ANOTO: Improving Automated Negotiation via Offline-to-Online Reinforcement Learning

Extended Abstract

Siqi Chen[*]
School of Information Science and Engineering, Chongqing Jiaotong University
Chongqing, China
siqichen@cqjtu.edu.cn

Jianing Zhao
College of Intelligence and Computing, Tianjin University
Tianjin, China

Kai Zhao
College of Intelligence and Computing, Tianjin University
Tianjin, China

Gerhard Weiss
Department of Advanced Computing Sciences, Maastricht University
Maastricht, the Netherlands

Fengyun Zhang
College of Intelligence and Computing, Tianjin University
Tianjin, China

Ran Su
College of Intelligence and Computing, Tianjin University
Tianjin, China

Yang Dong
School of Information Science and Engineering, Chongqing Jiaotong University
Chongqing, China

Daqian Li
School of Information Science and Engineering, Chongqing Jiaotong University
Chongqing, China

Kaiyou Lei
College of Computer and Information Science, Southwest University
Chongqing, China

ABSTRACT

Automated negotiation is a crucial component for establishing cooperation and collaboration within multi-agent systems. While reinforcement learning (RL)-based negotiating agents have achieved remarkable success in various scenarios, they still face limitations due to certain assumptions on which they are based. In this work, we proposes a novel approach called ANOTO to improve the negotiating agents' ability via offline-to-online RL. ANOTO enables a negotiating agent (1) to communicate with opponents using an end-to-end strategy that covers all negotiation actions, (2) to learn negotiation strategies from historical offline data without requiring active interactions, and (3) to enhance the optimization process during the online phase, facilitating rapid and stable performance improvements for the learned offline strategies. Experimental results, based on a number of negotiation scenarios and recent winning agents from the Automated Negotiating Agents Competitions (ANAC), are provided.

KEYWORDS

Automated negotiation, E-commence, Reinforcement learning

ACM Reference Format:

Siqi Chen, Jianing Zhao, Kai Zhao, Gerhard Weiss, Fengyun Zhang, Ran Su, Yang Dong, Daqian Li, and Kaiyou Lei. 2024. ANOTO: Improving Automated Negotiation via Offline-to-Online Reinforcement Learning: Extended Abstract. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024*, IFAAMAS, 3 pages.

1 INTRODUCTION

Thanks to the popularity of deep learning and reinforcement learning, these approaches have been successfully deployed in a wide range of scenarios [9, 16–19]. Automated negotiation [1, 2, 5, 6, 11] is now focusing on deep reinforcement learning-based negotiating agents [3, 8] because of their ability to adapt to different scenarios and opponents [4, 7, 12, 15, 20]. However, continuous interaction with opponents is often impractical and unrealistic, especially in real-world applications where data collection can be challenging and costly [10, 14]. Consequently, designing autonomous agents that can learn effective negotiation strategies without online interactions remains a significant open problem. Another important challenge is adapting to the behavior of opponents, as they may change their strategy in subsequent negotiations. The contributions of this work are two-fold: a novel approach is proposed for bilateral multi-issue negotiation via Offline-to-Online reinforcement learning; experimental results also show that our approach is effective.

2 METHODOLOGY

The diagram of our whole framework is shown in Fig. 1. The agent is based on an end-to-end policy network, which takes the historical bids of both parties as input and outputs the corresponding negotiation action. The negotiation action consists of bids (ω_s) and acceptance signals (χ). Thus, a negotiation action is denoted by a combination of bids $\omega = (o_1, o_2, ..., o_n)$ and acceptance signals $\chi \in \{0, 1\}$, where o_k means the agent's choice for the k-th issue. These actions form a factored action space $\mathcal{A}$, which can be formally represented

[*]Corresponding author: Siqi Chen

Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand. © 2024 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

