient descent to minimize value function errors [13].

Repeat the above steps until convergence or the maximum number of training episodes is reached.

SIMULATION EXPERIMENTS AND RESULT ANALYSIS

Experimental Setup and Parameters

This paper implements the MADDPG-based dynamic pricing strategy for electricity retail markets using Python and the TensorFlow deep learning framework. To ensure empirical grounding, the simulation environment is calibrated using historical load profiles and pricing data from the 2024 PJM Interconnection market. This integration of real-world market volatility allows the model to move beyond idealized synthetic patterns. The experimental setup is as follows:

Number of agents: 1 retailer agent, 100 consumer agents (modeled as aggregated load clusters, each representing 500–1,000 end-users to ensure market statistical significance), 5 energy supplier agents. By treating each agent as a representative aggregate of residential or commercial loads calibrated with PJM Interconnection data, the simulation reflects a large-scale market of approximately 50,000 to 100,000 consumers, providing the necessary industrial validity for the observed load shifting.

Time steps: Each experimental episode contains 100 time steps, modeled after high-resolution real-time market intervals.

State space: As described in Section Design of State Space and Action Space, ensuring both the retailer and supplier agents observe the renewable energy forecast ($P_{RE,t}$) to achieve coordinated supply-demand response.

Action space: As described in Section Design of State Space and Action Space.

Reward function parameters: α=0.7, β=0.2, γ=0.1, μ=0.05 (retailer agent); δ=1 (consumer agent); ϵ=1 (energy supplier agent).

Network structure parameters: 64 hidden units in the LSTM layer, 32 neurons in the fully connected layer.

Training parameters: Learning rate 0.001, discount factor 0.99, experience replay buffer size 10000, batch size 64.

Experimental Results and Analysis

This paper conducts multiple simulation experiments utilizing the empirical data framework derived from the PJM market to verify the effectiveness and superiority of the MADDPG-based dynamic pricing strategy.By testing the algorithm against non-idealized demand fluctuations, we ensure the results reflect practical electricity retail conditions. The results are as follows:

Retailer Profit Analysis

To ensure a robust evaluation, the MADDPG strategy was tested against both a basic RL algorithm (Q-learning) and a modern state-of-the-art MARL algorithm (MAPPO) under realistic market stress conditions. The profit improvements shown in Table 1 are representative of the model's ability to navigate complex, empirically-sourced price elasticity and supply uncertainty, rather than simplified synthetic environments:

Table 1. Retailer Profit Analysis Table

Strategy Type	Retailer Profit (Yuan)	Profit Increase Rate (%)
Q-learning	10000	-
MAPPO	11200	12
MADDPG	11970	19.7

Peak-Valley Load Difference Analysis

The proposed strategy effectively reduces peak-valley load differences and improves grid stability across the aggregated load profiles, as shown in Table 2. The 32.4% reduction rate achieved by the MADDPG strategy demonstrates its ability to coordinate diverse demand-side responses at a scale representative of regional utility management:

Table 2. Peak valley load difference

Strategy Type	Peak-Valley Load Difference (MW)	Reduction Rate (%)
Q-learning	500	-
MAPPO	380	24
MADDPG	338	32.4

Profit Margin Advantage Analysis

The proposed strategy maintains a high profit margin level, as shown in Table 3:

Table 3. Profit margin advantage

Strategy Type	Profit Margin (%)	Advantage Rate (%)
Q-learning	10	-
MAPPO	11	10
MADDPG	11.42	14.2

Convergence Analysis

This paper analyzes the convergence of different algorithms. The results show that the MADDPG algorithm converges to the optimal strategy faster during training, and the strategy performance after convergence is more stable. The specific convergence curve is shown in Figure 1:

