AI-Optimized Energy Management for More Efficient and Sustainable Microgrids

Sebastián López Flórez[1,2,3](✉), Guillermo Herniández[1,2,3,4],
Alfonso Gonziález-Briones[1,2,3], and Fernando de la Prieta[1,3]

[1] BISITE Digital Innovation Hub, University of Salamanca, Edificio Multiusos
I+D+i, Calle Espejo 2, 37007 Salamanca, Spain
{sebastianlopezflorez,guillehg,alfonsogb,fer}@usal.es
[2] Air Institute, IoT Digital Innovation Hub, 37188 Salamanca, Spain
[3] Universidad Tecnológica de Pereira, Cra. 27 N 10-02, Pereira, Risaralda, Colombia
[4] Institute for Artificial Intelligence and Big Data, Universiti Malaysia Kelantan,
City Campus, Pengkalan Chepa, 16100 Kota Bharu, Kelantan, Malaysia

Abstract. The implementation of renewable energy sources reduces dependence on fossil fuels and greenhouse gas emissions, but it comes at a significant cost to the grid due to the need for coupling multiple energy sources. This necessitates the optimization of energy production using artificial intelligence techniques that can reflect the characteristics of distributed energy management from the storage capacity of the microgrid while considering the consumer load. As such, this study develops an energy management algorithm based on the Deep-Q-Network (DQN) with prioritized experience replay (PER) utility function to model the operator's risk-aversion or risk-seeking behavior, generating optimal exchange strategies in a dynamic environment. The reinforcement learning-based agent provides an understanding of energy storage capacity constraints in aggregate load/discharge energy decision making in the microgrid, using a discrete action space that depends on a reward related to the value of the optimal online objective function of the microgrid. By analyzing microgrid management under different system state definitions, the results show that substantially improved performance can be achieved compared to a traditional model that assumes deterministic information. The findings provide solid experimental validation of the potential of deep reward-based learning techniques to enable more efficient, flexible, and cost-effective energy management, particularly with the integration of more intermittent renewable energy sources.

Keywords: Smart grid technologies · Energy efficiency optimization · contingency reserve · Reinforcement learning

The IoTalentum Project within the Framework of Marie Skłodowska-Curie Actions Innovative Training Networks (ITN)-European Training Networks (ETN), which is funded by the European Union Horizon 2020 Research and Innovation.

R. Mehmood et al. (Eds.): DCAI 2023, LNNS 741, pp. 438–447, 2023.
https://doi.org/10.1007/978-3-031-38318-2_43

1 Introduction

Due to the exponential rise in energy demand, efforts are being made to increase the production of clean, sustainable energy while improving energy efficiency to lower consumption through the use of renewable energy sources including solar, eolian, and geothermal energy. This is due to the intermittent nature of the energy along with the stochasticity of the loads, the cost of network energy, and the availability of the network, which makes the generation of energy from renewable sources and the demand for energy a challenge as modern electrical systems become more complex [5].

The incorporation of distributed energy into the electrical system gives rise to the concept of a microgrid, an energy supply system that may function independently or in conjunction with an energy distribution network. When connected to the energy distribution network, the Microgrid can operate either in an isolated mode to provide local services with reliability guarantees or in a connected mode to fully utilize the distributed energy resources, subject to the characteristics of the main network and providing supplementary services like redundant energy sales to the network or purchasing additional energy from the network when needed. It is anticipated that a combination of several cooperating Microgrid autonomous systems will become the dominating configuration in the next-generation intelligent network [16].

