IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 10 April 2024, accepted 22 May 2024, date of publication 28 May 2024, date of current version 4 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3406338

RESEARCH ARTICLE

Dynamic Demand-Aware Power Grid Intelligent Pricing Algorithm Based on Deep Reinforcement Learning

CHAO TANG, YUNCHUAN QIN, FAN WU, AND ZHUO TANG, (Member, IEEE)
College of Computer Science and Electronic Engineering, Hunan University, Changsha, Hunan 421082, China
Corresponding author: Zhuo Tang (ztang@hnu.edu.cn)

ABSTRACT The increasing integration of renewable energy sources and the growing complexity of modern power grids demand innovative solutions for efficient energy management. This paper introduces a novel dynamic demand-aware Power Grid Intelligent Pricing (PGIP) algorithm based on Deep Reinforcement Learning (DRL). The proposed PGIP algorithm aims to optimize energy consumption and pricing in real time by leveraging the capabilities of DRL to adapt to dynamic demand patterns and evolving grid conditions. PGIP employs a sophisticated neural network architecture to model the intricate relationships between various grid parameters, user demand, and pricing strategies. Through continuous learning and adaptation, the algorithm dynamically adjusts pricing structures to incentivize demand-side flexibility while ensuring grid stability. The reinforcement learning framework enables the algorithm to discover optimal policies for pricing in response to changing environmental factors and user behaviors. We used real-world data sets to assess its performance in diverse scenarios. Results demonstrate the algorithm's ability to optimize energy consumption, reduce peak demand, and enhance overall grid efficiency. Moreover, comparisons with traditional pricing models highlight the superior adaptability and responsiveness of PGIP in addressing the challenges posed by the evolving landscape of power grids. PGIP presents a promising approach to address the dynamic nature of power grids and the increasing demand for efficient energy management.

INDEX TERMS Deep reinforcement learning, dynamic demand, intelligent pricing, power grid.

I. INTRODUCTION

The current power grid landscape is significantly transforming, propelled by the escalating assimilation of renewable energy sources, advancements in smart grid technologies, and a rising demand for sustainable and efficient energy management solutions [1], [2]. Conventional power grid structures struggle to adjust to the volatile nature of energy production and consumption patterns, and the necessity for real-time optimization [3]. Given these challenges, there is an urgent requirement for innovative algorithms capable of intelligently managing energy demand, optimizing pricing strategies, and improving overall grid efficiency.

A pivotal factor affecting the performance of power grids is the demand-side consumer behavior [4], [5]. Traditional pricing models frequently lack the adaptability necessary to encourage consumers to relocate their energy usage to off-peak times or curtail consumption during high-demand periods [6]. Additionally, the incorporation of renewable energy sources incurs variability, making it crucial for the power grid to dynamically modify its operations.

The shifting dynamics of power grid management have sparked comprehensive research into intelligent algorithms and pricing strategies designed to confront the challenges of dynamic demand patterns, integration of renewable energy, and the necessity for real-time adaptability [7]. Existing literature demonstrates a myriad of strategies geared towards optimizing energy consumption, pricing structures, and grid stability [8]. Current solutions underscore a range of efforts to improve the efficiency and adaptability of power grid management [9]. However, this work aims to augment the existing knowledge base. It focuses principally on applying deep reinforcement learning for dynamic pricing in reaction to changing demand patterns, thus targeting a more responsive and sustainable power grid.

The associate editor coordinating the review of this manuscript and approving it for publication was Zhiyi Li.

Driven by these challenges, we propose a dynamic, demand-aware Power Grid Intelligent Pricing (PGIP) algorithm based on Deep Reinforcement Learning (DRL) in this paper. The rationale for this research is to exploit the potential of deep reinforcement learning to tackle the intricate and evolving dynamics of power grids. By utilizing the capabilities of DRL, the algorithm can determine optimal pricing strategies in real time, adapting to fluctuating demand patterns, grid conditions, and the increasing dominance of renewable energy sources.

This work's significance transcends are theoretical advancements, seeking to offer pragmatic solutions for power grid operators, utility companies, and policymakers. The proposed DDPG-IPA aims to amplify demand-side flexibility, mitigate peak demand, and enhance the overall stability and efficiency of power grids. Aspiring to contribute to the ongoing efforts in constructing smarter, more robust, and sustainable power grids for the future, the algorithm employs a data-driven approach and advanced machine-learning techniques. Results from this work hold the potential to inform policy decisions, direct the development of intelligent grid management systems, and lay the foundation for a more adaptive and responsive energy infrastructure. Our contributions comprise:

- An Innovative Power Grid Pricing Algorithm. We design an original PGIP algorithm catered to resolve the challenges stemming from the amplified integration of renewable energy sources and the escalating complexity of today's power grids. This algorithm employs DRL to adapt dynamically to real-time alterations in demand patterns and evolving grid conditions.
- An Advanced Neural Network Architecture for Grid Parameter Modeling. We integrate an advanced neural network architecture to model the complex relationships among various grid parameters, user demand, and pricing strategies. This addition enhances the algorithm's ability to capture the intricate dynamics of modern power grids accurately.
- Real-World Performance Evaluation and Superior Adaptability. To verify the effectiveness of PGIP, we utilize real-world data sets to conduct comprehensive performance evaluations across a myriad of scenarios. The results showcase the algorithm's exceptional ability to optimize energy consumption, lower peak demand, and bolster overall grid efficiency.

The remainder of the paper is organized as follows: Section II briefly discusses related work in this field. Section III describes the system model, introducing the electricity consumer and service provider models. Section IV presents our proposed dynamic demand-aware Power Grid Intelligent Pricing algorithm based on deep reinforcement learning. In Section V, we discuss the comparison experiments and their results. Finally, Section VI concludes this work and outlines potential directions for future research.

II. RELATED WORK

The changing dynamics of power grid administration have spurred in-depth exploration into smart algorithms and pricing tactics to tackle the complexities presented by fluctuating demand trends, the integration of renewable energy, and the imperative for immediate adaptability. Recent literature showcases a varied array of methods focused on enhancing energy usage efficiency, pricing frameworks, and the stability of the grid.

