Energy Proceedings
Vol 45, 2024
ISSN 2004-2965

Optimizing Participation in The Local Energy Market Using a Deep Reinforcement Learning Approach

Helder Pereira, Luis Gomes and Zita Vale

Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development (GECAD), Intelligent Systems Associated Laboratory (LASI), Polytechnic of Porto (P.PORTO), Rua Dr. António Bernardino de Almeida 431, 4200-072 Porto, Portugal

{hjvlp, lfg, zav}@isep.ipp.pt

ABSTRACT

The actors involved in the energy network have benefited much from its expansion in recent decades, but the network's management has become more difficult as a result of the sources' variability and unpredictability. Thus, it is essential to create models that can manage the current energy resources, which are becoming more and more dispersed. This study provides a new optimization model for participating in local energy markets based on peer-to-peer energy trading, using the twin-delayed deep deterministic policy gradient method and the double-auction trading mechanism. The model is integrated into an ecosystem based on agents, which enables the modeling of energy communities to produce a more plausible implementation scenario. The concept was used in a case study with 30 players in an energy community, and the findings revealed that each member saved an average of 1.54 EUR per week.

Keywords: deep reinforcement learning, local energy markets, multi-agent systems, peer-to-peer energy trading, twin-delayed deep deterministic policy gradient.

1. INTRODUCTION

The electrical grid has considerably incorporated distributed energy resources (DER), which benefits both the network and its users [1]. By 2050, it is expected that more than half of the world's generation will come from renewable energy sources (RES), which are also gaining ground quickly [2]. This development has brought about a paradigm change in favor of a distributed smart grid and enhanced flexibility toward a more dependable system. Power and energy systems (PES) planning and operation have been impacted by the output of RES's reduced predictability and stability [3].

The growth of local energy markets LEMs and peer-to-peer (P2P) energy sharing demands the optimization of the player's participation in these types of markets [4]. Reinforcement learning (RL) is a common approach in this domain. One of the most explored algorithms to do it is Q-Learning, and it can be used with a multi-agent variant to determine the optimal approaches for energy market pricing negotiations [5]. The most important issue with Q-Learning is that it cannot deal with continuous observation and action spaces in an original manner, particularly important in complex contexts, such as the smart grid, where most observations are continuous values. Another often used technique in multi-agent contexts is the multi-agent deep deterministic policy gradient (DDPG), which can be used to deal with continuous values observations and actions, an advantage compared to Q-learning [6].

The methodology proposed in this paper uses Twin Delayed DDPG (TD3) to optimize participation in LEMs and P2P based on Double Auction trading markets. The methodology was incorporated into the Agent-based ecosystem framework for Smart Grid (A4SG). This integration enables the methodology to be applied to real-world scenarios, hence enabling more rigorous testing. To evaluate the proposed methodology, an energy community of 30 players was used, with the results yielding an average savings of 1.54 EUR per player or 40% of the total potential savings.

2. PEER-TO-PEER ENERGY TRADING MODEL

The LEM used in this work implements a P2P energy trading model based on the Double Auction (DA) [7]. In DA buyers and sellers are paired with an effective method that benefits all traders according to their offerings. It is commonly used for trading stocks and energy [8].

A DA market's auction period is predetermined (e.g., hourly resolution in the electricity market). It lets traders place bids/offers at the beginning of an auction period, after which the auctioneer clears the market and announces the public market results (e.g., trading prices and quantities).

Specifically, a DA market consists of a set of buyers and a set of sellers that indicate the amount of energy

This is a paper for 15th International Conference on Applied Energy (ICAE2023), Dec. 3-7, 2023, Doha, Qatar.

they want to trade, in kWh, as well as the price, in EUR/kWh that they provide to the market. Then, an auctioneer creates a public order book in which the approved bids and offers are published, accordingly. Order book queues for purchase orders are ordered by reducing submitted buy prices, whereas order book queues for sell orders are ordered by rising submitted sell prices.