The objective of the Microgrid energy management strategy is to use distributed energy efficiently and, at the same time, reduce long-term operating costs by buying and selling electricity on the distribution grid. In addition, due to the large number of sources distributed by Microgrid, it is necessary to perform statistical analysis, load consumption forecasting and data mining in order to establish benefits of optimal generation dispatch, energy savings, increased reliability and reduced costs due to losses. A number of studies have addressed this problem focusing on energy management considering various approaches to achieve the optimal operation of microgrids based on the expected load and renewable energy resources. Clarke et al. [2] studied a standalone desalination renewable energy system with the intention of sizing and power management. In the end, it was discovered that MOPSO's optimization produced greater outcomes in lowering NPC and CO2 emissions. TACCARI et al. [15] propose comparing two optimization approaches to deal with short-term operational planning of the energy systems, such as electric power plants, calderas, heat storage facilities, and cogeneration units. In considering the option optimized operation of multiple microgrids, reference [10] uses the random optimization method of generation and scene reduction to cope with renewable energy uncertainty. In ILR Gomes et al. [6] focuses on a management system to support and plan the operation of a microgrid by the new entrant in the electricity market, the microgrid aggregator This system is based on mixed-linear entera estocatic programming. Zhang et al. [17] took a step forward, which both includes electric turbines and photovoltaic panels. They developed an energy management model based on uncertain factors represented by typical scenarios to coordinate the operation of the electrical and thermal systems. However, these methods are not guaranteed to fall into the local optimum easily, In addition, the lack of an ideal target for the correction makes it difficult to apply the methods now available to make practical control decisions and in some cases generates a significant computational burden.

The challenges surrounding energy consumption scheduling, energy trading, and storage management in microgrids entail intricate interdependencies among various components and nonlinear dynamics under constant changes in system parameters, such as real-time pricing, contingency reserve requirements, and operator risk preferences. Traditional optimization and rule-based techniques rely on strong assumptions of precise predictions, static models, and linear relationships that are invalid for such complex cyber-physical energy systems [9,12].

In contrast, reinforcement learning acquires the optimal policy through continuous trial-and-error interactions with the dynamic system. The learned policy adapts to changes in conditions without requiring retraining, providing a flexible and optimal approach [4]. Deep reinforcement learning further enhances its capabilities by allowing the handling of high-dimensional state and action spaces with complex nonlinear interactions. Empirical results also suggest that deep Q-learning(DQN) techniques can outperform rule-based and optimization-based baselines for complex decision-making problems under uncertainty.

Specifically, we propose a deep neural network (DNN) model with prioritized experience replay (PER) that prioritizes experiences with large errors, which has shown to substantially improve the accuracy and speed up the training of deep neural networks in high-dimensional spaces by updating the weights of relevant experiences in the buffer memory as shown in the Fig. 1. This technique enables the algorithm to focus on the most informative data points, rapidly improving the optimal control policy compared to a Q-Network approach, also validated in this research for this use case.

This document is structured as follows: Sect. 2 details the microgrid energy management problem considering flexible option sources and describes the deep RL framework. Section 3 presents comparative results of the methods used in the use case. Section 4 provides conclusions

2 Deep RL for Energy Microgrids Management

Microgrids are interconnected local distribution power systems that can operate in isolation or connected to the main distribution grid. These smart grids integrate distributed renewable generation, energy storage systems and loads, and allow optimizing the use of resources at the local level, as well as improving reliability and security of supply. They facilitate the integration of renewable energies by providing operational flexibility and active management of supply and demand. They can include multiple generation technologies such as solar photovoltaic, wind, biomass or mini-hydro. Storage systems, such as batteries or flywheels, provide backup capacity, smooth the variability of renewables and enable new grid management services. Finally, this paper assumes that the microgrid is not connected to the grid, where the objective function is the energy

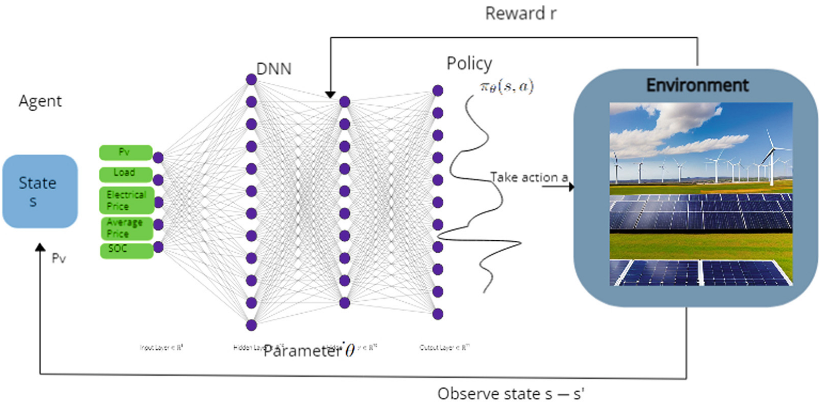


Fig. 1. Simple schematic of algorithm

management system that determines how the microgrid behaves in response to different scenarios. In these papers, the factors considered in the objective functions consist of an aggregate consumption load, a PV farm and an Energy Storage System, which is connected only to the main grid through a power distribution line. The microgrid is managed by a microgrid operator who has full control of all microgrid operations, including the power transaction operation with the main grid [1].