Traditional demand response methods involve incentivizing consumers to adjust their energy consumption in response to price signals. Time-of-use pricing and peak/off-peak pricing have been widely explored [10]. For example, in [11], Hung and Michailidis examined the attributes of power demand observed in actual grids to represent it as a steady level with variations using a scaled fractional Brownian motion during uniform peak periods. In [12], Yang et al. outlined a game-theoretic method to enhance the optimization of pricing strategies based on Time-of-use. In [13], Wesseh and Lin suggested an electricity market pricing model based on time-of-use (TOU) that can effectively represent the interplay among power plants, generation ramping, storage devices, electric vehicle charging, and fluctuations in electricity prices. While effective to some extent, these approaches often lack the adaptability required for real-time optimization and may not fully exploit the potential of emerging technologies.

Several studies have explored the application of Machine Learning (ML) techniques for demand forecasting. Predictive models, such as neural networks and ensemble methods, aim to forecast energy demand with high accuracy [14], [15]. For instance, in [16], Real et al. designed a blended framework comprising a convolutional neural network (CNN) paired with an artificial neural network. The primary goal is to leverage the strengths of both architectures: utilizing ANN's regression capabilities and tapping into CNN's feature extraction capacities. In [17], Almaghrebi et al. used three ML techniques to forecast the charging demand for Plug-in Electric Vehicle (PEV) users once a charging session commences. In [18], Haq et al. introduced a hybrid approach utilizing ML to predict appliance consumption and peak demand. The methodology involves the implementation of accelerated k-medoids clustering, support vector machine, and artificial neural network for forecasting both appliance consumption and peak demand among customers. However, these approaches often operate statically and may struggle to adapt to sudden changes in demand or grid conditions.

The integration of reinforcement learning (RL) in energy systems has gained attention for its ability to optimize decision-making in dynamic environments. RL algorithms, including deep reinforcement learning (DRL), have been applied to various aspects of power grid management [19]. For example, in [20], Li et al. proposed a prioritized experience replay automated RL (PER-AutoRL) to streamline the implementation of a customized DRL-based forecasting model. To tackle the uncertainties present in grid networks, Bahrami et al. [21] created a DRL algorithm employing an actor-critic method. In [22], the Q-learning methods were applied to predict the plug-in hybrid electric vehicle (PHEV) loads under different scenarios. However, the specific focus on dynamic demand-aware pricing algorithms for real-time grid optimization is an area that warrants further exploration.

Research on intelligent grid management systems emphasizes the development of comprehensive solutions that combine data analytics, optimization algorithms, and smart grid technologies. These systems often integrate ML components for adaptive decision-making. For instance, in [23], Aurangzeb et al. designed an equitable pricing scheme grounded in power demand prediction to minimize additional costs for consumers with low energy consumption. In [24], Rasheed et al. explored a comprehensive framework for constructing models for electricity retail pricing, utilizing load demand and market price data. The aim is to minimize the average system cost and mitigate rebound peaks through the incorporation of energy procurement prices, load scheduling, and the integration of renewable energy sources. However, the incorporation of deep reinforcement learning specifically for demand-aware pricing remains a relatively underexplored aspect. With the increasing share of renewable energy sources, researchers have investigated strategies for effectively integrating these sources into the grid [25]. Dynamic pricing mechanisms that incentivize consumption during periods of high renewable energy availability and discourage consumption during low availability have been explored.

III. PROPOSED METHOD

In this section, we propose a dynamic demand-aware Power Grid Intelligent Pricing (PGIP) algorithm that integrates strategies based on pricing and incentives, aiming to encourage consumers to alter their electricity consumption patterns while considering satisfaction with consumption. The objective is to establish an optimal pricing and incentive structure within the context of contemporary markets. Two markets are considered in our work, including a wholesale market and an adjustment market, as shown in Fig. 1.

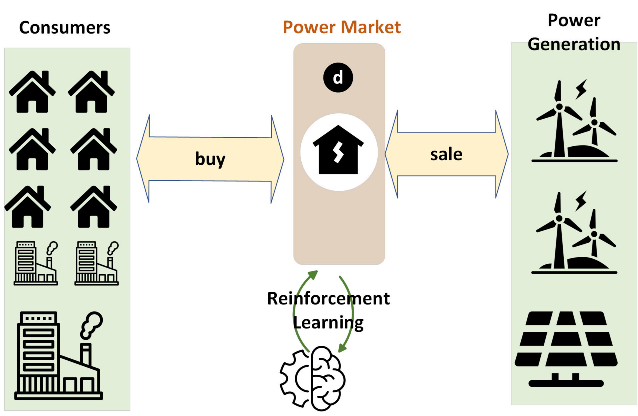


FIGURE 1. Market models considered in this work.

We use reinforcement learning techniques to increase the adoption of Renewable Energy Sources (RES) by adding flexibility to the power grid and decreasing uncertainty in long-term planning. This involves integrating signals from the wholesale market into pricing and incentive strategies, by optimizing their transmission to consumers. The primary objective is to achieve effective management of electrical energy, allowing consumers to adapt to emerging system components like electric mobility and demand aggregators, while maximizing the utilization of available RES.

A. MULTI-PERIOD HIERARCHICAL ELECTRICITY PRICE MODEL

We construct a multi-period hierarchical electricity price model $P = \{P^m, P^w, P^d\}$, where P^m represents the electricity price set of one month, P^w represents the electricity price set of one week, and P^d represents the electricity price set of one day. In addition, each day's electricity price set consists of hourly electricity prices for 24 hours, that is,

$$P_i^d = \{P_{i,1}, P_{i,2}, \ldots, P_{i,t}, \ldots, P_{i,24}\}. \tag{1}$$

Similarly, each week's electronic price set consists of the prices of each day of the week, that is,

$$\begin{aligned} P_j^w = \{&P_1^d(\text{Sun}), P_2^d(\text{Mon}), P_3^d(\text{Tue}), \\ &P_4^d(\text{Wed}), P_5^d(\text{Thu}), P_6^d(\text{Fri}), P_7^d(\text{Sat})\}. \end{aligned} \tag{2}$$

The price for each month consists of the price for each day of the month. $P_k^m = \{P_{k,1}^d, P_{k,2}^d, \ldots\}$. An example of the proposed multi-period hierarchical electricity price model is illustrated in Figure 2.

