Proceedings of the 2019 Winter Simulation Conference
N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, eds.

BIDDING STRATEGY SIMULATION IN THE DAY AHEAD ELECTRICITY MARKET VIA REINFORCEMENT LEARNING

Kristen R. Schell

Rensselaer Polytechnic Institute
Department of Industrial & Systems Engineering
Rensselaer Polytechnic Institute
Troy, NY 12180 USA

Fangyuan Chang

University of Michigan - Dearborn
Department of Electrical and Computer Engineering
University of Michigan - Dearborn
Dearborn, MI 48128 USA

ABSTRACT

The day-ahead electricity market is the central planning mechanism for dispatching electric power. In a perfectly competitive market, the equilibrium strategy is to bid at marginal cost, driving down profit to zero. The recently restructured electricity market, however, is an oligopolistic market at best, where a few entrenched utilities can exercise market power to manipulate the DAM price. Traditionally, such a market is modeled as reaching an optimal, Nash equilibrium price for electricity. We simulate market players' bidding strategies via reinforcement learning. We show that a Q-learning algorithm accurately models the Nash equilibrium, no matter the number of Nash equilibria. However, it takes players over one year of experimenting with bidding strategies to achieve these optimal outcomes. Future work is focused on replicating this result with real market data, in order to assess market power and the existence of Nash equilibria in the real world.

1 INTRODUCTION

The day-ahead electricity market (DAM) is the main mechanism for electricity pricing in restructured electricity markets, which are being developed all over the world - from the European Union to California to Ontario. Such restructuring has come under scrutiny, however, for the ease with which a poor market design can enable firms to obtain and severely abuse market power (Borenstein and Bushnell 2015). In order to assess and prevent market power, it would be helpful to know firms' bidding strategies. Such information, however, is proprietary. Past work on analyzing strategic behavior in restructured markets has focused on modeling the existence of equilibria via game theory (Hobbs and Metzler 2000). More recent work has analyzed various decision rules within agent- based models (Krause et al. 2006, Aliabadi et al. 2017). We focus on analyzing a Q-learning algorithm to model firms' bidding strategies in the DAM, to evaluate whether this algorithm can accurately find the optimal strategy for each player - the Nash equilibrium (or Nash equilibria, depending on the number of players in the market). If the Q-learning algorithm is successful at finding the Nash equilibrium/a this model can be utilized in real-world market settings to evaluate market power, and develop policies to mitigate against it.

2 DATA AND METHODS

The power system under study is the modified benchmark IEEE 30-bus system (Christie 1993). In this system, generating firms at bus 2 and bus 13 operate with the same equipment and therefore have the same set of bidding strategies from which to choose. In a real-world setting, producers submit price-quantity bids into the DAM for every hour of the next 24-hour period. In our simulation, we assume players in the market can submit three price-quantity bids that represent production over an eight-hour period.

Each firm selects its bidding strategy from a pre-defined set of strategies, which represent an increasing mark-up in price from marginal cost. We assume each player selects a strategy according to an ε-Greedy policy, defined in Figure 1. The player will choose action a_i which maximizes its expected profit, $Q_i(a^t{}_i)$, with probability ε. As the player does not know if the strategy that maximizes his reward will actually maximize his profit in the market, against other players' strategies, he explores the reward space by randomly choosing a different strategy with probability 1 − ε. Players learn the rewards associated with their strategies after the market clears. The DAM clears according to the classic market clearing optimization problem, which prioritizes the lowest price bids that can meet demand.

Algorithm 1: Q-learning with ϵ-Greedy policy

initialization $t = 0$;
initialization $Q_i(a_i) = 0 \quad \forall i \in N,\ a_i \in A_i$;
while $t < 3000$ **do**
  $t \leftarrow t + 1$;
  Choose action a_i^t with ϵ-Greedy policy, $\forall i \in N$;
  Clear market with joint actions $(a_1^t, ..., a_n^t)$, $\forall i \in N$,;
  Calculate profit $r_i(a_1^t, ..., a_n^t)$, $\forall i \in N$;
  Update expected reward function $Q_i(a_i^t) \leftarrow Q_i(a_i^t) + \alpha_i^t(r_i(a_1^t, ..., a_n^t) - Q_i(a_i^t))$,
end

Figure 1: Q-learning algorithm.

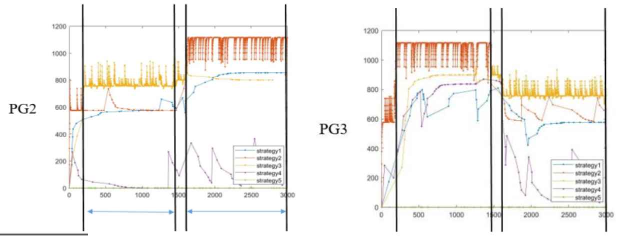

Figure 2: Market players find both Nash equilibria.

3 RESULTS

We consider various bidding strategies for each player in the market. For the simulation results shown in the graphs above, we consider only two market players/generating firms (PG2 and PG3). Initially, each player, PG2 and PG3, can choose from selection of five strategies - [0, 20, 40, 60, 80] ($/MW) - which represents a mark-up of 20 ($/MW) in each block/time period. Simulating over 3,000 trials, we find that PG2 finds a high reward strategy after around 200 attempts, as seen in Figure 2. Similarly, PG3 plays its own high reward strategy in response to PG2 (strategy 2). Eventually, PG2 finds a higher reward strategy (2), and continues to play this strategy. Simultaneously, PG3 plays its best response strategy (3). By explicitly enumerating the Nash equilibria for the available bidding strategies, it shown that both players find the Nash equilibria using the Q-learning algorithm. By comparing the strategies to the enumerated rewards, we find that three out of the four Nash equilibria are found. It is possible that the fourth equilibrium could be found if the simulation was allowed to run for more than 3,000 hourly time steps. In a sensitivity analysis of the learning rate, α, and the ε value, we find that the Nash equilibria are found faster when $0.1 \leq \alpha \leq 0.5$, and $\varepsilon \leq 0.3$. These rates represent the optimal tradeoff between exploitation and exploration.

4 CONCLUSIONS

We find that a Q-learning algorithm can accurately represent strategic behavior in the restructured day-ahead electricity market. If the players are allowed to explore the strategy space for long enough, all Nash equilibria are likely to be found. It is hypothesized that Q-learning can be utilized to assess market power and design market rules to prevent exploitation of this power.

REFERENCES

Aliabadi, D.S., M. Kaya, and G. Sahin. 2017. "An Agent-Based Simulation Of Power Generation Company Behavior In Electricity Markets Under Different Market-Clearing Mechanisms." *Energy Policy* 100:191 – 205.

Borenstein, S. and J. Bushnell. 2015. "The U.S. Electricity Industry After 20 Years Of Restructuring." *Annual Review of Economics* 7:437-463.

Hobbs, B.F. and C.B. Metzler. 2000. "Strategic Gaming Analysis For Electric Power Systems: An MPEC Approach." *IEEE Transactions on Power Systems* 15(2):638 – 645.

Krause, T., E.V. Beck, R. Cherkaoui, A. Germond, G. Andersson, and D. Ernst. 2006. "A Comparison Of Nash Equilibria Analysis And Agent-Based Modeling For Power Markets." *Electrical Power & Energy Systems* 28:599 – 607.

Christie, R. 1993. *Power Systems Test Case Archive*. https://labs.ece.uw.edu/pstca/pf30/pg_tca30bus.htm, accessed 16th June 2018.