The following assumptions are made about microgrid systems each time interval is assumed to last 1 h. Therefore, a system divided into N time slots of 1 h duration each is considered. Each time slot, denoted as 4t, has a duration of 1 h, indicating that there are 24 power scheduling time slots per day. Each of these 24 time slots represents a 1-h period, so the optimal energy scheduling must be determined hour by hour over a full day. Second, a simplified (quasi-static) model of the dynamics of the microgrid energy system is assumed to facilitate its study and optimization, third, it is assumed that the generation and consumption of energy in the microgrid remains constant within each time interval (slot) considered, but changes from one slot to another to capture its variability, third, it is assumed that the generation and consumption of energy in the microgrid remains in equilibrium, with no power deficit or surplus [3].

2.1 Reinforcement Learning

Reinforcement learning (RL) is a field of artificial intelligence focused on the sequential interaction of decisions between an agent and an environment. The agent pursues an optimal policy of actions that maximizes a reward received from the environment over time.

Formally, we define a state space S, an action space A and a family of reward functions $r(s,a,s') : S \times A \times S \longrightarrow R$, where R are the reals. A sequence of states, actions and rewards forms a trajectory (also called episode, history or roll-out) $\tau = (s_0, a_0, r_0, s_1, a_1, r_1, ...)$ defined by the dynamics of the Markov decision model (MDP).Each transition has a likelihood of happening $p(s'|s,a)$ and offers a specific amount of benefit determined by $r(s,a,s')$. In episodic tasks, the horizon τ is finite, while in continuing tasks τ is infinite.

The agent's policy of actions is defined as $\pi : S \longrightarrow A$. Given a trajectory and a policy, the cumulative reward is computed as $G(\tau,\pi) = \sum_{k=0}^{\tau} \gamma r_t$. The optimal policy π_* maximizes the expected reward $G\pi$. If the task is episodic (τ is finite, the trajectories ends after a finite number of transitions), γ can be set to 1, but if the task is continuing ($\tau = \infty$, trajectories have no end), γ must be chosen smaller than 1.

The iterative optimization process of the policy π_* is known as reinforcement learning (RL) algorithm. It can be either value-based or policy-based. The former update a value function $V\pi(s)$, while the latter directly update the policy $\pi(s)$. In Policy-Based methods, we learn a policy function directly.

Learning takes place by exploring the space of states and actions to find the optimal paths, and exploiting the knowledge gained to obtain the highest possible reward. After many iterations, the policy converges to an optimal solution π_*.

2.2 Deep RL Solutions for Sequential Decision Making

This work uses a Deep Learning version of Q-Learning [13] called Deep Q-Network (DQN) and Prioritized Experience Replay (PER) as configured and conditioned in [5,14]. Another feature of some of these algorithms is the ability to work with continuous and discrete action spaces, but since DQN only works with discrete action spaces, these algorithms require discretization of continuous action space environments.

Deep Q-Network (DQN). To formally define the DQN algorithm, let $Q(s,a;\theta)$ be the Q function that we estimate with a deep neural network with parameters θ. The goal of the algorithm is to find the parameters $\theta*$ that minimize the difference between the estimated Q function and the true Q function. The algorithm updates the Q function as follows.

$$q_\star(s,a) = E\left[R_{t+1} + \gamma\ max_{a'} q_\star(s',a')\right] \tag{1}$$

After calculating the loss, the weights are updated in the network using SGD and backpropagation, like any other normal network. This process is repeated for each state of the medium until the losses are sufficiently reduced and an approximately optimal Q function is obtained.

