A Project Report

On

Multi-agent Deep Reinforcement Learning for Peer to Peer Distributed Energy Marketplace in Smart Grids

BY

SIDDHARTH SINGH

2020B5A32029H

Under the supervision of

DR. ALIVELU MANGA PARIMI

SUBMITTED IN COMPLETE FULFILLMENT OF THE REQUIREMENTS OF

EEE F376: DESIGN PROJECT

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI (RAJASTHAN)

HYDERABAD CAMPUS

(SEPTEMBER 2023)

ACKNOWLEDGMENTS

I would like to express my profound gratitude to the individuals and organizations whose unwavering support and invaluable contributions made the realization of this project possible.

My heartfelt thanks go to our supervisor Dr. Alivelu Manga Parimi for their sagacious guidance, encouragement, and profound insights throughout the journey. Your mentorship has been instrumental in shaping our understanding and refining our approach in the intricate field of Multiagent Deep Reinforcement Learning.

I am immensely grateful to the academic institutions whose financial support and resources empowered us to conduct cutting-edge research and push the boundaries of knowledge in the domain of Smart Grids and Distributed Energy Marketplaces. Your investment in my vision has been pivotal to our success.

My appreciation extends to the participants in our research, whose willingness to engage, share insights, and provide critical feedback enriched our understanding and strengthened the empirical foundation of this project. Your participation was invaluable.

I would also like to acknowledge the broader scientific community whose groundbreaking work and discoveries continue to inspire and inform our research. Your pioneering efforts have laid the groundwork for the advancements we have made.

Finally, to my family and loved ones, who stood by us with unwavering support and patience during the long hours and intense moments of this project, we offer our heartfelt thanks. Your understanding and encouragement were our pillars of strength.

In conclusion, this project stands as a testament to the power of collaboration, dedication, and innovation. With deep gratitude, we acknowledge the collective efforts and commitment that have brought us to this milestone.

Thank you all for being an integral part of our journey toward a smarter, more sustainable energy future.

Birla Institute of Technology and Science-Pilani,

Hyderabad Campus

Certificate

This is to certify that the project report entitled "**Multiagent Deep Reinforcement Learning for Peer to Peer Distributed Energy Marketplace in Smart Grids"** submitted by Mr. Siddharth Singh (2020B5A32029H) in complete fulfillment of the requirements of the course EEE F376, Design Project Course, embodies the work done by them under my supervision and guidance.

Date: 7$^{\text{TH}}$ DECEMBER, 2023

(Dr. Alivelu Manga Parimi)
BITS- Pilani, Hyderabad Campus

ABSTRACT

This research introduces an innovative approach called Multi-Agent Deep Reinforcement Learning (MADRL) tailored for complex energy markets in smart grids. It's designed for microgrids where consumers and producers (prosumers) coexist. The key idea is to have real-time, responsive pricing, reducing grid costs and benefiting prosumers economically.

The Deep Q-Network (DQN) framework, a powerful reinforcement learning technique that uses neural networks to approximate the Q-function, is at the core of this ground-breaking MADRL approach. The Q-function acts as a compass, directing agents' choices by calculating the expected future benefits connected to particular behaviors inside different states. The MADRL technique orchestrates the orchestration of optimal pricing and energy consumption strategies for each participating agent using a multi-agent ensemble made up of the grid agent (GA) and numerous prosumer agents (PAs).

This report presents a strong framework for a thorough simulation to be built on the DQN architecture. There are notable increases in the 24-hour cumulative earnings for prosumers and the grid operator, as well as notable decreases in the use of grid reserve power. These results highlight the potential of the proposed MADRL strategy as a game-changing method for developing effective and scalable energy marketplaces within the context of the smart grid.

CONTENTS

CHAPTER - 1

Introduction

The proliferation of Distributed Energy Resources (DERs) is changing the way electricity is produced and posing problems for traditional power grid operations. A sizeable portion of previous customers have adopted DERs and are now proactive "prosumers" who produce and consume energy. Solar photovoltaic (PV) systems for homes are a typical example.

