Exploring market power using deep reinforcement learning for intelligent bidding strategies

Alexander J. M. Kell, Matthew Forshaw, A. Stephen McGough

School of Computing
Newcastle University
Newcastle upon Tyne, U.K.
{a.kell2, matthew.forshaw, stephen.mcgough}@newcastle.ac.uk

Abstract—Abstract goes here.
Index Terms—deep reinforcement learning, bidding strategy, multi-agent system, electricity markets

I. Introduction

Under perfectly competitive electricity markets, generator companies (GenCos) tend to bid their short-run marginal costs (SRMC) when bidding into the day-ahead electricity market. SRMC is the cost to produce a single MWh of electricity and excludes capital costs. However, electricity markets are often oligopolistic, where a small subset of GenCos provide a majority of the capacity to the market. Under these conditions, it is possible that the assumption that GenCos are price-takers does not hold. That is, large GenCos artificially increase the price of electricity to gain increasing profit using their market power.

Reduced competition within electricity markets can lead to higher prices to the consumers, for no increased societal benefit. It is, therefore, within the interests of the consumer and that of government to maintain a competitive market. Low energy costs can enable innovation in other industries reliant on electricity, and in turn, make a more productive economy. Competition within electricity markets can be decreased through having multiple entities with a relatively small control over capacity in electricity markets.

In this paper, we explore the effect of total controlled capacity on electricity prices. Specifically, we use deep reinforcement learning (RL) to calculate a bidding strategy for GenCos in a day-ahead market. These GenCos are modelled as agents within the agent-based model, ElecSim [1], [2]. We use the UK electricity market instantiated in 2018 as a case study, similar to our work in [3]. That is, we model each GenCo with their respective power plants in the year 2018 to 2019. In total, we model 60 GenCos with 1085 power plants.

We use the deep deterministic policy gradient (DDPG) deep RL algorithm, which allows for a continuous action space [4]. Conventional RL methods require discretization of state or action spaces and therefore suffer from the curse of dimensionality [5]. As the number of discrete states and actions increases, the computational cost grows exponentially. However, too small a number of discrete states and actions will reduce the information available to the GenCos, leading to sub-optimal bidding strategies. Additionally, by using a continuous approach, we allow for GenCos to consider increasingly complex bidding strategies.

Other work considers a simplified model of an electricity market by modelling a small number of GenCos or plants [6], [7]. We, however, model each GenCo as per the UK electricity market with their respective power plants in a day-ahead market. Additionally, other work focuses on a bidding strategy to maximize profit for a GenCo. In our work, we focus on the impact of large GenCos, as well as collusion, between GenCos on total the electricity price.

Our approach does not require GenCos to formulate any knowledge of the information informing the market clearing algorithm or rival GenCo bidding strategies, unlike in game-theoretic approaches [8]. This enables a more realistic simulation where the strategy of rival GenCos are unknown.

In Section II we review the literature, and explore other approaches of RL in electricity markets. In Section III we introduce the agent-based model used and the DDPG algorithm. Section IV explores the methodology taken for our case study. We discuss and conclude our work in Sections VI and VII respectively.

II. Literature Review

Intelligent bidding strategies for day-ahead electricity markets can be divided into two broad categories: game-theoretic models and those based upon simulation and agent-based models (ABMs). ABMs allow for the simulation of heterogenous irrational actors with imperfect information. Additionally, ABMs allow for the modelling of learning and adaption within a dynamic environment [6]. Game-theoretic approaches may struggle in complex electricity markets where Nash equilibriums do not exist [8].

Kumar *et al.* propose a Shuffled Frog Leaping Algorithm (SFLA) [9] to find bidding strategies for GenCos in electricity markets. SFLA is a meta-heuristic that is based on the evolution of memes carried by active individuals, as well as a global exchange of information among the frog population. They test the effectiveness of the SFLA algorithm on an IEEE 30-bus system and a practical 75-bus Indian system. They find superior results when compared to particle swarm optimization and the genetic algorithm with respect to total profit and

convergence with CPU time. They assume that each GenCo bids a linear supply function, and they model the expectation of bids from rivals as a joint normal distribution. In contrast to their work, we do not require an estimation of the rivals bids.