The Q-function is updated as follows:

$$q_\star(s,a) = q_\star(s,a) + \alpha[r + \gamma\ max_{a'} q_\star(s',a') - q_\star(s,a)] \tag{2}$$

Prioritized Experience Replay (PER). The Prioritized Experience Replay (PER) algorithm is an extension of the memory replay method used in the DQN algorithm. In PER, experiences stored in the replay memory are prioritized based on their importance for improving learning. Each experience is associated with a weight indicating its importance, and a non-uniform selection criterion is used to sample experiences from the replay memory, ensuring that the probability of being sampled is monotonic in the priority of a transition, while ensuring a non-zero probability even for the lowest priority transition. The most important experiences (i.e., those with higher weight) are sampled more frequently, helping the algorithm to focus on the most relevant experiences and improve learning more efficiently. The stochastic sampling method interpolates between purely greedy prioritization and uniform random sampling. PER prioritizes them according to their expected importance for future learning. Transitions with higher temporal difference (TD) errors, which indicate greater learning potential, are replayed more frequently. This allows learning to quickly focus on regions of the state-action space where the approximate value function differs most from the optimal.

The sampling probability of a transition is defined as follows:

$$P(i) = \frac{p_i^\alpha}{\sum_k p_k^\alpha} \tag{3}$$

Here, $p_i > 0$ represents the priority of the transition, and the exponent determines the extent of prioritization, where $\alpha = 0$ corresponds to the uniform case.

Importance sampling weights rectified the uncontrolled shift in sampling distribution under prioritized experience replay. By reweighting priorities, IS stabilized the solution that value estimates would converge to rather than diverging due to bias from prioritization.

$$w_i = \left(\frac{1}{N} \cdot \frac{1}{P\left(i\right)}\right)^\beta \tag{4}$$

which fully explains the non-uniform probabilities $P\left(i\right)$ if $\beta = 1$.

We use the absolute value of the magnitude of our TD error:

$$p_i = |\delta_i| + \epsilon \tag{5}$$

2.3 Application of RL in the Microgrid

State Definition. In the case of microrgrids, the state of the system can be described using a tuple of five components that captures the relevant details of the system's state for each time step. This state representation is then used to construct a state-based window, represented as st, which captures the system's current and past states for a given time interval.

$$s_t \in S = \left\{P_t^{WG}, P_t^{Load}, Pr_t, Pr_t^{avg}, SOC_t\right\}$$

where electrical load P_t^{Load}, wind power P_t^{PV} and electrical price Pr_t are obtained from the time series dataset, while SOC and average electrical price Pr_t^{avg} for the past 24 h need to be calculated at each time step [11].

An excerpt from a dataset encompassing load consumption data and dynamic prices, sourced from [8] and employed for model evaluation, exhibits the trends of photovoltaic solar energy generation, electricity consumption, and prices over a 24-h period.

Action Definition. In the context of managing lithium-ion battery energy, it is crucial to determine the amount of energy to be charged or discharged using a reinforcement learning algorithm. This algorithm enables the agent's action at time t to be defined, incorporating fundamental aspects into the utility function [3].

- Maintaining a minimum level of contingency reserves or target state of charge (SOC) in the ESS to ensure critical operations. The utility function will be piecewise, depending on if the SOC is below, at or above this target level.
- Below the target SOC, the agent must charge the ESS as soon as possible regardless of price. And the price must be high enough for discharging at this level.
- Above the target SOC, the criteriachanges. At higher SOC levels, the agent should charge at lower prices and discharge at higher prices to better utilize the remaining storage capacity. Closer to the target SOC, the agent should do the opposite.
- A negative exponential multiplier (M) is used to incorporate these criteria into the utility function. Parameters can be tuned to control the agent's behavior below and above the target SOC.
- Two penalties are also introduced to further control the agent's actions:
  - PntESSt: Penalizes actions that violate ESS operating limits. This helps preserve the ESS.
  - PntPVt: Penalizes not storing surplus solar power when there is spare capacity. This aims to maximize use of free energy within limits.

The overall utility function (u) incorporates all these elements to guide the agent's trading strategies regarding the ESS in a way that maintains necessary reserves, utilizes storage effectively and prolongs the ESS lifespan.