Despite the enormous potential of DER integration, current grid plans frequently fall short because they lack sophisticated control techniques. For instance, household PV systems commonly generate extra electricity during off-peak hours, which may cause the grid to become unstable. Energy storage can be implemented to solve this problem, enabling prosumers to store excess energy for times of high demand and provide crucial grid support.

Regrettably, the existing net metering compensation schemes do not encourage prosumer involvement in grid support activities. Prosumers pay a set retail price for their energy independent of the time of day or the state of the grid, which discourages grid support donations.

This research offers a real-time, demand-responsive, dynamic pricing structure for the distributed energy market. The goal is to establish a win-win situation that encourages prosumer involvement in grid support while boosting financial gains for both prosumers and grid operators.

The proposed energy market makes use of a multi-agent RL framework, which consists of a single grid operator agent and a network of distributed prosumer agents, to achieve these attributes. The goal of the grid agent is to maximize economic gain by choosing how to divide up the purchase of electricity between traditional producing facilities and a group of prosumers with dispatchable generation capabilities. In making this choice, we compared the retail price of buying power from prosumers to the added cost of the producing facilities. Grid agent simultaneously changes retail energy rates in a dynamic manner to encourage prosumers to alter their level of generation. Prosumer agents decide whether to participate in grid assistance based on variables like as energy retail prices, the state of charge of storage devices, the amount of PV generation, household usage, and more in an effort to maximise their economic gain.

CHAPTER - 2

Multiagent Deep R-L Framework

2.1 The Model

A multi-agent deep Reinforcement Learning (MARL) architecture is the foundation for our suggested electricity market model. The Grid Agent (GA) and Prosumer Agents (PAs), in particular, are capable of effective coordination and decision-making thanks to this novel technique.

Agent Interactions

The learning environment, which includes the dynamics of the power grid and prosumer physical systems, is where the GA and PAs interact. Reinforcement learning controls these interactions, allowing agents to make well-informed choices that maximize their own economic gains while considering the larger goals of the electricity market.

Grid Agent (GA)

The GA is essential to the optimisation of the power market because it serves as the major coordinating body. Its main goal is to increase financial gain while maintaining the dependability and stability of the electrical grid. To do this, the GA dynamically distributes power procurement across the set of PAs with dispatchable generation capabilities and conventional generation facilities. When making a decision, it is necessary to weigh the added cost of producing electricity at conventional facilities against the retail price of buying power from PAs. Additionally, the GA dynamically modifies retail energy pricing to encourage PAs to modify their generation levels, hence encouraging effective grid operation.

Prosumer Agents (PAs)

PAs are independent organizations that speak for prosumers in the electricity sector. They want to maximize their personal economic gains while supporting grid-related initiatives. The state of charge (SoC) of their energy storage devices, the amount of PV generation, household energy usage, and other relevant variables are among the important considerations that PAs consider when making judgements. Their level of participation in grid support operations, such as modifying energy generation levels to match grid demand or injecting excess energy into the system during peak hours, is determined by these choices.

Key Features

1. Real-time Decision Making: Agents make decisions in real-time and continuously modify their plans in response to shifting circumstances, such as shifts in the demand and supply for electricity.
2. Demand-Responsive Pricing: Dynamic retail energy pricing keeps energy costs in line with the state of the market, encouraging PAs to modify their patterns of production and consumption as necessary.
3. Intelligent Coordination: The architecture encourages intelligent coordination, which improves grid support services, by efficiently aggregating and coordinating dispatchable prosumer generation capacity.
4. RL-Based Adaptation: The MARL framework uses Reinforcement Learning (RL) approaches to give agents the tools they need to deal with the electrical market's high-dimensional, non-stationary, and stochastic characteristics without the use of intricate explicit modeling.
5. Agent Autonomy: PAs are free to choose their own actions based on their particular conditions, which promotes a diversified and adaptable market environment.