Wang *et al.* propose an evolutionary imperfect information game approach to analyzing bidding strategies with price-elastic demand [8]. Their evolutionary approach allows for GenCos to adapt and update their beliefs about an opponents' bidding strategy during the simulation. They model a 2-bus system with 3 GenCos. Our work, however, models a simulation with 60 GenCos modelled across the entire UK.

The previously discussed approaches take a game-theoretic view. Next, we explore reinforcement learning approaches used to make intelligent bidding decisions in electricity markets. RL is a suitable method for analyzing the dynamic behaviour of complex systems with uncertainties. RL can therefore be used to identify optimal bidding strategies in energy markets [10].

Aliabadi *et al.* utilize an ABM and the Q-learning algorithm to study the impact of learning and risk aversion on GenCos in an oligopolistic electricity market with five GenCos [6]. They find that some level of risk aversion is beneficial, however excessive risk degrades profits by causing an intense price competition. Our paper focuses on the impact of the interaction of many GenCos within the UK electricity market. In addition, we extend the Q-learning algorithm to use the DDPG algorithm, which uses a continuous action.

Bertrand *et al.* use RL in a continuous intraday market. Specifically, they use the REINFORCE algorithm to optimize the choice of price thresholds. They demonstrate an ability to outperform the traditionally used method, the rolling intrinsic method, by increasing profit per day by 4.2%. The rolling intrinsic method accepts any trade, which gives a positive profit if the contracted quantity remains in the bounds of capacity. In our paper, we model a day-ahead market and use a continuous action for price bids, as opposed to a discrete threshold.

Ye *et al.* propose a novel deep RL based methodology which combines the DDPG algorithm with a prioritized experience replay (PER) strategy. The PER samples from previous experience, but samples from the "important" ones more often [11]. They use a day-ahead market as a case study with hourly resolution and show that they are able to achieve approximately 41%, 20% and 11% higher profit for the GenCo than the MPEC, Q-learning and DQN methods, respectively. In our paper, we inspect the effect on electricity price within the UK for increasing sized groups of GenCos.

In [12], Zhao *et al.* propose a modified RL method, known as the gradient descent continuous Actor-Critic (GDCAC) algorithm. This algorithm is used in a double-side day-ahead electricity market simulation. Where in this case, a double-side day-ahead market refers to GenCos selling their supply to distribution companies, retailers or large consumers. Their approach performs better in terms of participant's profit or social welfare compared with traditional RL methods.

III. MATERIAL

In this section we describe the RL methodology used for the intelligent bidding process as well as the simulation model used as the environment.

A. Reinforcement Learning

1) RL Background: Generally in RL, an agent interacts with an environment to maximize its cumulative reward. Generally, RL can be described as a Markov Decision Process (MDP). An MDP includes a state space $\mathcal{S}$, action space $\mathcal{A}$, a transition dynamics distribution $p(s_{t+1}|s_t, a_t)$ and a reward function, where $r : S \times \mathcal{A} \to \mathbb{R}$. At each time-step an agent receives an observation.

An agent's behaviour is defined by a policy, π. π maps states to a probability distribution over the actions $\pi : \mathcal{S} \to \mathcal{P}(\mathcal{A})$. The return from a state is defined as the sum of discounted future reward $R_t = \sum_{i=t}^{T} \gamma^{(i-t)} r(s_i, a_i)$. Where γ is a discounting factor $\gamma \in [0, 1]$. The return is dependent on the action chosen, which is dependent on the policy π. The goal in reinforcement learning is the learn a policy that maximizes the expected return from the start distribution $J = \mathbb{E}_{r_i, s_i \sim E, a_i \sim \pi}[R_1]$.

The expected return after taking an action a_t in state s_t after following policy π can be found by the action-value function. The action-value function is used in many reinforcement learning algorithms and is defined in Equation 1

$$\mathcal{Q}^\pi(s_t, a_t) = \mathbb{E}_{r_{i \geq t}, s_{i > t} \sim \mathcal{E}, a_{i > t} \sim \pi}[R_t | s_t, a_t] \qquad (1)$$

An optimal policy can be derived from the optimal Q-values $Q_*(s_t, a_t) = \max_\pi Q_\pi(s_t, a_t)$ by selecting the action corresponding to the highest Q-value in each state.