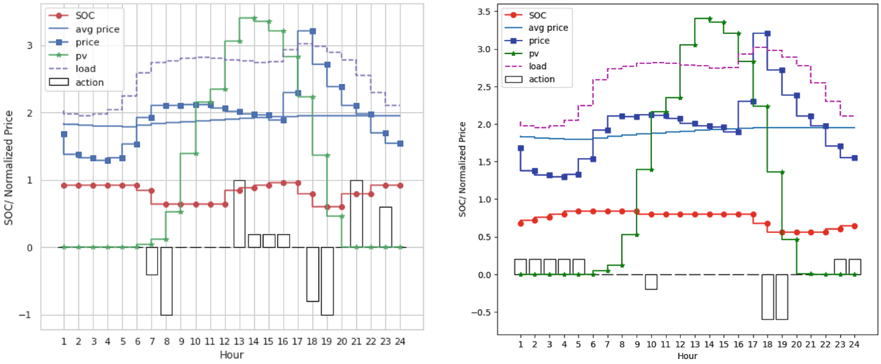
3 Results

An energy storage system utilizing lithium-ion batteries (ESS) is employed, with a nominal power of 1000 kW and a nominal capacity of 5000 kWh. The ESS boasts a cycle life of 4996, a δ value of 1, and a respective η_c and η_d of 90%. It is established that C_i is 171/kWh. The critical microgrid operations require a target SOC of 0.5, equivalent to 2500 kWh [7]. Other hyperparameters include (Table 1).

Table 1. Hyperarameters Reinforcement Learning Model

ρ	d	err	α	β
0.6	0.0001	0.01	0.001	0.4

Simulations were conducted to assess the impact of risky and prudent behaviors of the MGO on the ESS operation decisions and performance. Furthermore, the performance of the proposed DQN-PER and DQN algorithm was compared with other algorithms. The numerical results of the simulations were compiled and tabulated in Table 2. The ESS operations and discharge patterns during a single day using different algorithms are depicted in Fig. 2a and 2b., respectively, under restricted conditions of surplus solar energy and ESS capacity reserve.



(a) PER ESS actions .

(b) DQN ESS actions

Fig. 2. Optimal trading strategy for a 24 h period

The average daily costs for DQN amount to $62.62. By taking perfect knowledge into account, DQN is capable of making more informed decisions when compared to relying solely on forecast knowledge. Conversely, the average daily costs for PER are $22.75, respectively. With a risk-averse perspective, the Microgrid tends to be less aggressive in energy trading compared to a risk-seeking perspective. This implies that, when considering potential risks, Microgrid takes a more cautious approach in its energy marketing activities.

Table 2. Hyperarameters Reinforcement Learning Model

	DQN-PER	DQN	DQN [7]
Benefit($)	22.75	62.62	53.45

There are two main observations from the results:

- The probability of the state of charge (SOC) falling below the target SOC is lowest using the DQN algorithm, and then the DQN-PER algorithm
- The daily monetary benefit obtained is highest using the DQN surpassing that of the state of the art

The ESS actions on how DQN-PER algorithm acted as an energy optimizer that made intelligent data-driven decisions. It charged the battery to maintain a high state of charge when the load was low and discharged the stored capacity during periods of high demand, thus fulfilling the reward function designed. Thus, ESS acts more meticulously in the DQN algorithm compared to the DQN-PER algorithm.

4 Conclusions

This research proposes a deep reinforcement learning approach to optimize the performance of microgrids under simulated conditions. The results demonstrate that an efficient control strategy can be applied to new scenarios using a neural network representation. The proposed approach has the advantage of allowing the algorithm to learn from its own experience. The use of algorithms such as DQN-PER and DQN can improve the efficiency and flexibility of energy trading, resulting in greater benefits for prosumers and a more sustainable energy system overall. The presented results highlight the importance of considering perfect knowledge when making informed energy trading decisions. The DQN algorithm was able to achieve higher daily costs compared to DQN-PER, indicating a more optimal and profitable energy trading strategy. Furthermore, the study revealed the impact of risk aversion on energy trading decisions, with the Microgrid taking a more cautious approach to energy marketing activities. These findings demonstrate the importance of considering risk in energy trading and the potential benefits of implementing adaptive algorithms that can respond to changing market conditions. Overall, these results provide valuable insights into the effective management of energy trading activities and the potential for AI-based approaches to enhance energy system sustainability and profitability.