2.2 <u>Algorithm</u>

2.2.1 The Deep Q Network

A crucial element of our reinforcement learning-based agents in the context of the power market is the effective deployment of a Deep Q-Network (DQN). The neural network model for calculating Q-values, a crucial component of our agents' decision-making, is the DQN. This section gives a high-level overview of our DQN implementation's architecture and main features for both prosumer and grid agents.

DQN Model Architecture

The QNetwork class, which houses our DQN model, is carefully constructed to make it easier to estimate Q-values, which stand for the predicted future rewards for particular actions performed in a given state. The structure of it is as follows:

```
class QNetwork(tf.keras.Model):
    def __init__(self, input_dim, output_dim, hidden_layer_sizes):
        super(QNetwork, self).__init__()
        self.input_layer = tf.keras.layers.InputLayer(input_shape=(input_dim,))
        self.hidden_layers = []
        for size in hidden_layer_sizes:
            self.hidden_layers.append(tf.keras.layers.Dense(size, activation='relu'))
        self.output_layer = tf.keras.layers.Dense(output_dim, activation='linear')

    def call(self, inputs):
        x = self.input_layer(inputs)
        for layer in self.hidden_layers:
            x = layer(x)
        return self.output_layer(x)
```

Our DQN design comprises an input layer(which is nothing but states or features), one or more hidden layers (each one contains arbitrary number of neurons which can be later optimized by hyper parameter tuning of number of neurons), and activation functions for rectified linear units (ReLUs). The network's ability to accurately approximate Q-values is improved by this configuration. To anticipate Q-values, the output layer uses a linear activation function.

Agent Functions Utilizing DQN

For making educated decisions, forecasting Q-values, and continuously updating these values during training iterations, our agent functions—deep_q_learning_prosumer_agent and deep_q_learning_grid_agent—rely significantly on the DQN.

These agent functions' principal capabilities include:

1. **Epsilon-Greedy Exploration:** The agents use an exploration method called epsilon-greedy exploration, which strikes a balance between exploration and exploitation. As learning advances, the epsilon value gradually decreases, favoring exploitation.
2. **State Representation:** Current states are retrieved from the environment (s_t_PAj for prosumers and s_t_GA for the grid agent). The DQN is given these states in order to anticipate the Q-value.
3. **Action Selection:** Based on anticipated Q-values, agents select their actions (a_t). They adopt an epsilon-greedy policy, choosing random actions to explore the surroundings with a chance of **'ε',** and we exploit the environment with a chance of **'1-ε'** i.e., we choose from the action which maximizes the cumulative reward at the end of a particular time period or iteration.
4. **Calculation of Rewards:** Based on the feedback from the environment, immediate rewards (r_t_plus_1_PAj for prosumers and r_t_plus_1_GA for the grid agent) are computed.
5. **Next State:** The following states, denoted by the environment as s_t_plus_1_PAj for prosumers and s_t_plus_1_GA for the grid agent, respectively, reflect the states following the execution of the selected actions.
6. **Q-Value Update:** The Q-learning update rule is the cornerstone of our reinforcement learning strategy. With the immediate reward and the highest anticipated Q-value of the following state, we compute the target Q-value. The Q-value linked to the selected action is then updated using this target value. The target value is dynamic, therefore , in each iteration, we are trying to converge our Q-value, corresponding to a particular state, to the target value for that particular state.
7. **Training:** The new goal Q-value and the current condition are used to train the DQN. The target value is dynamic. Therefore, in each iteration, we try to converge our Q-value, corresponding to a particular state, to the target value for that particular state. The neural network's weights are modified during each training iteration, which consists of one epoch every time step, to improve its Q-value estimations.

2.2.1.1 States

In our Multiagent Deep Reinforcement Learning Framework, we emphasize the crucial function of state representation. A "state" in this context refers to an extensive snapshot of the relevant factors and circumstances defining the current environment. It is the basis for the judgments of our prosumer and grid agents. We examine the design and organization of state data for prosumer and grid agents, highlighting key elements and ramifications.