The main motivation for energy players to participate in local energy markets is the financial savings that can exist, both when buying and selling energy. In this way, and to ensure that everyone involved feels motivated to participate, bids are restricted to a range of values, given by the following equation:

$$Price_t^{Sell} < Bid < Price_t^{Buy} \quad (1)$$

where $Price_t^{Sell}$ represent the price that the grid will pay in period t if a player sells energy, and $Price_t^{Buy}$ is the price to pay to the grid in period t to buy energy, both in EUR/kWh.

3. PROPOSED DEEP REINFORCEMENT LEARNING MODEL FOR PEER-TO-PEER ENERGY TRADING

The increasing interest in LEMs and P2P demands optimizing the behavior of the players who participate in them. Thus, the primary purpose of the model proposed in this paper is to improve the players' decision-making regarding the price to be paid and the amount of energy to be transacted. To deal with potential excesses or shortages of energy, several market and individual player parameters, such as the demand forecast and its related error, are evaluated.

Using Deep RL to optimize actions done in energy markets is a frequent practice, mainly because it is a problem that can be modeled by the Markov Decision Process and resembles trial-and-error activities. Thus, this paper proposes the use of the TD3 algorithm to discover the optimal policies for double auction-based energy markets, and its integration in A4SG, a smart grid representation, agent-based ecosystem. The proposed methodology centralizes training in order to consider the actions of all agents, rendering the environment stationary throughout training. On the other hand, real-time execution may be decentralized, with each agent using just local information to perform actions without access to the knowledge of other agents. TD3, which was proposed in [9], is utilized as the Deep RL algorithm in the methodology proposed in this study. Even though DDPG can achieve optimal performance in some contexts, it is often susceptible to the values of hyperparameters and other forms of tuning. A common failure scenario for DDPG is when the learned Q-function begins to significantly overestimate Q-values, resulting in policy violations caused by the exploitation of Q-function mistakes. The TD3 method addresses this issue by proposing three significant strategies: a pair of critic networks, delayed updates, and action noise regularization.

One of the main characteristics of an RL model is its environment, and in this case, it is partially observable since each of the agents has a non-total view of the entire state. Bearing in mind that the environment is competitive (i.e., each agent will learn independently), agents wish to conceal their bids so that other agents do not recognize and duplicate their strategy. The observation of the environment's state in a period t by a player p is given by:

$$o_t^p = (Forecast_t^p, Price_t^{Buy}, Price_t^{Sell}, Transactions_{t-1}^p) \quad (2)$$

where $Forecast_t^p$ represents the demand forecast obtained for player p for period t, in kWh, and $Transactions_{t-1}^p$ represents the transactions completed in the previous period of the energy market by player p. This list only contains the price and quantity of energy traded in each transaction, not containing information about the buyers and sellers.

The action space of the environment, represented in Eq. 3 is composed of two continuous actions, in the range [0,1], one regarding the price, and the other regarding the amount of energy to trade.

$$a_t^p = (aPrice_t^p, aQuantity_t^p) \quad (3)$$

The calculation of the agent's bid price is straightforward. The $aPrice_t^p$ value can be considered a percentage value, placed between the minimum and maximum prices that are accepted in the market as represented in Eq. 4.

$$BidPrice_t^p = aPrice_t^p * \left(Price_t^{Buy} - Price_t^{Sell}\right) + Price_t^{Sell} \quad (4)$$

Regarding the amount of energy to be transacted, there is one more variable to consider, which is the possible error in the player's forecast. The uncertainty that the error brings must be considered so that an agent can learn the best strategy to deal with it. The error is calculated using the forecast model's evaluation metrics at the time of testing. After determining the error, $aQuantity_t^p$ is applied within the possible range of the forecast, represented as follows, where MAE_t^p represents the mean absolute error (MAE), in the moment of training, for player p in period t, in kWh:

$$BidQuantity_t^p = aQuantity_t^p * ((Forecast_t^p + MAE_t^p) - (Forecast_t^p - MAE_t^p)) + (Forecast_t^p - MAE_t^p) \quad (5)$$