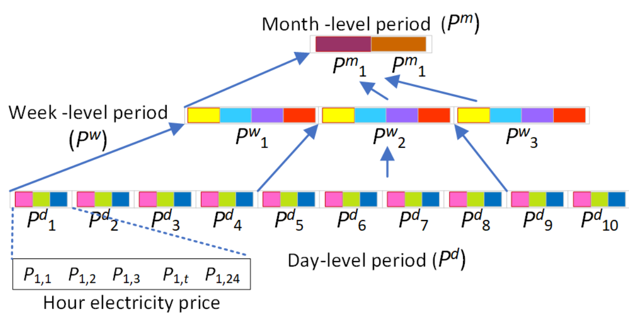


FIGURE 2. Multi-period hierarchical electricity price model.

B. ELECTRICITY CONSUMER MODEL

We establish the consumption behavior and income model of electricity consumers to provide theoretical support for electricity price prediction. The electricity consumption behavior and benefit model is a theory that describes how individuals, businesses, or social organizations use electricity and obtain benefits from it. This model covers the behavior of electricity consumers, the services of electricity suppliers, and the working mechanism of energy markets. The model mainly includes consumer consumption behavior, which consists of consumer usage needs and renewable energy preferences, and consumer benefits, which consist of cost savings and environmental benefits.

1) CONSUMER USAGE NEEDS

The electricity demand pattern of individuals or organizations refers to their demand characteristics for electricity in different periods and different scenarios. For example, electricity demand may be different during the day and night, and electricity usage patterns may also be different during workdays and rest days. The electricity usage demand of a certain consumer u_a in each hour on the i-th day can be defined as:

$$R_i(u_a) = \{R_{i,1}(u_a), \ldots, R_{i,24}(u_a)\}. \tag{3}$$

2) RENEWABLE ENERGY PREFERENCE

Some electricity consumers may prefer to choose electricity from renewable sources due to environmental or sustainability considerations. Suppose there is a set of renewable energy power options $G = \{g_1, \ldots, g_N\}$, a certain consumer u_a's preference for each renewable energy power is set to $\gamma(u_a) = \{\gamma_{u_a,g_1}, \ldots\}$.

3) COST SAVINGS

By optimizing energy usage and selecting appropriate power supply contracts, electricity consumers can achieve cost savings. Since the price of electricity may be different every hour, the user's cost savings in one day is the sum of the cost savings in 24 hours that day. For each hour, the cost savings for an hour is the difference between the highest market price in that hour and the user's actual purchase price multiplied by the actual power usage. The cost saving of the certain consumer u_a in the i-the day is defined as $\mathrm{CS}_i(u_a)$:

$$\mathrm{CS}_i(u_a) = \sum_{t=1}^{24} c_{i,t} \times \left(max(P^d_{i,t}) - P^d_{i,t}(u_a)\right), \tag{4}$$

where $c_{i,t}$ is the actual power usage of u_a in the period of t-th hour, $max(P^d_{i,t})$ is the highest market price in that hour, and $P^d_{i,t}(u_a)$ is the user's actual purchase price.

4) ENVIRONMENTAL BENEFITS

Choosing renewable energy or taking energy-saving measures can reduce carbon emissions and reduce the impact on the environment, thereby obtaining environmental benefits. The environmental benefit is calculated by the user's use of renewable energy multiplied by unit energy-saving income. The environmental benefits of the certain consumer u_a in the i-the day is defined as $\mathrm{EB}_i(u_a)$:

$$\mathrm{EB}_i(u_a) = \sum_{t=1}^{24} (c_{i,t} \times \varrho_e), \tag{5}$$

where ϱ_e is the unit energy-saving income.

C. SERVICE PROVIDER MODEL

The behavioral and revenue model of electricity providers involves how they produce, distribute, and sell electricity and derive profits from these activities. Below are the key elements of electricity provider behavior and revenue models, including supply contracts and revenue models. Among them, the supply contract consists of a fixed-price contract and a floating-price contract. The revenue model consists of sales revenue and renewable energy subsidies.

1) ELECTRICITY SUPPLY CONTRACT

Electricity supply contracts include fixed-price contracts and floating-price contracts. In a fixed-price contract, providers have stable electricity prices, and consumers can lock in electricity costs based on the price within the contract period. In a floating-price contract, the price of each provider is adjusted according to market price fluctuations, consumers may face price risks, but they also have the opportunity to gain price advantages.

2) SALES REVENUE

Revenue is obtained from selling electricity and providing related services. Since the price of electricity may be different every hour, the supplier's sales revenue in a day is the sum of the sales revenue in the 24 hours that day. For each electricity supplier s_a, its sales revenue on the i-th day is calculated as follows:

$$\mathrm{SR}_i(s_a) = \sum_{t=1}^{24} (s_{i,t} \times P^d_{i,t}(s_a)), \tag{6}$$

where $s_{i,t}$ is the quantity of electricity sale in the t-th hour in the i-th day and $P^d_{i,t}(s_a)$ is the price of s_a provided in that hour.

3) SUBSIDIES AND INCENTIVES

The government may encourage the development of renewable energy through subsidies and incentives to increase the motivation of power providers to obtain revenue from renewable energy. Therefore, if a supplier sells electricity from renewable sources, he will receive corresponding subsidies. For each renewable energy source in a set of renewable energy power options $G = \{g_1, \ldots, g_N\}$, assume that the government subsidies are $S(G) = \{s_1, \ldots, s_N\}$. Assume that supplier U_a sells renewable energy g1, and the quantity sold per hour is $v_{i,t}$, then the renewable energy subsidy it can obtain is:

$$\mathrm{ES}_i(s_a) = \sum_{t=1}^{24} (s_{i,t} \times \varpi_{g_j}), \tag{7}$$

where ϖ_{g_j} is the unit renewable energy subsidy provided by governments.