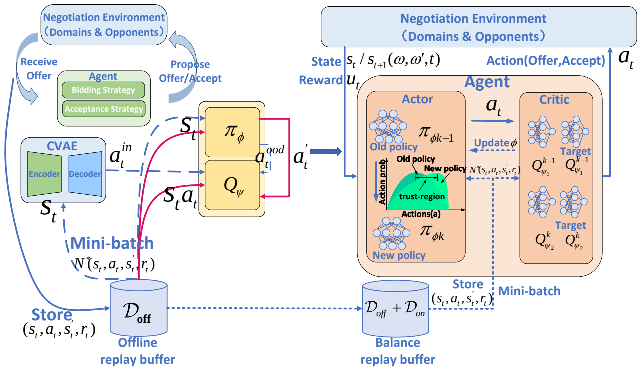

Figure 1: Framework of the proposed approach.

as the Cartesian product $\mathcal{A} = \chi \bigotimes_{i=1}^{N} \mathcal{A}_n = \chi \times \mathcal{A}_1 \times \cdots \times \mathcal{A}_N$, where $\mathcal{A}_k = \{o_m | m = 1, \ldots, |i_k|\}$ denotes the choice on the k-th issue i_k. We design a new neural network architecture in which we distribute the representation of actions across multiple sub-policy networks. Each sub-policy network is responsible for a specific action dimension while sharing a decision module to extract features from input states and coordinate with other sub-policy networks. Each network is responsible for a specific action dimension, at the same time, maintaining a shared decision module among them to extract features from input states and coordinate each sub-policy networks.

The offline pre-training phase serves as a foundation for our online optimization. We employ a strategy to boost the Q-values of OOD actions [13], which are actions that deviate from the observed data distribution. By doing so, we aim to improve the generalization capabilities of the offline policy, allowing it to handle a wider range of negotiation scenarios. Building upon the pre-trained offline policy, we focus on the online optimization phase to further refine the negotiation performance. In this work, offline data is collected in the real-world applications considered in the ANAC competitions. Once offline datasets have been obtained, we can leverage offline RL algorithms to train our offline negotiation policy. To offline learning a strategy from fixed negotiation datasets, we introduce a special treatment for OOD actions during this offline phase, rather than excessively penalizing them. Specifically, we adopt a moderately optimistic evaluation of their Q-values, ensuring that their Q-values remain below the maximum value. The estimated values for OOD actions are allowed to be high as long as it does not affect the learning for the optimal policy supported by the dataset. This approach prevents the selection of OOD actions during the offline phase, thereby maintaining the performance of the offline policy. Simultaneously, it appropriately elevates the Q-values of OOD actions, facilitating their exploration during the online phase. Moreover, we introduce effective mechanisms to facilitate this process. These mechanisms leverage the knowledge and experience gained from the offline pre-training phase to achieve stable and rapid performance improvements. Through careful design and incorporation of these mechanisms, we aim to enhance the agent's ability to negotiate efficiently and effectively in real-time interactions against various opponents.

3 RESULTS

Table 1: Tournament results. The best is marked in bold.

Agent name	avg utility	avg social welfare	agreement
RLBOA	0.49	1.33	98.7%
TitForTat	0.56	1.16	83.33%
RandomAgent	0.57	1.26	88.63%
LinearAgent	0.59	1.42	**100.00%**
ChargingBoul	0.62	1.14	75.75%
AntAllianceAgent	0.63	**1.43**	96.96%
Comp.Agent	0.64	1.42	98.48%
RGAgent	0.66	1.26	84.84%
SmartAgent	0.68	1.35	92.42%
DreamTeam109Agent	0.69	1.32	90.91%
MiCRO	0.71	1.27	85.61%
AntHeartAgent	0.71	1.30	86.36%
SuperAgent	0.71	1.26	86.36%
ANOTO	**0.76**	1.38	93.80%

Then, to have a comprehensive view of the performance, Table 1 presents the ANAC-like tournament results. In this setup, recent ANAC agents are considered, and each agent negotiates with all other agents in a number of ANAC domains. As depicted in the table, the ANOTO agent behaved better than all other agents, leading the mean utility of ANAC agents by a margin of 23.9%. This results demonstrate the at agent's excellent negotiation skills against a variety of opponents. On the other hand, the RLBOA agent received a low score of 0.49, which did not exceed the mean utility of collecting agents. Interestingly, the ANOTO agent achieved a fairly high social welfare, despite not considering it in its objective. We attribute this satisfying performance to the fact that the end-to-end policy network can predict the opponents' issue weights, increasing the chances of finding offers closer to the Pareto frontier. Additionally, the ANOTO agent benefits from the effective strategy learned during the offline training phase, which observed opponent behaviors. As a result, the ANOTO agent exhibited higher agreement rates than most of the winning ANAC agents.