For an agent to know whether he performed well in a particular period, he must learn from a reward that represents the consequence of its actions. The proposed reward is directly connected to the decrease of cost or growth of profit of a player with the market's participation. The objective is to maximize the reward, which is provided by the following equations:

$$CostGrid_t^p = Demand_t^p * \begin{cases} Price_t^{Buy}, if\ Role_t^p = Buyer \\ Price_t^{Sell}, if\ Role_t^p = Seller \end{cases} \quad (6)$$

$$CostMarket_t^p = \sum_{i=0}^{N} (TradedEnergy_i * Price_i) \quad (7)$$

$$EnExtra_t^p = \sum_{i=0}^{N} (TradedEnergy_i) - Demand_t^p \quad (8)$$

$$CostExtra_t^p = EnExtra_t^p * \begin{cases} Price_t^{Sell}, if\ Role_t^p = Buyer\ AND\ EnExtra_t^p \geq 0 \\ Price_t^{Sell}, if\ Role_t^p = Seller\ AND\ EnExtra_t^p < 0 \\ Price_t^{Buy}, if\ Role_t^p = Seller\ AND\ EnExtra_t^p \geq 0 \\ Price_t^{Buy}, if\ Role_t^p = Buyer\ AND\ EnExtra_t^p < 0 \end{cases} \quad (9)$$

$$r\_buyer_t^p = 1 - \frac{(CostMarket_t^p - CostExtra_t^p - MinCost_t^p)}{CostGrid_t^p - MinCost_t^p} \quad (10)$$

$$r\_seller_t^p = \frac{(CostMarket_t^p - CostExtra_t^p - CostGrid_t^p)}{MaxProfit_t^p - CostGrid_t^p} \quad (11)$$

$$r_t^p = \begin{cases} r\_buyer_t^p, if\ Role_t^p = Buyer \\ r\_seller_t^p, if\ Role_t^p = Seller \end{cases} \quad (12)$$

where $MinCost_t^p$ is the minimum cost for the buyer p in the period t (i.e., computed by multiplying the demand by the minimum price of the market), and $MaxProfit_t^p$ is the maximum profit for seller p in the period t (i.e., computed by multiplying the surplus energy by the maximum price of the market). To simulate, test, and evaluate the use of the double auction P2P energy trading model in the proposed deep RL algorithm, the A4SG will be used. This solution is able to represent energy communities with a decentralized and distributed, agent-based approach, combining Multi-Agent Systems (MAS) and Agent Communities (ACOM), where a MAS can contain numerous ACOMs, which represent aggregation entities. A4SG's agents use novel approaches to manage the smart grid's dynamic interactions. Branching creates agents that are extensions of the main agents, while mobility allows agents to move to distinct MAS or hosts, allowing a greater adaptation to their context. Representation agents from A4SG have a knowledge base with all their data and behaviors. The architecture of the developed A4SG MAS hosts three ACOMs: one to host the main agents, one to enable agents to engage in forecast services, and one to represent an energy community. The last one, from the energy community, has an additional ACOM to allow RL training. To implement the proposed solution, the ACOM for P2P training is created using the OpenAI Gym framework, enabling the agents that represent energy players to go to this ACOM to train new strategies of P2P trading.

4. CASE STUDY

In this case study, it is tested an energy community with 30 players. The training is carried out from the A4SG ecosystem, where the community representative (ACOM main agent) initiates the training by the players of the community. The agent responsible for training receives data from all agents and starts training. In training, each of the players will have an independent policy that will be used to train their behavior in a LEM. The main goal of the case study is to optimize the participation of the players in the LEM, minimizing their energy bill costs.