As a future work, implementing a more descriptive model where the output is characterized by a distribution is a promising direction. This approach can provide a more accurate representation of the uncertainty involved in energy trading and enable more informed decision-making. Furthermore, incorporating probabilistic models can facilitate the development of risk-averse strategies that take into account the potential outcomes of different decisions. The use of distribution-based models can also enable the modeling of complex systems with multiple sources of uncertainty and provide a more comprehensive understanding of the underlying dynamics. Overall, the implementation of distribution-based models represents a promising avenue for enhancing the accuracy and robustness of energy trading algorithms.

References

1. Adhikari, R.S., Paudyal, P.: Handbook on Microgrids for Power Quality and Connectivity, 1st edn. Elsevier, Amsterdam (2019)
2. Affi, S., Cherif, H., Belhadj, J.: Smart system management and techno-environmental optimal sizing of a desalination plant powered by renewables with energy storage. Int. J. Energy Res. **45**(5), 7501–7520 (2021)
3. Aih, H.C., Kumar, P.P.S., Srinivasan, P.D.: Energy management economic evaluation of grid-connected microgrid operation. Department of Electrical Computer Engineering, National University of Singapore (2020)
4. Du, Y., Li, F.: Intelligent multi-microgrid energy management based on deep neural network and model-free reinforcement learning. IEEE Trans. Smart Grid **11**(2), 1066–1076 (2019)
5. Gassi, K.B., Baysal, M.: Improving real-time energy decision-making model with an actor-critic agent in modern microgrids with energy storage devices. Energy **263**, 126105 (2023)
6. Gomes, I., Melicio, R., Mendes, V.: A novel microgrid support management system based on stochastic mixed-integer linear programming. Energy **223**, 120030 (2021)
7. Hau, C., Radhakrishnan, K.K., Siu, J., Panda, S.K.: Reinforcement learning based energy management algorithm for energy trading and contingency reserve application in a microgrid. In: 2020 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe), pp. 1005–1009. IEEE (2020)
8. Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A., Hyndman, R.: Probabilistic energy forecasting: global energy forecasting competition 2014 and beyond. Int. J. Forecast. **32**(3), 896–913 (2016). https://EconPapers.repec.org/RePEc:eee:intfor:v:32:y:2016:i:3:p:896-913
9. Ji, Y., Wang, J., Xu, J., Fang, X., Zhang, H.: Real-time energy management of a microgrid using deep reinforcement learning. Energies **12**(12), 2291 (2019)
10. Kargarian, A., Rahmani, M.: Multi-microgrid energy systems operation incorporating distribution-interline power flow controller. Electr. Power Syst. Res. **129**, 208–216 (2015)
11. Liu, F., Liu, Q., Tao, Q., Huang, Y., Li, D., Sidorov, D.: Deep reinforcement learning based energy storage management strategy considering prediction intervals of wind power. Int. J. Electr. Power Energy Syst. **145**, 108608 (2023)
12. Nakabi, T.A., Toivanen, P.: Deep reinforcement learning for energy management in a microgrid with flexible demand. Sustain. Energy Grids Netw. **25**, 100413 (2021)
13. Ong, H.Y., Chavez, K., Hong, A.: Distributed deep Q-learning, arXiv preprint arXiv:1508.04186 (2015)
14. Saglam, B., Mutlu, F.B., Cicek, D.C., Kozat, S.S.: Actor prioritized experience replay, arXiv:2209.00532 (2022)
15. Taccari, L., Amaldi, E., Martelli, E., Bischi, A.: Short-term planning of cogeneration power plants: a comparison between MINLP and piecewise-linear MILP formulations. In: Computer Aided Chemical Engineering, vol. 37, pp. 2429–2434. Elsevier (2015)
16. Wu, N., Wang, H.: Deep learning adaptive dynamic programming for real time energy management and control strategy of micro-grid. J. Clean. Prod. **204**, 1169–1177 (2018)
17. Zhang, Y., Meng, F., Wang, R., Kazemtabrizi, B., Shi, J.: Uncertainty-resistant stochastic MPC approach for optimal operation of CHP microgrid. Energy **179**, 1265–1278 (2019)