D. OBJECTIVE DEFINITION

Based on the periodic electricity price model, consumer behavior and revenue model, and supplier revenue model, we propose the objective function of this work. For consumers, the optimization goal is to maximize consumption benefits and minimize power purchase costs.

$$\max \sum_{u_a \in U} \sum_{t \in (1,24)} (\alpha \mathrm{CS}_i(u_a) + \beta \mathrm{CS}_i(u_a)), \tag{8}$$

where the role of coefficients α and β is to control the importance ratio between cost savings and environmental benefits.

$$\min \sum_{u_a \in U} \sum_{t \in [1,24]} (c_{i,t} \times P^d_{i,t}(u_a)), \tag{9}$$

where $c_{i,t}$ is the actual power usage of u_a in the period of t-th hour and $P^d_{i,t}(u_a)$ is the user's actual purchase price. For service providers, the optimization goal is to maximize revenue.

$$\max \sum_{s_a \in S} SR_i(s_a) + \mathrm{ES}_i(s_a). \tag{10}$$

IV. INTELLIGENT PRICING ALGORITHM BASED ON DEEP REINFORCEMENT LEARNING

In this section, we design a novel dynamic demand-aware Power Grid Intelligent Pricing (PGIP) algorithm based on Deep Reinforcement Learning (DRL) [21]. The proposed PGIP algorithm aims to optimize energy consumption and pricing in real time by leveraging the capabilities of DRL to adapt to dynamic demand patterns and evolving grid conditions.

The PGIP algorithm tackles the challenge of optimizing an agent's benefits within an environment. This is achieved by observing the response (reward) a state receives as a result of an action taken by the agent. The agent's goal is to acquire knowledge about which set of actions (policy) will yield the most favorable outcome (return) from the environment. It's important to note that each action has the potential to alter the environment, making the process reliant on interactive learning between the agent and its surroundings.

A. DYNAMIC DEMAND-AWARE SCHEME

Initially, we have the environment, which includes the energy usage data of each consumer in this context. Conversely, there's the agent, which encompasses processing from the perspectives of the service provider, aggregator, and marketer. These two components engage in interactions within a discrete-time sequence denoted by $t \in T$. As the agent perceives alterations in the Reinforcement Learning (RL) environment caused by an action, it generates state observations. A state S encompasses all the parameters obtained by the usage demand aggregator $c_{s,t}$ from the consumers and the sale behaviors $s_{i,t}$ and price $P^d_{i,t}$ from the providers, as defined as:

$$\begin{aligned} S_U &= [c_{s,t}, P^d_{i,t}], \\ S_S &= [s_{i,t}, P^d_{i,t}]. \end{aligned} \tag{11}$$

The set of actions A of each agent relates to the incentives provided to the users, as defined as:

$$A_{u,h} = [\eta_{u,h}, r_{u,h}]. \tag{12}$$

Ultimately, the overall reward for the approach is defined as the set of rewards or benefits rt.

Based on the mentioned actions, it is essential to define a policy, referred to as π. This policy acts as a set of rules that the agent adheres to decide its actions depending on the current state of the environment. Essentially, it serves as the function that links the action A with the state S:

$$\begin{aligned} Q(S_{U,i,t}, a_{i,t}) = Q(S_{i,t}, a_{i,t}) + \alpha[r(S_{i,t}, a_{i,t}) \\ + \gamma(S_{i,t+1}, a_{i,t+1})] \end{aligned} \tag{13}$$

$$\begin{aligned} Q(S_{S,i,t}, a_{i,t}) = Q(S_{i,t}, a_{i,t}) + \alpha[r(S_{i,t}, a_{i,t}) \\ + \gamma(S_{i,t+1}, a_{i,t+1})] \end{aligned} \tag{14}$$

B. DEEP REINFORCEMENT LEARNING

We develop the DRL algorithm to obtain the optimal policies. The collection of policies acts as a foundation, allowing the agent $S - t$ to initiate without initially relying on an e-greedy policy. Alternatively, it initiates by examining this predetermined collection and then proceeds with iterations to optimize rewards. Once the agent achieves an optimal policy, the S-t Q-learning algorithm registers the policy set to strike a balance between short-term and long-term tactics. As depicted in Fig. 3, Agent $L - t$ gathers data from the environment, encompassing states and rewards (s_t, a_t, r_t, s_{t+1}) resulting from actions a_{t+1}.

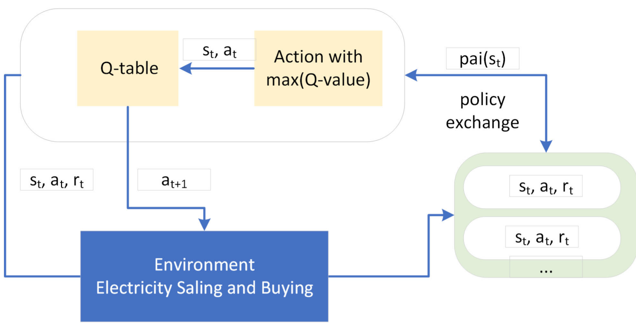


FIGURE 3. Process of the deep reinforcement learning for PGIP.

As shown in Fig. 3, utilizing the above data, the Q-table is generated. Moreover, utilizing the policy interaction approach, the algorithm identifies actions that maximize the Q-values, representing the rewards r_h of $Q^*(s_t, a_t)$. The result of the $L-t$ Q-learning algorithm is the optimal policy $\pi^*(S_t)$, which is then stored in the experiential memory replication. Upon temporal shift, the agent seeks actions that maximize the reward $max(Q - value)$. In this scenario, the agent utilizes its existing knowledge and incorporates iterations of the $L-t$ agent as input for searching the optimal policy $\pi^*(S_t)$. The step-by-step procedure of the proposed DRL-based PGIP algorithm is outlined in Algorithm 1.