B. Q-Learning

- Introduction to RL and DDPG (model-free approach and continuous action space)
- Problems with Q-learning not using entire continuous space.
- Explore our parameters of DDPG
-Note that in the model, the GenCo does not take any strategic action, that is, the GenCo does not consider the actions of other GenCos explicitly in its decision process. In fact, it does not have information on other GenCos. The GenCo is modeled as a simple agent that learns only from its own experi- ence. GenCos' collective behavior, however, may lead to strategic outcomes. [6]

- Market Structure of ElecSim (yearly outlook)
- Real life structure of the UK
- Assumed that all capacity is bid
- We change the price of the bid (from SRMC to Bid cap in the market)
- The GenCo has no information about the generation capacity, marginal cost, bid prices or profits of other GenCos, or the total number of GenCos in the system. [6]

- Do not consider flow constraints due to the large nature of the simulation
- Amount of time taken to run 1 episode is...

IV. Methodology

- Grouping agents based upon size and seeing results
- Observation and action space
- Allowing them to bid maximum of £600 and a market cap of £150

V. Results

- Show time-steps vs. reward for both scenarios
- Show the step change in reward after a certain amount of controlled capacity

VI. Discussion

- Make suggestions based upon optimal level of competition
- Importance of understanding market power and having a regulator otherwise prices can significantly increase

VII. Conclusion

- Future work (withold capacity)

VIII. Acknowledgment

References

[1] A. Kell, M. Forshaw, and A. S. Mcgough, "ElecSim : Monte-Carlo Open-Source Agent-Based Model to Inform Policy for Long-Term Electricity Planning," *The Tenth ACM International Conference on Future Energy Systems (ACM e-Energy)*, pp. 556–565, 2019.

[2] A. J. M. Kell, M. Forshaw, and A. S. McGough, "Long-Term Electricity Market Agent Based Model Validation using Genetic Algorithm based Optimization," *The Eleventh ACM International Conference on Future Energy Systems (e-Energy'20)*, 2020.

[3] A. Kell, M. Forshaw, and A. S. McGough, "Modelling carbon tax in the UK electricity market using an agent-based model," *e-Energy 2019 - Proceedings of the 10th ACM International Conference on Future Energy Systems*, no. Ldc, pp. 425–427, 2019.

[4] J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous learning control with deep reinforcement," *International Conference on Learning Representations (ICLR)*, 2016.

[5] Y. Ye, D. Qiu, M. Sun, D. Papadaskalopoulos, and G. Strbac, "Deep Reinforcement Learning for Strategic Bidding in Electricity Markets," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1343–1355, 2020.

[6] D. Esmaeili Aliabadi, M. Kaya, and G. Sahin, "Competition, risk and learning in electricity markets: An agent-based simulation study," *Applied Energy*, vol. 195, pp. 1000–1011, 2017. [Online]. Available: http://dx.doi.org/10.1016/j.apenergy.2017.03.121

[7] A. C. Tellidou and A. G. Bakirtzis, "Agent-based analysis of monopoly power in electricity markets," *2007 International Conference on Intelligent Systems Applications to Power Systems, ISAP*, 2007.

[8] J. Wang, Z. Zhou, and A. Botterud, "An evolutionary game approach to analyzing bidding strategies in electricity markets with elastic demand," *Energy*, vol. 36, no. 5, pp. 3459–3467, 2011. [Online]. Available: http://dx.doi.org/10.1016/j.energy.2011.03.050

[9] J. Vijaya Kumar and D. M. Kumar, "Generation bidding strategy in a pool based electricity market using Shuffled Frog Leaping Algorithm," *Applied Soft Computing Journal*, vol. 21, pp. 407–414, 2014. [Online]. Available: http://dx.doi.org/10.1016/j.asoc.2014.03.027

[10] T. Yang, L. Zhao, W. Li, and A. Y. Zomaya, "Reinforcement learning in sustainable energy and electric systems: a survey," *Annual Reviews in Control*, vol. 49, pp. 145–163, 2020.

[11] T. Schaul, J. Quan, I. Antonoglou, and D. Silver, "Prioritized experience replay," *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*, pp. 1–21, 2016.

[12] H. Zhao, Y. Wang, S. Guo, M. Zhao, and C. Zhang, "Application of a Gradient Descent Continuous Actor-Critic Algorithm for Double-Side Day-Ahead Electricity Market Modeling," *Energies*, vol. 9, no. 9, p. 725, 2016.