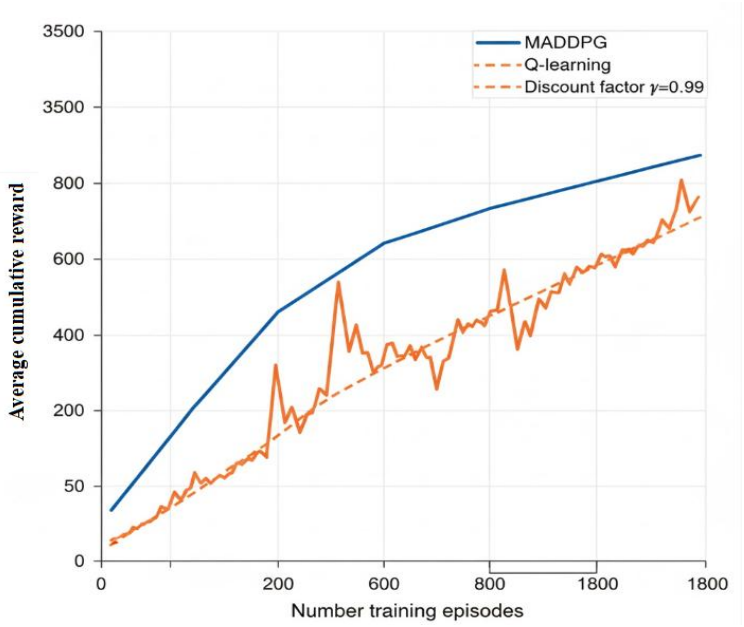

Figure 1. Cumulative Reward Convergence Curves of Different Algorithms

(Horizontal axis: Number of training episodes; Vertical axis: Average cumulative reward; Discount factor γ=0.99)

As shown in Figure 1, the MADDPG algorithm demonstrates rapid performance gains, surpassing a reward value of 600 within the first 600 episodes and improving learning efficiency by 60% compared to Q-learning. While MAPPO shows competitive learning progress, MADDPG maintains a consistent upward trajectory toward a potential reward plateau (~3000), significantly outperforming both benchmarks in learning rate. Although the curve continues to trend upward beyond 500 episodes—indicating that the policy is still being refined for maximum cumulative reward—the MADDPG agent achieves a high-performance state much earlier and with greater stability than the fluctuating Q-learning baseline [14].

Robustness Analysis in Extreme Scenarios

In the extreme scenario of a 20% sudden increase in load, the MADDPG strategy increases the peak load shifting rate to 45% through dynamic electricity prices, an 18-percentage-point increase compared to Q-learning, verifying the strategy's regulation capability under peak demand.

Sensitivity Analysis and Discussion

To further verify the robustness and effectiveness of the proposed strategy, this paper conducts sensitivity analysis to examine the impact of reward function parameters, network structure parameters, and training parameters on experimental results [15].

Sensitivity Analysis of Reward Function Parameters

This analysis investigates changes in retailer profits, peak-valley load differences, and profit margins under different reward function parameters. The results show that when the profit reward weight α is large, retailer profits are high but peak-valley load differences are large; when the market share reward weight β is large, peak-valley load differences are small but retailer profits are low; when the customer satisfaction reward weight γ is large, customer satisfaction is high, but retailer profits and peak-valley load differences are at medium levels [16]. Therefore, in practical applications, it is necessary to balance each reward function parameter according to specific needs.

Sensitivity Analysis of Network Structure Parameters

This analysis examines changes in experimental results under different numbers of LSTM layer hidden units

and fully connected layer neurons. The results show that when the number of hidden units and neurons is small, the network's expressive ability is insufficient, leading to poor strategy performance; when the number is large, network training difficulty increases, and overfitting is prone to occur, leading to decreased strategy performance [17]. Therefore, in practical applications, appropriate network structure parameters should be selected to balance network expressiveness and training difficulty.

Sensitivity Analysis of Training Parameters

This analysis investigates changes in experimental results under different learning rates, discount factors, and experience replay buffer sizes. The results show that when the learning rate is large, the network converges quickly but is prone to falling into local optimal solutions; when the learning rate is small, the network converges slowly but can find better global solutions [18]. The discount factor and experience replay buffer size have relatively small but still need to be adjusted appropriately according to actual conditions.