Algorithm 1 DRL-Based PGIP Algorithm

Require:
U: the set of electricity customers;
SP: the set of service providers;
Z_{da}: the unit price of each provider.
Ensure:
$Q^*(S_{u,h}, a_{u,h})$: the optimal solution for grid pricing.
for each customer u_i in U **do**
  initialize the value of $Q(S_{u,h}, a_{u,h}) \leftarrow$;
  while $a_{u,h} = maxQ^*(S_{u,h}, a_{u,h})$ **do**
    for each epoch τ **do**
      set the initialization value to $s_{u0,h0}$;
      calculate policy $r(s, a) \leftarrow Q(S_{u,h}, a_{u,h})$;
      perform action a based on state s and policy $r(s, a)$;
    end for
  end while
  save the solution $Q(S_{u,h}, a_{u,h})$;
  update action based on the current optimal solution;
end for
return $Q^*(S_{u,h}, a_{u,h})$.

V. EXPERIMENTS

In this section, we first introduce our experimental setup. Subsequently, we evaluate the effectiveness of our proposed PGIP algorithm by comparing it with the state-of-the-art methods under different cases.

A. EXPERIMENTAL SETUP

In our experiments, we utilize input data obtained from smart meters installed in the consumer population of a distribution network grid. The electrical grid in question is a crucial component of the "Caucete Smart Grid" innovation project [26]. The main aim of this project is to modernize a section of the current electrical distribution network in the City of Caucete, situated in the San Juan province of Argentina, into an advanced and modernized network.

The intended overhaul seeks to improve the operational effectiveness, regulatory mechanisms, and general electrical functionality of the grid. The project's primary objectives include enhancing energy efficiency in electricity usage and elevating service quality to maximize overall benefits for consumers, the utility company, and society as a whole, ultimately boosting social welfare. Additionally, our experiments are designed to promote the utilization of Renewable Energy Sources (RES) for electricity generation, with a specific emphasis on photovoltaic solar energy. Moreover, our experiments aspire to establish an advanced measurement infrastructure that illuminates consumption patterns, facilitating the development of innovative strategies for a sustainable system. In essence, the initiative seeks to create a forward-looking and sustainable electrical distribution system that integrates modern technologies and promotes the utilization of renewable energy, contributing to both energy efficiency and the overall well-being of the community.

1) METRICS

We assess the effectiveness of the suggested pricing strategies using two primary measures. Firstly, we examine the overall alteration in electrical energy usage per day, termed as the variation in demand. This factor, known as DV (Demand Variation), measures the relationship between consumption prior to any demand response action and consumption following the implementation of a demand response program. The formula representing this factor is:

$$\mathrm{DV} = \frac{C_u^o - C_u^n}{C_u^o}, \tag{15}$$

where C_u^o is the original price and C_u^n is the new price. This metric provides valuable insights into the effectiveness of the demand response scheme in influencing and optimizing consumer energy consumption patterns.

Furthermore, we assessed the average load factor of consumers to conduct a comparative analysis of pricing formulations and to determine whether, despite efforts to address demand responses, there was an improvement in the load factor. This metric acts as an extension, reflecting both the significant peak consumer demand and the efficiency of the pricing strategy in altering electricity usage patterns. Furthermore, the assessment was expanded to assess the average load factor of consumers to examine the influence of the pricing strategy and its effectiveness in improving the load factor. This metric offers insights into managing high peak consumer demand and assessing the efficacy of the pricing approach in redirecting electricity usage patterns.

B. COMPARISON WITH RELATED METHODS

We evaluate the effectiveness of the PGIP algorithm by conducting experiments to compare it with the SGDM and RMSProp methods. The experimental results of algorithm comparison are shown in Fig. 4.

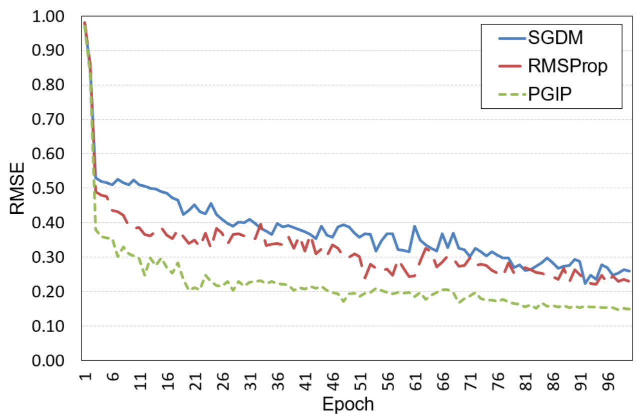


FIGURE 4. Comparison with related methods.

To predict future time step values, a Stochastic Gradient Descent with Momentum (SGDM) model was utilized. This model was trained using responses as training steps, with values adjusted by a single action. The dataset was divided into 90% for training and the remaining 10% for testing.

The Long Short-Term Memory (LSTM) layer was equipped with 128 hidden units to enhance learning and forecasting capabilities. Following a comprehensive evaluation, it was found that the Adam algorithm produced the most favorable results in terms of reducing error, particularly the Root Mean Square Error (RMSE). Consequently, RMSProp was selected as the preferred optimization algorithm for the LSTM network within this predictive modeling framework.

C. CONSUMER GROUPING RESULTS

We discuss the results of residential consumers in whole days. All-time electricity demand records are used in the experiments. The consumer experiences their peak consumption from 9:00 (am) to 11:00 (am) and from 20:00 (pm) to 22:00 (pm). The experimental results of consumer grouping are shown in Fig. 5.