4 CONCLUSIONS AND FUTURE WORK

We proposes a novel approach for improving Automated Negotiation via Offline-to-Online reinforcement learning (ANOTO). ANOTO allows the negotiating agent to learn an effective end-to-end RL-based strategy from historical offline data, and to adjust its strategy in an adaptive and efficient manner. Overall, the experimental results show that the ANOTO Agent, leveraging a more advanced technical framework, outperformed state-of-the-art negotiating agents. The exceptional results justify further research efforts into this approach. In the future, we will extend the proposed approach to negotiation settings where autonomous agents use natural language to bargain. Additionally, it is of great interest to examine the performance of the negotiation approach against human negotiators.

ACKNOWLEDGMENTS

The work is supported by the National Natural Science Foundation of China (Grant Nos.: 61602391, 62222311).

SOCIETAL IMPACT

This work proposes a highly effective and efficient RL-based approach for learning negotiation strategies from previously collected data. The deployment of this technology will enhance market efficiency and also assist human negotiators in achieving better outcomes. It can compensate for the limited computational abilities of humans, especially in complex negotiations with a wide range of possible outcomes. However, since it is still in the early stages, we do not anticipate any significant negative social impact, such as massive job loss.

REFERENCES

[1] Siqi Chen, Yonghao Cui, Cong Shang, Jianye Hao, and Gerhard Weiss. 2019. ONECG: Online Negotiation Environment for Coalitional Games. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS '19, Montreal, QC, Canada, May 13-17, 2019*, Edith Elkind, Manuela Veloso, Noa Agmon, and Matthew E. Taylor (Eds.). International Foundation for Autonomous Agents and Multiagent Systems, 2348–2350. http://dl.acm.org/citation.cfm?id=3332108

[2] Siqi Chen, Jianye Hao, Gerhard Weiss, Shuang Zhou, and Zili Zhang. 2015. Toward Efficient Agreements in Real-Time Multilateral Agent-Based Negotiations. In *27th IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2015, Vietri sul Mare, Italy, November 9-11, 2015*. IEEE Computer Society, 896–903. https://doi.org/10.1109/ICTAI.2015.130

[3] Siqi Chen and Ran Su. 2022. An autonomous agent for negotiation with multiple communication channels using parametrized deep Q-network. *Mathematical Biosciences and Engineering* 19, 8 (2022), 7933–7951. https://doi.org/10.3934/mbe.2022371

[4] Siqi Chen, Qisong Sun, Heng You, Tianpei Yang, and Jianye Hao. 2023. Transfer Learning based Agent for Automated Negotiation. In *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2023, London, United Kingdom, 29 May 2023 - 2 June 2023*. ACM, 2895–2898. https://doi.org/10.5555/3545946.3599115

[5] Siqi Chen and Gerhard Weiss. 2014. An Intelligent Agent for Bilateral Negotiation with Unknown Opponents in Continuous-Time Domains. *ACM Trans. Auton. Adapt. Syst.* 9, 3 (2014), 16:1–16:24. https://doi.org/10.1145/2629577

[6] Siqi Chen and Gerhard Weiss. 2015. An approach to complex agent-based negotiations via effectively modeling unknown opponents. *Expert Syst. Appl.* 42, 5 (2015), 2287–2304. https://doi.org/10.1016/j.eswa.2014.10.048

[7] Siqi Chen, Tianpei Yang, Heng You, Jianing Zhao, Jianye Hao, and Gerhard Weiss. 2023. Transfer Reinforcement Learning Based Negotiating Agent Framework. In *Advances in Knowledge Discovery and Data Mining - 27th Pacific-Asia Conference on Knowledge Discovery and Data Mining, PAKDD 2023, Osaka, Japan, 2023, Proceedings, Part II (Lecture Notes in Computer Science, Vol. 13936)*. Springer, 386–397. https://doi.org/10.1007/978-3-031-33377-4_30