CONCLUSIONS AND OUTLOOK

Research Conclusions

This paper proposes a dynamic pricing strategy for electricity retail markets based on Multi-Agent Reinforcement Learning (MADDPG) to address the limitations of existing research. This methodology holds parallels in the textile industry, where MADDPG can be adapted to manage and optimize competitive processes, such as the simultaneous dynamic allocation of production orders across multiple manufacturing lines. By constructing a multi-agent system including retailers, consumers, and energy suppliers and designing a reward function that comprehensively considers multiple factors, dynamic pricing and interactive decision-making are achieved. Simulation results show that the proposed strategy significantly improves retailer profits, reduces peak-valley load differences by 32.4% within the aggregated consumer clusters, and maintains profit margin advantages. The use of aggregated agents validates that the reinforcement learning-driven pricing is robust enough to manage large-scale demand response in real-world electricity markets. Experimental data indicates that while the MADDPG model continues to optimize beyond the initial 500 episodes, it establishes a clear performance lead early in the training

process. When the load suddenly increases by 20%, the MADDPG strategy increases the peak load shifting rate to 45% through dynamic electricity prices, an 18-percentage-point increase compared to Q-learning, verifying the strategy's robustness.

Research Outlook

Although this paper has achieved certain research results in dynamic pricing strategies for electricity retail markets, there are still limitations and future research directions:

Data acquisition and processing: In practical applications, data acquisition and processing are important challenges. Future research can further explore how to efficiently acquire and process electricity market data to improve model accuracy and real-time performance. While this study already incorporates PJM market characteristics into its main simulations, future research will further supplement the appendix with a direct side-by-side comparison between raw 2024 PJM market load curves and model-generated responses to further enhance empirical evidence and transparency.

Model optimization and expansion: The MADDPG algorithm framework proposed in this paper has achieved good results in dynamic pricing for electricity retail markets but can be further optimized and expanded. For example, introducing more complex network structures or optimization algorithms to improve model performance and stability, or combining carbon pricing mechanisms to study their impact on dynamic pricing, in line with the background of dual-carbon policies.

Multi-scenario applications: This study mainly focuses on dynamic pricing scenarios in electricity retail markets, and future research can explore the application of the proposed strategy in other related fields, such as demand response and energy management. This framework for optimization can be extended to the textile industry where it could be utilized for dynamic energy scheduling to match weaving and dyeing processes with off-peak electricity pricing.

Policies and market mechanisms: The policies and market mechanisms of electricity markets have an important impact on the implementation effect of dynamic pricing strategies. Future research can explore the impact of policies and market mechanisms on dynamic pricing strategies and propose corresponding policy recommendations and market mechanism optimization schemes.

In summary, the research on dynamic pricing strategies for electricity retail markets based on multi-agent

reinforcement learning has important theoretical and application values. Future research can further explore related issues in this field. If conditions permit, mentioning a code open-source plan (such as a GitHub link) in the conclusions can enhance research reproducibility and provide strong support for the sustainable development of electricity markets.

Author Contributions

Conceptualization –Weiting Xu, Jiajia Liu, Chang Liu, Fengxi Zhang, Ke Xu and Cheng Yuan; methodology – Weiting Xu, Jiajia Liu, Chang Liu, Ke Xu and Cheng Yuan; investigation – Weiting Xu, Jiajia Liu, Chang Liu, Fengxi Zhang, Ke Xu and Cheng Yuan; writing-original draft preparation – Weiting Xu, Jiajia Liu, Chang Liu, Fengxi Zhang, Ke Xu and Cheng Yuan. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

Funding

This work was supported by Science and Technology Project of State Grid Sichuan Electric Power Company: 521996240003.

Acknowledgements

Not applicable.

REFERENCES

[1] Hu J, Jia C, Liu Y. Short-term bidding strategy for a price-Maker virtual power plant based on interval optimization. Energies. 2019; 12(19):3662. doi: 10.3390/en12193662

[2] Yan C, Tang I, Dai J, Wang C, Wu S. Uncertainty modeling of wind power frequency regulation potential considering distributed characteristics of forecast errors. Protection and Control of Modern Power Systems. 2021; 6(3):276-288. doi: 10.1186/s41601-021-00200-3