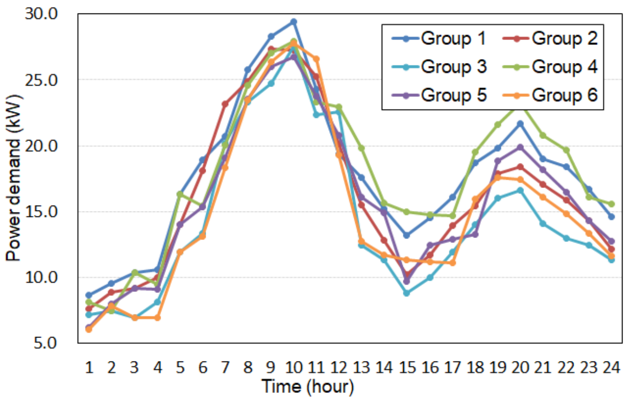


FIGURE 5. Consumer grouping results.

As depicted in Fig. 5, if the consumers use electricity not in the period of system peak, they are selected. However, despite this, the consumer encounters their peak consumption period from 13:00 (pm) to 21:00 (pm). Consequently, the grouping mechanism needs to discern and categorize these consumption patterns, adhering to the specific Demand Response (DR) requirements. In this scenario, based on the group results, the period of system peak is deliberately excluded from consideration. This strategic exclusion ensures that the grouping process accurately captures and classifies the distinct consumption behavior, aligning to optimize demand response strategies for consumers with varying peak hours.

D. PRICING RESULTS WITH COINCIDENCE FACTORS

We further consider the pricing results of our algorithm according to the coincidence factors. We analyze the incorporation of a wholesale market price along with associated elements, demonstrating the effective decrease of the system peak by regulating consumer demand. The experimental results are illustrated in Fig. 6.

As shown in Fig. 6, it's crucial to emphasize that the static price is inherently inefficient in the experiment, as it fails to adapt to the dynamic nature of demand behavior–an

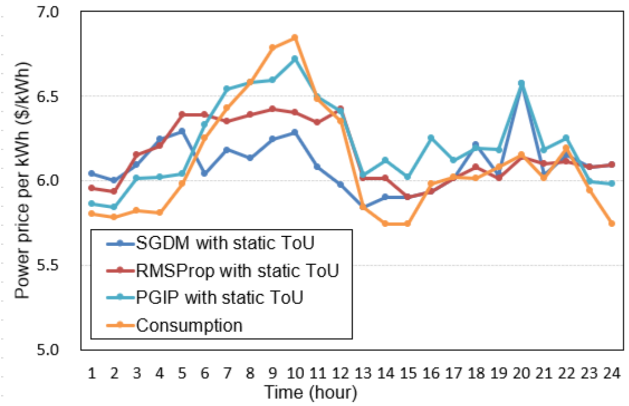


FIGURE 6. Code-related instruction-following experiment output examples for various models.

aspect that is known in advance. Notably, the approach no longer targets the reduction of consumer peak exclusively, and as a consequence, the hours with consumer demand peaks exhibit a more tempered price signal compared to the scenario without considering coincidence factors. Furthermore, thanks to the incorporation of bidirectional satisfaction considerations, the algorithm adeptly devises prices that incentivize consumers to increase their consumption. The feature is particularly evident during the use time of the day. Thus, the experiments demonstrate how the consumer is presented with prices that reflect an elasticity, fostering a more responsive and adaptive approach to energy consumption.

E. EVALUATION ON SYSTEM PEAK

We discuss the effectiveness of the PGIP algorithm during the period of system peak. The electricity consumers' adept reduction in consumption precisely during the designated peak hours of the system. The experimental results are illustrated in Fig. 7.

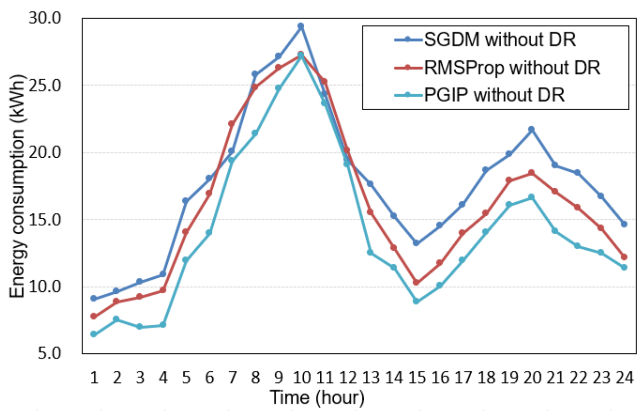


FIGURE 7. Demands and pricing results on system peak.

Fig. 7 illustrates the consumer's adept reduction in consumption precisely during the designated peak hours of the system, thereby successfully accomplishing the demand response objective. Conversely, owing to the model's provision of a price lower than a small rate, and the related electricity consumption tends to experience a more moderate increase. Nevertheless, it effectively serves the purpose of mitigating the impact on the system peak. This demonstrates the nuanced influence of pricing strategies on consumer behavior, striking a balance between incentivizing demand reduction during peak times and encouraging consumption at a rate that aligns with overall system efficiency.

F. EVALUATION UNDER DIFFERENT CONSUMER GROUPS

We discuss the reward behaviors of our algorithm under different consumer groups. Fig. 8 illustrates the cumulative reward patterns concerning the formulation of demand response prices, particularly concentrating on different 6 consumer groups.

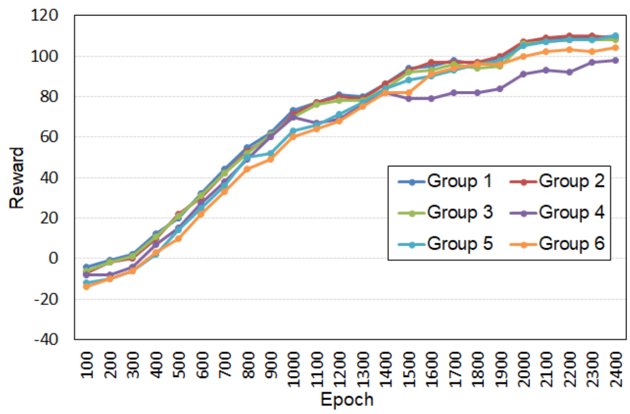


FIGURE 8. Cumulative reward behaviors under different consumer groups.

