Recurrent Deep Multiagent Q-Learning for Autonomous Agents in Future Smart Grid

Extended Abstract

Yaodong Yang, Jianye Hao*, Zan Wang
yydapple@gmail.com,{jianye.hao,wangzan}@tju.edu.cn
School of Computer Software, Tianjin University
Tianjin, China

Mingyang Sun, Goran Strbac
{mingyang.sun11,g.strbac}@imperial.ac.uk
Imperial College London
London, England

ABSTRACT

The broker mechanism is widely applied to serve for interested parties to derive long-term policies to reduce costs or gain profits in smart grid. However, brokers are faced with a number of challenging problems such as balancing demand and supply from customers and competing with other coexisting brokers to maximize profits. In this paper, we develop an effective pricing strategy for brokers in local electricity retail market based on recurrent deep multiagent reinforcement learning and sequential clustering. We use real household electricity consumption data to simulate the retail market for evaluating our strategy. The experiments demonstrate the superior performance of the proposed pricing strategy and highlight the effectiveness of our reward shaping mechanism.

KEYWORDS

Deep Multiagent Reinforcement Learning; Smart Grid; Pricing Strategy

ACM Reference Format:
Yaodong Yang, Jianye Hao*, Zan Wang and Mingyang Sun, Goran Strbac. 2018. Recurrent Deep Multiagent Q-Learning for Autonomous Agents in Future Smart Grid. In *Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), Stockholm, Sweden, July 10–15, 2018,* IFAAMAS, 3 pages.

1 INTRODUCTION

Traditional power grid is faced with fundamental changes with the advent of decentralized power generation technologies and the increasing number of active electricity customers. One critical objective of smart grid is to guarantee its stability and reliability in terms of the real-time balance between demand and supply. Nevertheless, with the increasing penetration of renewable energy resources, existing centralized control mechanisms are unable to simultaneously accommodate the vast number of small-scale intermittent producers and dynamic changes in demand of customers in response to price variations [5, 13].

One promising way of maintaining real-time balance in local tariff market is to apply electricity retail brokers, which offer tariff contracts for both local consumers and small-scale producers. To satisfy the demand of the contracted customers, retail brokers are challenged by optimizing their trading strategies to maximize their profits while balancing demand and supply [21]. Power TAC [9], as a rich, competitive, open-source simulation platform, is commonly used to design and evaluate various broker strategies. However, brokers developed on Power TAC mainly purchase electricity from remote power plants, in which traditional fossil fuel is still the main generation resource via a wholesale market. Small-scale producers are usually overlooked in local retail market [10, 18, 19].

In a retail market, as the environment can be modeled as a Markov decision process (MDP) [14], reinforcement learning techniques have been applied to learn electricity broker strategies for customers and retailers [1, 3, 13–16]. However, existing works of retail brokers [13, 14, 20] are based on the simple Q-table structure or a linear function approximation, where features are approximated as discrete values and may need to be constructed manually which result in information loss. On the other side, electricity customers exhibit various electricity consumption or producing patterns. In [20], its broker framework assigns each type of customers with an independent pricing agent. However, they use independent SARSA for different customers and regard the whole broker's profit as each agent's immediate reward during its Q-value update process. It does not distinguish each agent's unique contribution to the broker's profits and thus does not encourage the learning towards optimal strategy.

To address above problems, we propose a recurrent deep multiagent reinforcement learning (RDMRL) framework augmented with sequential clustering and reward shaping to coordinate internal sub-brokers. Experimental results show that superior performance of our RDMRL broker and highlight the effectiveness of our reward shaping mechanism.

2 RDMRL FRAMEWORK

Figure 1 shows the structure of our multiagent-based reinforcement learning framework. Customers with various electricity consumption patterns are clustered into groups, detailed in Section 2.2. Then an individual recurrent Deep Q-Network (DQN) is employed to handle the continuous state space explosion problem for each type of customers detailed in Section 2.1. Finally, a reward shaping mechanism (Section 2.3) is proposed to allocate the correct reward signal for each sub-broker.

2.1 Learning Framework of Sub-brokers

Deep reinforcement learning (DRL) has recently been shown to master numerous complex problem domains that suffer from the curse of dimensionality [6, 12]. It is expected to learn more efficient pricing policies by employing DRL into the broker pricing domain.

* Corresponding author: Jianye Hao.

Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), M. Dastani, G. Sukthankar, E. André, S. Koenig (eds.), July 10–15, 2018, Stockholm, Sweden.