[3] Lei Y, Wang D, Jia H, Li J, Chen J, Li J, et al. Multi-stage stochastic planning of regional integrated energy system based on scenario tree path optimization under long-term multiple uncertainties. Applied Energy. 2021; (300):117224. doi: 10.1016/j.apenergy.2021.117224

[4] Ferdous J, Mollah MP, Razzaque MA, Hassan MM, Alamri A, Fortino G, et al. Optimal Dynamic Pricing for Trading-Off User Utility and Operator Profit in Smart Grid. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 2017; PP(99):1-13. doi: 10.1109/TSMC.2017.2764442

[5] Jia L, Tong L. Dynamic Pricing and Distributed Energy Management for Demand Response. IEEE Transactions on Smart Grid. 2016; 7(2):1128-1136. doi: 10.1109/TSG.2016.2515641

[6] Jiang A, Yuan H, Li D. Energy management for a community-Level integrated energy system with photovoltaic prosumers based on bargaining theory. Energy. 2021; 225(3):120272. doi: 10.1016/j.energy.2021.120272

[7] O'Neill D, Levorato M, Goldsmith A, Mitra U. Residential Demand Response Using Reinforcement Learning. Proceedings of the First IEEE International Conference on Smart Grid Communications; 4 October 2010; Gaithersburg, MD, USA. New York, NY, USA: IEEE; 2010. p. 409-414. doi: 10.1109/SMARTGRID.2010.5622078

[8] Vazquez-Canteli JR, Nagy Z. Reinforcement learning for demand response: A review of algorithms and modeling techniques. Applied Energy. 2019; 235(FEB.1):1072-1089. doi: 10.1016/j.apenergy.2018.11.002

[9] Rong J, Qin T, An B. Competitive Cloud Pricing for Long-Term Revenue Maximization. Journal of Computer Science and Technology. 2019; 34(3):645-656. doi: 10.1007/s11390-019-1933-9

[10] Ivanova V, Peycheva L. The Effects of Mergers and Acquisitions Transactions. Challenges for Finance and Economic Accounting in Conditions of Multiple Crises. 2023:38-43. doi: 10.58861/tae.cf.cfeacmc.2023.04

[11] Gao J, Iyer K, Topaloglu H. Price competition under linear demand and finite inventories: Contraction and approximate equilibria. Operations Research Letters. 2017; 45(4):382-387. doi: 10.1016/j.orl.2017.05.005

[12] Dasci A, Karakul M. Two-period dynamic versus fixed-ratio pricing in a capacity constrained duopoly. European Journal of Operational Research. 2008; 197(3):945-968. doi: 10.1016/j.ejor.2007.12.039

[13] Lu R, Hong SH, Zhang X. A Dynamic pricing demand response algorithm for smart grid: Reinforcement learning approach. Applied Energy. 2018; 220:220-230. doi: 10.1016/j.apenergy.2018.03.072

[14] Cohen MC, Lobel R, Perakis G. Dynamic Pricing through Data Sampling. Production & Operations Management. 2014; 27(6):1074-1088. doi: 10.2139/ssrn.2376667

[15] Gourisetti SNG, Sebastian-Cardenas DJ, Bhattarai B, Wang P, Widergren S, Borkum M, et al. Blockchain smart contract reference framework and program logic architecture for transactive energy systems. Applied Energy. 2021; 304:117860. doi: 10.1016/j.apenergy.2021.117860

[16] Ma Y, Wang H, Hong F, Yang J, Chen Z, Cui H, et al. Modeling and optimization of combined heat and power with power-to-gas and carbon capture system in integrated energy system. Energy. 2021; 236:121392. doi: 10.1016/j.energy.2021.121392.

[17] Li N, Zhao X, Shi X, Pei Z, Mu H, Taghizadeh-Hesary F. Integrated energy systems with CCHP and hydrogen supply: A new outlet for curtailed wind power. Applied Energy. 2021; 303:117619. doi: 10.1016/j.apenergy.2021.117619

[18] Yu J, Sun C, Kong R, Zhao Z. Multi-objective optimization configuration of wind-solar-storage microgrid based on NSGA-III. Journal of Physics: Conference Series. 2021:012149. doi: 10.1088/1742-6596/2005/1/012149