In the context of reinforcement learning and our argument, "state" refers to a comprehensive understanding of the surroundings at a certain time. Consider it as a snapshot of all the important data and factors affecting agent decisions. A state serves as the basis for the observations and judgments of intelligent agents since it is dynamic and changes with the environment.

Prosumer State Representation

The state representation for prosumers in our framework is carefully planned to include important factors that significantly impact their decision-making processes in the dynamic electricity market. It includes the following crucial elements:

1. Battery State: This component offers real-time data on the prosumer battery's current charge status. It is an important factor to consider when choosing how to store and use energy.
2. PV Generation: The display of the electricity produced by a prosumer solar (PV) system provides historical information over a specified time period. This indicates the prosumer's ability to produce renewable energy.
3. Market Price: Prosumers are given information about the current cost of power by the "market_price" component. It influences them by considering economic aspects while deciding on energy generation and use.

Our Multiagent Deep Reinforcement Learning Framework incorporates these meticulously designed state representations, enabling our agents to make context-aware, data-driven decisions. While the grid agent skillfully manages resources to suit the shifting demands of the electrical market, prosumers can optimize their energy use, storage, and generating techniques.

2.2.1.2 Reward

A crucial component of our Multiagent Deep Reinforcement Learning Framework is the calculation of rewards. Our prosumer and grid agents receive input from rewards, which directs their learning and influences the way they make decisions. We explain how these rewards are calculated below by delving into the nuances of reward computation for both prosumer and grid agents.

Prosumer Agent Reward Calculation

The basis of reward calculation is the assessment of the financial benefits or losses arising from prosumer agents' actions and energy-related decisions. The reward for the prosumer agent is determined by the function get_reward_prosumer. The summary is as follows:

1. Virtual Transaction Gain (vt_Hj): The term "virtual transaction gain" (vt_Hj) refers to the total of the electricity price at the current time step and the prosumer's electricity generation (Pt_Hj). It estimates the money made by selling extra energy to the grid or to the consumers or prosumers (we call it revenue).
2. Prosumer Cost: The function determines the price the prosumer must pay for the electricity they use, which is represented by the expression Pt_Hj * rho_t_s. This covers the costs related to using the grid to purchase electricity.
3. Prosumer Agent Reward (rt_PAj): The Prosumer Agent Reward (rt_PAj) is the difference between the virtual transaction gain (vt_Hj) and the prosumer's overall energy-related

costs, which includes the cost of power use. A reward that is positive or negative denotes a profit or loss in money.

Our Multiagent Deep Reinforcement Learning Framework's reward calculation is focused on evaluating the economic effects of agent decisions. In order for our agents to improve their tactics and optimize their behaviour within the dynamic electricity market, it measures the benefits or losses incurred as a result of energy transactions and consumption.

2.2.1.3 Policy

The creation of efficient policies is crucial to our Multiagent Reinforcement Learning Framework because it helps our intelligent agents make decisions in the volatile electricity market. This section explores the subtleties of how we formulate our policies, with an emphasis on the best_buy_sell_price function and how it influences agent behavior.

Policy Formulation

The selection of the best purchasing and selling prices in response to the current market conditions is the basis of our policy formulation. In this process, the best_buy_sell_price function is essential. This is how our policy is put together:

1. Assessment of Market Price variation: We start by analysing the market price variation (data_mp). Variance measures how much the power market's prices fluctuate.
2. Current Market Price (CMP): Using the avg_curerent_market_price function, we determine the average current market price (CMP). This serves as a starting point for the development of our policy.
3. Determining Buy-Sell Price Range: Our approach aims to provide a range of purchase and sell prices that are centred around the current market price (cmp). To establish a baseline, we initialise j as cmp - variance.
4. Iterative Price Assignment: We assign buy and sell prices to a range of values using an iterative procedure that covers a spectrum around the current market price. Every iteration of the loop, which lasts for 2*variance iterations, updates the buy and sell prices appropriately.
5. Storage in Dictionary: For convenience, we keep these purchase and sell price pairings in a dictionary (buy_sell_prices), where the keys stand in for the action (a_t) connected to each pair of prices.
6. Policy Decision: Finally, the policy retrieves the appropriate price from the dictionary to help agents make educated and adaptive decisions when they need a purchase or sell price (as indicated by a_t).