[8] Siqi Chen, Yang Yang, and Ran Su. 2022. Deep reinforcement learning with emergent communication for coalitional negotiation games. *Mathematical Biosciences and Engineering* 19, 5 (2022), 4592–4609. https://doi.org/10.3934/mbe.2022212

[9] Siqi Chen, Yang Yang, Haoran Zhou, Qisong Sun, and Ran Su. 2023. DNN-PNN: A parallel deep neural network model to improve anticancer drug sensitivity. *Methods* 209 (2023), 1–9. https://doi.org/10.1016/j.ymeth.2022.11.002

[10] Siqi Chen, Jianing Zhao, Gerhard Weiss, Ran Su, and Kaiyou Lei. 2023. An effective negotiating agent framework based on deep offline reinforcement learning. In *Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence (Proceedings of Machine Learning Research, Vol. 216)*, Robin J. Evans and Ilya Shpitser (Eds.). PMLR, 324–335. https://proceedings.mlr.press/v216/chen23c.html

[11] Dave de Jonge. 2022. An Analysis of the Linear Bilateral ANAC Domains Using the MiCRO Benchmark Strategy. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022*, Luc De Raedt (Ed.). ijcai.org, 223–229. https://doi.org/10.24963/ijcai.2022/32

[12] Ryota Higa, Katsuhide Fujita, Toki Takahashi, Takumu Shimizu, and Shinji Nakadai. 2023. Reward-based Negotiating Agent Strategies. In *Proceedings of the 37th AAAI Conference on Artificial Intelligence, AAAI-23*. 297–304.

[13] Aviral Kumar, Aurick Zhou, George Tucker, and Sergey Levine. 2020. Conservative q-learning for offline reinforcement learning. *Advances in Neural Information Processing Systems* 33 (2020), 1179–1191.

[14] Jiafei Lyu, Xiaoteng Ma, Xiu Li, and Zongqing Lu. 2022. Mildly Conservative Q-Learning for Offline Reinforcement Learning. In *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (Eds.), Vol. 35. Curran Associates, Inc., 1711–1724.

[15] Zhiyu Mou, Yusen Huo, Rongquan Bai, Mingzhou Xie, Chuan Yu, Jian Xu, and Bo Zheng. 2022. Sustainable Online Reinforcement Learning for Auto-bidding. In *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (Eds.), Vol. 35. Curran Associates, Inc., 2651–2663. https://proceedings.neurips.cc/paper_files/paper/2022/file/11faf17bf7e5412d9cded369f97db23d-Paper-Conference.pdf

[16] Ran Su, Huichen Wu, Bo Xu, Xiaofeng Liu, and Leyi Wei. 2019. Developing a Multi-Dose Computational Model for Drug-Induced Hepatotoxicity Prediction Based on Toxicogenomics Data. *IEEE ACM Trans. Comput. Biol. Bioinform.* 16, 4 (2019), 1231–1239. https://doi.org/10.1109/TCBB.2018.2858756

[17] Ran Su, Haitang Yang, Leyi Wei, Siqi Chen, and Quan Zou. 2022. A multi-label learning model for predicting drug-induced pathology in multi-organ based on toxicogenomics data. *PLOS Computational Biology* 18, 9 (09 2022), 1–28. https://doi.org/10.1371/journal.pcbi.1010402

[18] Yizheng Wang, Yixiao Zhai, Yijie Ding, and Quan Zou. 2023. SBSM-Pro: Support Bio-sequence Machine for Proteins. *arXiv preprint arXiv:2308.10275* (2023). https://doi.org/10.48550/arXiv.2308.10275

[19] Gerhard Weiss (Ed.). 2013. *Multiagent Systems, 2nd edition.* MIT Press, USA.

[20] Leling Wu, Siqi Chen, Xiaoyang Gao, Yan Zheng, and Jianye Hao. 2021. Detecting and Learning Against Unknown Opponents for Automated Negotiations. In *PRICAI 2021: Trends in Artificial Intelligence*, Duc Nghia Pham, Thanaruk Theeramunkong, Guido Governatori, and Fenrong Liu (Eds.). Springer International Publishing, Cham, 17–31.