As shown in Fig. 8, the rewards derived from the DL method are outcomes of optimizing the benefits for each participant. Our algorithm effectively achieves its objective in approximately 600 episodes, although it is configured to run for 1000 episodes in this instance. Notably, the algorithm exhibits a tendency to seek analogous strategies across the three cases, indicating a consistent pattern of growth. This suggests a convergence towards similar optimal solutions for the different scenarios.

G. PRICING RESULTS FOR DYNAMIC DEMANDS

We further evaluate the performance of the PGIP algorithm under dynamic demands. A time frame of four days was set to ascertain the model's prediction, and reference consumer data was utilized to validate the algorithm's understanding of long-term dynamics. Moreover, the algorithm is designed to transfer data via shared information between different periods. This design improvement aims to boost efficiency and streamline the quest for the most favorable solution. The pricing results for dynamic demands are shown in Fig. 9.

As illustrated in Fig. 9, the algorithm excels in identifying optimal pricing strategies, not only to alleviate long-term system peaks but also to incentivize the identification of long-term peaks. This capability shows that could potentially facilitate a direct load control scheme. As the energy landscape

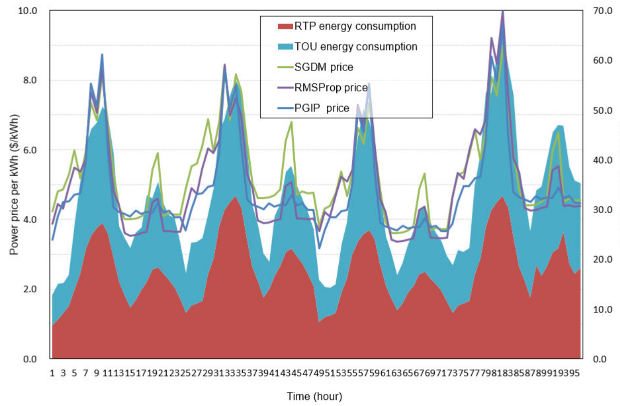


FIGURE 9. Pricing results for dynamic demands in four days.

continues to transform, PGIP is a potential of DRL, in shaping the future of efficient energy management. This research contributes not only a practical and effective algorithm but also insights that can inform the development of intelligent grid management systems. PGIP presents a crucial step towards a more sustainable, adaptive, and resilient power grid infrastructure, addressing the dynamic nature of energy demand and fostering the efficient use of renewable resources.

VI. CONCLUSION

This paper presented the Power Grid Intelligent Pricing (PGIP) algorithm, driven by Deep Reinforcement Learning (DRL), to address the pressing challenges arising from the integration of energy sources and modern power grids. Through an innovative approach, PGIP optimizes energy consumption and pricing in real-time by dynamically adapting to the intricate interplay of dynamic demand patterns and evolving grid conditions. The incorporation of a sophisticated neural network architecture enables PGIP to model the complex relationships among various grid parameters, user demand, and pricing strategies with unprecedented accuracy. The continuous learning and adaptation capabilities of the algorithm empower it to dynamically adjust pricing structures, fostering demand-side flexibility while ensuring grid stability. Empirical validation using real-world datasets showcases PGIP's exceptional performance across diverse scenarios. The algorithm proves its efficacy by optimizing energy consumption, reducing peak demand, and enhancing overall grid efficiency. Comparative analyses against traditional pricing models underscore PGIP's superior adaptability and responsiveness, positioning it as a promising solution to meet the evolving challenges of modern power grids.

REFERENCES

[1] S. Janko and N. G. Johnson, "Reputation-based competitive pricing negotiation and power trading for grid-connected microgrid networks," *Appl. Energy*, vol. 277, Nov. 2020, Art. no. 115598.

[2] A. Amin, W. U. K. Tareen, M. Usman, H. Ali, I. Bari, B. Horan, S. Mekhilef, M. Asif, S. Ahmed, and A. Mahmood, "A review of optimal charging strategy for electric vehicles under dynamic pricing schemes in the distribution charging network," *Sustainability*, vol. 12, no. 23, p. 10160, Dec. 2020.

[3] G. Yuan, Y. Gao, B. Ye, and R. Huang, "Real-time pricing for smart grid with multi-energy microgrids and uncertain loads: A bilevel programming method," *Int. J. Electr. Power Energy Syst.*, vol. 123, Dec. 2020, Art. no. 106206.

[4] B. Xu, J. Wang, M. Guo, J. Lu, G. Li, and L. Han, "A hybrid demand response mechanism based on real-time incentive and real-time pricing," *Energy*, vol. 231, Sep. 2021, Art. no. 120940.

[5] K. Hanauer, M. Henzinger, L. Ost, and S. Schmid, "Dynamic demand-aware link scheduling for reconfigurable datacenters," 2023, *arXiv:2301.05751*.

[6] M. E. Kabir, C. Assi, H. Alameddine, J. Antoun, and J. Yan, "Demand-aware provisioning of electric vehicles fast charging infrastructure," *IEEE Trans. Veh. Technol.*, vol. 69, no. 7, pp. 6952–6963, Jul. 2020.

[7] D. S. N. K. Pg. Ali Kumar, S. H. S. Newaz, F. H. Rahman, G. M. Lee, G. Karmakar, and T.-W. Au, "Green demand aware fog computing: A prediction-based dynamic resource provisioning approach," *Electronics*, vol. 11, no. 4, p. 608, Feb. 2022.

[8] M. Pacut, W. Dai, A. Labbe, K.-T. Foerster, and S. Schmid, "Improved scalability of demand-aware datacenter topologies with minimal route lengths and congestion," *Perform. Eval.*, vol. 152, Dec. 2021, Art. no. 102238.

[9] H.-M. Chung, S. Maharjan, Y. Zhang, and F. Eliassen, "Distributed deep reinforcement learning for intelligent load scheduling in residential smart grids," *IEEE Trans. Ind. Informat.*, vol. 17, no. 4, pp. 2752–2763, Apr. 2021.