Since TD3 is a robust method regarding hyperparameters, i.e., it does not require a significant tuning process, only one run of the model was performed, with the policies and the reward data being saved after every 500 training episodes. Each episode depicts a whole cycle of the market, which operates only when there are both buyers and sellers. The training was complete with 4000 episodes, and the evolution of rewards is shown in Fig. 1 and Fig. 2.

Observing Fig. 1, all agents were able to sustain their positive reward despite the training being competitive, which is the first good conclusion. Examining Fig. 2, it is feasible to discern that the agents' rewards have a positive trend in connection to the three measures under study (maximum, minimum and average). But in fact, these rewards have a real meaning, which is the savings of the players participating in the market. Thus, in Fig. 3 it is possible to observe the decrease in costs or increase in profit for each player. Analyzing the figure, it is possible to see that all players had positive results, including some that went from having costs to making profits (average 1.54EUR savings).

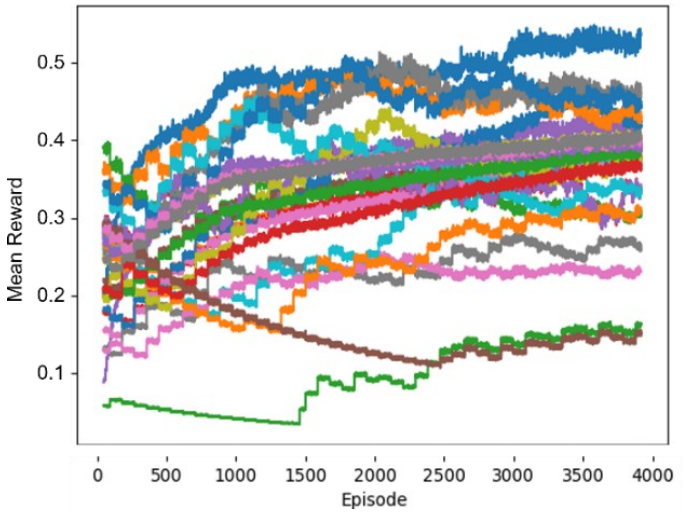

Fig. 1 Rewards evolution: Agents' mean reward.

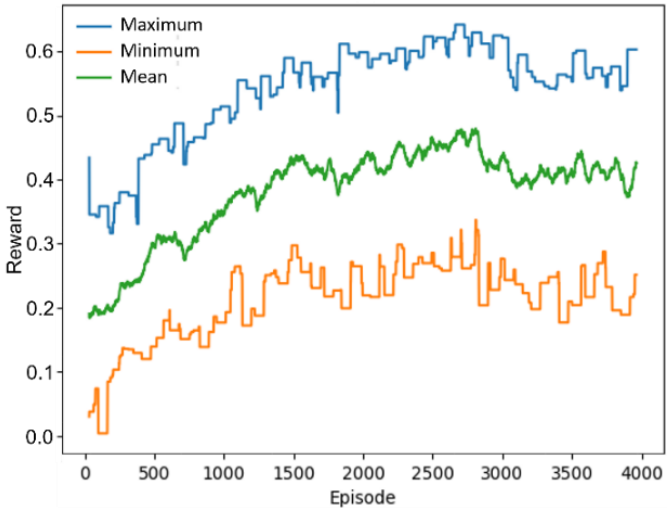

Fig. 2 Rewards evolution: reward per episode

5. CONCLUSIONS

The use of a current reinforcement learning algorithm to maximize energy players' engagement in regional energy markets and peer-to-peer energy sharing is examined in this work. Through a case study in the energy community, the methodology – which employs the twin delayed deep deterministic policy gradient – proved its capacity to improve participation. Each participant saved an average of 1.54 Euros over a week or 40 % of all feasible price reductions.