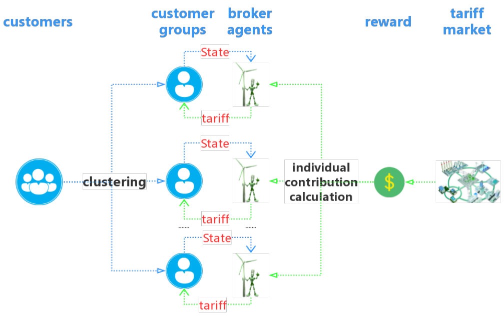


Figure 1: RDMRL broker framework

Meanwhile, we apply recurrent neural units such as Long Short-Term Memory (LSTM) [7] to capture the temporal information of the tariff market.

2.2 Clustering Consumers

We cluster consumers according to their temporal electricity consumption patterns by K-Means [11] with Dynamic Time Warping (DTW) distance criterion [8], which is a state-of-the-art clustering algorithm for measuring temporal sequence similarity.

2.3 RDMRL Broker with Reward Reshaping

To address the multiagent credit assignment problem [2] of a cooperative multiagent system, we calculate each sub-broker's individual credit by difference rewards [17]:

$$r_t^i = r_t - (\sum_{j \neq i} p_t^j \psi_{t,C}^j - \sum_{k \neq i} p_t^k \psi_{t,P}^k - \Phi_t^i), j \in C, k \in P \qquad (1)$$

where i represents the customer type charged by sub-broker i, r_t is the broker's reward, $\psi_{t,C}^j$ denotes consumptions of consumers of type j at time t, $\psi_{t,P}^k$ denotes outputs of producers of the type k at time t. Also, p_t^j and p_t^k are the broker's current tariff prices for C_j and P_k respectively. Φ_t^i is current imbalance fee:

$$\Phi_t = \begin{cases} \phi_-(\sum_{j \neq i} \psi_{t,C} - \sum_{k \neq i} \psi_{t,P}), & if \sum_{j \neq i} \psi_{t,C} \geq \sum_{k \neq i} \psi_{t,P} \\ \phi_+(\sum_{j \neq i} \psi_{t,C} - \sum_{k \neq i} \psi_{t,P}), & otherwise \end{cases} \qquad (2)$$

where ϕ_- and ϕ_+ are the imbalance fee for shortsupply and oversupply respectively.

3 EXPERIMENTS AND ANALYSIS

In this section, we first compare a DQN based broker with a Q-table based broker by following the settings in [14] to demonstrate the superior performance of DQN. The other co-existing strategies are set to be the same as previous work [14]: *Balanced Strategy*, *Greedy Strategy*, *Random Strategy* and *Fixed Strategy*. Table 1 shows the detailed results of the first experiment. We can see the profit of DQN based B_L is 105% higher than Q-table based B_L and its imbalance amount is reduced by 10%.

In the second set of experiments, we compare the RDMRL broker with two baseline brokers (single-agent recurrent DQN and multiple independent recurrent DQNs) against the same set of competing brokers as the first experiment. We consider more realistic settings by introducing real-world data [4] to model consumer consumption patterns. First, Figure 2 shows the accumulated profits of single agent broker during the evaluation episodes, which fails to gain profits because of the single pricing method. The result indicates that the broker using only one recurrent DQN cannot learn effective pricing strategies in a realistic retail market with various consumers of different electricity consumption patterns.

Table 1: Q-table Based B_L and Other Brokers' Total Profits

Broker	Profits	ShortSupply	OverSupply
$Tabular - Q$	1327482$	-244764kWh	313536kWh
DQN	2721828$	-226826kWh	275564kWh

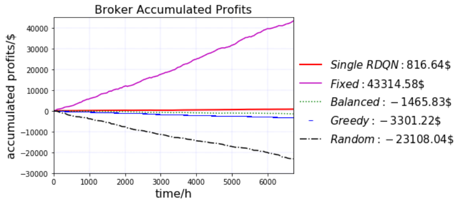
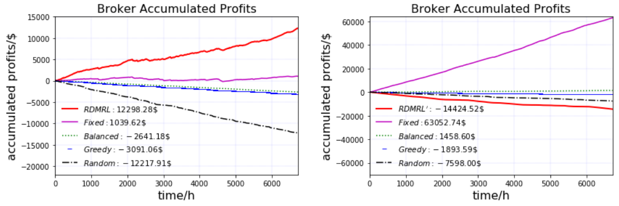


Figure 2: Brokers' Profits of the evaluation episodes.

Second, Figure 3(a) shows the profits of RDMRL broker. We can observe that RDMRL broker gains the most profits. Lastly, Figure 3(b) shows the profits of an incomplete version of RDMRL broker (denoted as RDMRL') by removing reward shaping. The result shows that RDMRL' using the global reward instead of reward shaping performs worst.



(a) RDMRL with reward shaping (b) RDMRL' without reward shaping

Figure 3: Brokers' profits of the evaluation episodes.