[10] R. Huang, Y. Chen, T. Yin, X. Li, A. Li, J. Tan, W. Yu, Y. Liu, and Q. Huang, "Accelerated derivative-free deep reinforcement learning for large-scale grid emergency voltage control," *IEEE Trans. Power Syst.*, vol. 37, no. 1, pp. 14–25, Jan. 2022.

[11] Y.-C. Hung and G. Michailidis, "Modeling and optimization of time-of-use electricity pricing systems," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 4116–4127, Jul. 2019.

[12] P. Yang, G. Tang, and A. Nehorai, "A game-theoretic approach for optimal time-of-use electricity pricing," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 884–892, May 2013.

[13] P. K. Wesseh and B. Lin, "A time-of-use pricing model of the electricity market considering system flexibility," *Energy Rep.*, vol. 8, pp. 1457–1470, Nov. 2022.

[14] I. Antonopoulos, V. Robu, B. Couraud, D. Kirli, S. Norbu, A. Kiprakis, D. Flynn, S. Elizondo-Gonzalez, and S. Wattam, "Artificial intelligence and machine learning approaches to energy demand-side response: A systematic review," *Renew. Sustain. Energy Rev.*, vol. 130, Sep. 2020, Art. no. 109899.

[15] Y. Liu, X. Guan, J. Li, D. Sun, T. Ohtsuki, M. M. Hassan, and A. Alelaiwi, "Evaluating smart grid renewable energy accommodation capability with uncertain generation using deep reinforcement learning," *Future Gener. Comput. Syst.*, vol. 110, pp. 647–657, Sep. 2020.

[16] A. J. del Real, F. Dorado, and J. Durán, "Energy demand forecasting using deep learning: Applications for the French grid," *Energies*, vol. 13, no. 9, p. 2242, May 2020.

[17] A. Almaghrebi, F. Aljuheshi, M. Rafaie, K. James, and M. Alahmad, "Data-driven charging demand prediction at public charging stations using supervised machine learning regression methods," *Energies*, vol. 13, no. 16, p. 4231, Aug. 2020.

[18] E. U. Haq, X. Lyu, Y. Jia, M. Hua, and F. Ahmad, "Forecasting household electric appliances consumption and peak demand based on hybrid machine learning approach," *Energy Rep.*, vol. 6, pp. 1099–1105, Dec. 2020.

[19] T. Wang, Y. Tang, Y. Huang, X. Chen, S. Zhang, and H. Huang, "Automatic adjustment method of power flow calculation convergence for large-scale power grid based on knowledge experience and deep reinforcement learning," in *Proc. IEEE 4th Conf. Energy Internet Energy Syst. Integr. (EI2)*, Oct. 2020, pp. 694–699.

[20] Y. Li, R. Wang, and Z. Yang, "Optimal scheduling of isolated microgrids using automated reinforcement learning-based multi-period forecasting," *IEEE Trans. Sustain. Energy*, vol. 13, no. 1, pp. 159–169, Jan. 2022.

[21] S. Bahrami, Y. C. Chen, and V. W. S. Wong, "Deep reinforcement learning for demand response in distribution networks," *IEEE Trans. Smart Grid*, vol. 12, no. 2, pp. 1496–1506, Mar. 2021.

[22] M. Dabbaghjamanesh, A. Moeini, and A. Kavousi-Fard, "Reinforcement learning-based load forecasting of electric vehicle charging station using Q-learning technique," *IEEE Trans. Ind. Informat.*, vol. 17, no. 6, pp. 4229–4237, Jun. 2021.

[23] K. Aurangzeb, S. Aslam, S. M. Mohsin, and M. Alhussein, "A fair pricing mechanism in smart grids for low energy consumption users," *IEEE Access*, vol. 9, pp. 22035–22044, 2021.

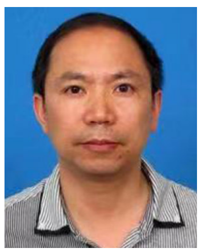
[24] M. B. Rasheed, M. A. Qureshi, N. Javaid, and T. Alquthami, "Dynamic pricing mechanism with the integration of renewable energy source in smart grid," *IEEE Access*, vol. 8, pp. 16876–16892, 2020.

[25] H. T. Dinh, J. Yun, D. M. Kim, K.-H. Lee, and D. Kim, "A home energy management system with renewable energy and energy storage utilizing main grid and electricity selling," *IEEE Access*, vol. 8, pp. 49436–49450, 2020.

[26] G. Coria, F. Penizzotto, and R. Pringles, "Economic analysis of rooftop solar PV systems in Argentina," *IEEE Latin Amer. Trans.*, vol. 18, no. 1, pp. 32–42, Jan. 2020.

CHAO TANG received the M.S. degree in computer science from Purdue University, West Lafayette, IN, USA, in 2015. He is currently pursuing the Ph.D. degree with the College of Computer Science and Electronic Engineering, Hunan University. His research interests include Machine learning, deep learning, and power systems.

YUNCHUAN QIN received the Ph.D. degree in computer science and technology from Hunan University, Changsha, China, in 2019. He is currently the Director of the Experiment Center for Information Technology, College of Computer Science and Electronic Engineering, Hunan University. His research interests include autonomous unmanned systems and high-performance edge computing.



FAN WU received the M.S. degree in computer science from Wuhan University, Wuhan, China, in 2004, and the Ph.D. degree in computer science from Hunan University, Changsha, China, in 2017. He is currently the Vice Professor of the College of Computer Science and Electronic Engineering, Hunan University. His research interests include parallel and distributed computing, machine learning, and DNA computing.

ZHUO TANG (Member, IEEE) received the Ph.D. degree in computer science from Huazhong University of Science and Technology, China, in 2008. He is currently a Professor with the College of Computer Science and Electronic Engineering, Hunan University. He is also the Chief Engineer of the National Supercomputing Center, Changsha. His majors are distributed computing systems, cloud computing, and parallel processing for big data, including distributed machine learning, security models, parallel algorithms, and resource scheduling and management in these areas. He has published almost 140 journal articles and book chapters. He is a member of ACM and CCF.