Agent Behavior Guidance

The best_buy_sell_price function, which carries out our policy, directs agent behaviour by giving them suggested buy and sell prices that are based on the current state of the market. To maximize their financial profits, agents utilise this data to fine-tune their purchasing and selling tactics.

In conclusion, the creation of our policies is essential to ensure that our agents function effectively in the electrical market. They are given the tools necessary to make decisions that are in line with their goals, whether they are buying or selling electricity, and to adjust to market swings. The overall effectiveness of our Multiagent Reinforcement Learning Framework is improved by this policy-driven methodology.

Dataset

We used a simulated dataset to mimic the dynamics of a smart grid system for our energy management project. This dataset was created to replicate the kinds of realistic patterns and variances that one could see in an actual energy system. This dataset's main elements include market prices, prosumer power generation, grid power consumption, and buy and sell prices.

- **Grid Buy Prices:** Varied between 9.5 to 11.5, simulating normal market fluctuations.
- **Grid Sell Prices:** Ranged from 12.0 to 13.3, typically higher than buy prices, representing the profit margin for the grid.
- **Grid Power Demand:** Our dataset includes values ranging from 48 to 62, reflecting normal variations in power demand.
- **Market Prices:** The dataset includes a series of market prices ranging from 10.2 to 11.5.
- **Prosumer Power Generation:** This part of the dataset varies from 18 to 26.

Results And Analysis

The energy management project's deep Q-learning model's results offer intriguing new perspectives on the system's dynamic pricing strategy and decision-making process. The model had to figure out the best sell prices across a number of iterations using a variety of inputs that represented various grid and market circumstances.

```
[[-0.2549804   3.197515   -0.6503892  -1.3692111  -0.53796536]]
first itr
Iteration: 0, Sell Price: 10.996000000000004
[[-0.86024535  5.0942774  -1.3112661  -0.85969335 -0.44059134]]
first itr
Iteration: 1, Sell Price: 10.681111111111111
[[-0.15642072  7.527804   -1.004197   -1.4361974   0.41814002]]
first itr
Iteration: 2, Sell Price: 11.019591836734694
[[-0.9556985 10.490094  -2.1127095 -0.8124784  0.5264237]]
first itr
Iteration: 3, Sell Price: 10.8453125
[[-0.84456205 12.84592    -2.4195359  -0.8461314   1.0265332 ]]
first itr
Iteration: 4, Sell Price: 11.01679012345679
[[-1.0252098  14.858348   -3.0823479  -0.36154634  1.4185594 ]]
first itr
Iteration: 5, Sell Price: 11.1304
[[-0.9092402  16.100056   -3.2185178  -0.11038841  1.9653088 ]]
first itr
Iteration: 6, Sell Price: 11.218399999999997
[[-0.35356116 18.948719   -3.5346804  -0.08752727  2.8863688 ]]
first itr
Iteration: 7, Sell Price: 11.206400000000004
[[-1.1565198  22.052988   -4.635723    0.32746598  3.2140434 ]]
first itr
Iteration: 8, Sell Price: 11.042800000000002
[[-0.92253006 22.43575    -4.573546    0.31536308  3.4031851 ]]
first itr
Iteration: 9, Sell Price: 10.959999999999999
[[-1.4949018 26.414051  -5.7244225  0.8926483  3.9205039]]
first itr
Iteration: 10, Sell Price: 10.72
[[-1.3054217 29.3157    -5.7655864  0.8336673  4.074108 ]]
first itr
Iteration: 11, Sell Price: 10.74
[[-2.1573627 35.459007  -7.960169   1.5318393  5.015747 ]]
first itr
Iteration: 12, Sell Price: 10.86
[[-2.2362792 38.07849   -8.522973   1.6820806  5.272852 ]]
first itr
Iteration: 13, Sell Price: 10.800000000000002
[[-2.4500265 40.482647  -9.43632    2.022882   5.6552134]]
first itr
Iteration: 14, Sell Price: 10.94
[[-2.4555733 41.249187  -9.490452   2.214749   5.643989 ]]
```