4 CONCLUSIONS

In this paper, we firstly apply recurrent DQN for retail broker design to learn more accurate and effective pricing strategies in smart grid domain. Also, we design a multiagent broker framework with reward shaping to publish tariffs for each group of customers. Extensive simulations validate the the superior effectiveness of our broker framework.

ACKNOWLEDGMENTS

The work is supported by the National Natural Science Foundation of China under Grant No.: 61702362.

REFERENCES

[1] Angelos Angelidakis and Georgios Chalkiadakis. 2015. Factored MDPs for Optimal Prosumer Decision-Making in Continuous State Spaces. In *the Proceedings of the 13th European Conference on Multi-Agent Systems*. 91–107.

[2] Yu Han Chang, Tracey Ho, and Leslie Pack Kaelbling. 2003. All Learning is Local: Multi-agent learning in global reward games. In *Proceedings of the 16th Advances in Neural Information Processing Systems*. 807–814.

[3] Moinul Morshed Porag Chowdhury, Russell Y. Folk, Ferdinando Fioretto, Christopher Kiekintveld, and William Yeoh. 2015. Investigation of Learning Strategies for the SPOT Broker in Power TAC. In *Proceedings of the 17th International Workshop on Agent-Mediated Electronic Commerce and Trading Agents Design and Analysis*.

[4] Energy Consumption Data 2015. Electricity Consumption in a Sample of London Households. (2015). https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households.

[5] Xi Fang, Satyajayant Misra, Guoliang Xue, and Dejun Yang. 2012. Smart Grid - The New and Improved Power Grid: A Survey. *IEEE Communications Surveys and Tutorials* 14, 4 (2012), 944–980.

[6] Shixiang Gu, Ethan Holly, Timothy Lillicrap, and Sergey Levine. 2017. Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates. In *Proceedings of IEEE International Conference on Robotics and Automation*. 3389–3396.

[7] S Hochreiter and J Schmidhuber. 1997. Long short-term memory. *Neural Computation* 9, 8 (1997), 1735–1780.

[8] Eamonn Keogh and Chotirat Ann Ratanamahatana. 2005. Exact indexing of dynamic time warping. *Knowledge and Information Systems* 7, 3 (2005), 358–386.

[9] Wolfgang Ketter, Markus Peters, and John Collins. 2013. Autonomous agents in future energy markets: the 2012 power trading agent competition. In *Proceedings of the 27th Conference on Artificial Intelligence*. 1298–1304.

[10] Bart Liefers, Jasper Hoogland, and La Poutré Han. 2014. A Successful Broker Agent for Power TAC. *Lecture Notes in Business Information Processing* (2014), 99–113.

[11] J Macqueen. 1967. Some Methods for Classification and Analysis of MultiVariate Observations. In *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*. 281–297.

[12] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin A. Riedmiller, Andreas Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. 2015. Human-level control through deep reinforcement learning. *Nature* 518 7540 (2015), 529–33.

[13] Markus Peters, Wolfgang Ketter, Maytal Saar-Tsechansky, and John Collins. 2013. A reinforcement learning approach to autonomous decision-making in smart electricity markets. *Machine Learning* 92 (2013), 5–39.

[14] Prashant P. Reddy and Manuela M. Veloso. 2011. Strategy learning for autonomous agents in smart grid markets. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence*. 1446–1451.

[15] Prashant P. Reddy and Manuela M. Veloso. 2013. Negotiated learning for smart grid agents: entity selection based on dynamic partially observable features. In *the Proceedings of 27th AAAI Conference on Artificial Intelligence*. 1313–1319.

[16] Valentin Robu, Meritxell Vinyals, Alex Rogers, and Nicholas R. Jennings. 2014. Efficient Buyer Groups for Prediction-of-Use Electricity Tariffs. In *the Proceedings of the 28th AAAI Conference on Artificial Intelligence*. 451–457.

[17] Kagan Tumer and Adrian Agogino. 2007. Distributed agent-based air traffic flow management. In *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems*. 255.

[18] Daniel Urieli and Peter Stone. 2014. TacTex'13: a champion adaptive power trading agent. In *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems*. 1447–1448.

[19] Daniel Urieli and Peter Stone. 2016. An MDP-Based Winning Approach to Autonomous Power Trading: Formalization and Empirical Analysis. In *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems*. 827–835.

[20] Xishun Wang, Minjie Zhang, and Fenghui Ren. 2016. A Hybrid-learning Based Broker Model for Strategic Power Trading in Smart Grid Markets. *Knowledge-Based Systems* 119 (2016), 142–151.

[21] Kazem Zare, Mohsen Parsa Moghaddam, and Mohammad Kazem Sheikh-El-Eslami. 2011. Risk-Based Electricity Procurement for Large Consumers. *IEEE Transactions on Power Systems* 26, 4 (2011), 1826–1835.