ACKNOWLEDGEMENT

This work has received funding from the EU Horizon 2020 research and innovation program under project TradeRES (grant agreement No 864276). The authors acknowledge the work facilities and equipment provided by GECAD research center (UIDB/00760/2020) to the project team.

REFERENCE

[1] P. Faria and Z. Vale, "Distributed Energy Resource Scheduling with Focus on Demand Response Complex Contracts," *Journal of Modern Power Systems and Clean Energy*, vol. 9, no. 5, pp. 1172–1182, Sep. 2021, doi: 10.35833/MPCE.2020.000317.

[2] IRENA, "IRENA (2018), Global Energy Transformation: A Roadmap to 2050," 2018.

[3] D. Sun, Z. Liu, J. Shao, and Z. Lin, "Review on low carbon planning and operation of integrated energy systems," *Energy Sci Eng*, Apr. 2022, doi: 10.1002/ESE3.1167.

[4] L. Gomes, Z. A. Vale, and J. M. Corchado, "Multi-Agent Microgrid Management System for Single-Board Computers: A Case Study on Peer-to-Peer Energy Trading," *IEEE Access*, vol. 8, pp. 64169–64183, 2020, doi: 10.1109/ACCESS.2020.2985254.

[5] W. Y. Chiu, C. W. Hu, and K. Y. Chiu, "Renewable Energy Bidding Strategies Using Multiagent Q-Learning in Double-Sided Auctions," *IEEE Syst J*, vol. 16, no. 1, pp. 985–996, Mar. 2022, doi: 10.1109/JSYST.2021.3059000.

[6] J. Li, J. Geng, and T. Yu, "Grid-area coordinated load frequency control strategy using large-scale multi-agent deep reinforcement learning," *Energy Reports*, vol. 8, pp. 255–274, Nov. 2022, doi: 10.1016/J.EGYR.2021.11.260.

[7] D. Friedman, *The double auction market: institutions, theories, and evidence*. Routledge, 2018.

[8] D. Qiu, J. Wang, J. Wang, and G. Strbac, "Multi-Agent Reinforcement Learning for Automated Peer-to-Peer Energy Trading in Double-Side Auction Market," *IJCAI International Joint Conference on Artificial Intelligence*, vol. 3, pp. 2913–2920, Aug. 2021, doi: 10.24963/IJCAI.2021/401.

[9] S. Fujimoto, H. Van Hoof, and D. Meger, "Addressing Function Approximation Error in Actor-Critic Methods," *35th International Conference on Machine Learning, ICML 2018*, vol. 4, pp. 2587–2601, Feb. 2018, doi: 10.48550/arxiv.1802.09477.

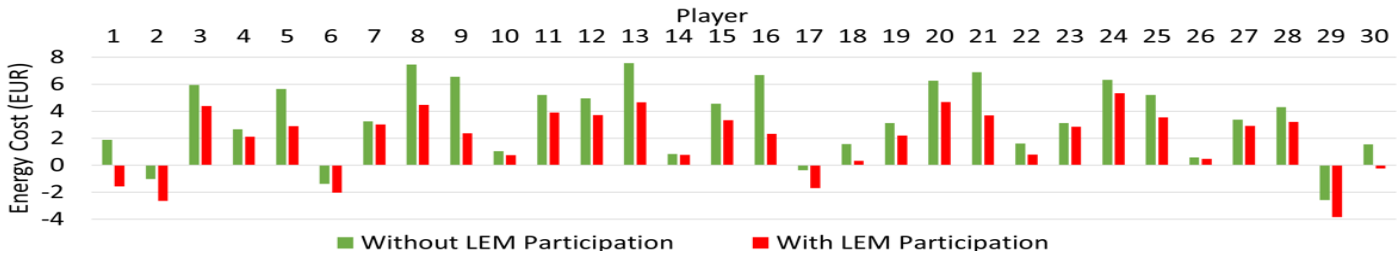

Fig. 3 Players' energy bills using the proposed methodology (with and without LEM)