- Through several iterations, the sell prices varied between 10.6811 and 11.1599.
- The sell prices showed a discernible tendency of fluctuating in reaction to variations in the input data, suggesting that the model was sensitive to the dynamics of the market.
- The model regularly produced sell prices that fell within a fair range, indicating some stability and dependability in its decision-making.

The outcomes show that in a simulated energy market setting, the deep Q-learning model can adapt and make wise decisions. The model's capacity to adapt to shifting market conditions is demonstrated by the variation in sell prices between iterations and the associated Q-values. This encouraging result highlights the models' potential in practical energy management and trading situations.

FUTURE PLANS

1. Data extraction for current market price, jth prosumer power, grid power, grid buy price, grid buy price.
2. Implementation of the real time dataset on the current MADRL model.

CONCLUSION

We set out on a revolutionary trip across the MADRL (Multi-Agent Deep Reinforcement Learning) domains within the framework of a distributed energy market in smart grids in this extensive paper. Our investigation resulted in the creation of a MADRL framework designed specifically to address the difficulties presented by prosumer-dominated microgrids.

Key Learnings:

1. Innovative Framework: To equitably handle the complex interactions of prosumers who concurrently produce and consume electricity, we presented a new MADRL framework. This framework lays the path for a more intelligent and adaptable energy market.
2. Foundation for Deep Q-Networks: The Deep Q-Network (DQN) framework, a reinforcement learning algorithm that uses neural networks to approximate the Q-function, provides a strong basis upon which our strategy is built. The DQN's adaptability is a key component of intelligent agent judgment.
3. Multi-Agent Collaboration: In our system, a variety of agents, such as the grid agent (GA) and prosumer agents (PAs), cooperate to learn and improve pricing and energy consumption strategies. This cooperative endeavor aims to open up new efficiencies in the energy sector.
4. Promising Simulation Results: Our simulations, which are based on the DQN architecture, have produced encouraging findings. Notably, the cumulative revenues over a 24-hour cycle have significantly increased for both prosumers and the grid operator. Furthermore, considerable decreases in the use of grid reserve power emphasize the possibility of cost savings and sustainability.

References

1. A Multi-Agent Deep Reinforcement Learning Approach for a Distributed Energy Marketplace in Smart Grids

   Arman Ghasemi, Amin Shojaeighadikolaei, Kailani Jones, Morteza Hashemi, Alexandru G. Bardas, Reza Ahmadi

2. Peer-to-Peer Energy Trading Mechanism Based on Blockchain and Machine Learning for Sustainable Electrical Power Supply in Smart Grid

   FAISAL JAMIL , NAEEM IQBAL , IMRAN , SHABIR AHMAD , AND DOHYEUN KIM

3. US Energy Department. Consumer vs prosumer: What's the difference? Accessed 5/2020. [Online].Available:https://www.energy.gov/eere/articles/consumer-vs-prosumer-whats-difference.
4. "Annual energy outlook 2019 with projections to 2050," US Energy Information Administration, Tech. Rep., 2019. [Online]. Available: https://www.eia.gov/outlooks/aeo/pdf/aeo2019.pdf.
5. C. Fang, H. Lu, Y. Hong, S. Liu, and J. Chang, "Dynamic pricing for electric vehicle extreme fast charging," IEEE Transactions on Intelligent Transportation Systems, pp. 1–11, 2020.
6. A. Shojaeighadikolaei, A. Ghasemi, K. R. Jones, A. G. Bardas, M. Hashemi, and R. Ahmadi, "Demand responsive dynamic pricing framework for prosumer dominated microgrids using multiagent rein-forcement learning," in The 52nd North American Power Symposium.