Scalable Negotiating Agent Strategy via Multi-Issue Policy Network (Student Abstract)

Takumu Shimizu[1,2], **Ryota Higa**[2,3], **Toki Takahashi**[1,2], **Katsuhide Fujita**[1,2], **Shinji Nakadai**[2,3]
[1] Tokyo University of Agriculture and Technology, 2-24-16 Naka-cho, Koganei-shi, Tokyo 184-8588, Japan
[2] National Institute of Advanced Industrial Science and Technology
[3] NEC Data Science Research Laboratories
shimizu@katfuji.lab.tuat.ac.jp, r-higaryouta@nec.com, takahashi@katfuji.lab.tuat.ac.jp, katfuji@cc.tuat.ac.jp, nakadai@nec.com

Abstract

Previous research on the comprehensive negotiation strategy using deep reinforcement learning (RL) has scalability issues of not performing effectively in the large-sized domains. We improve negotiation strategy via deep RL by considering an issue-based represented deep policy network to deal with multi-issue negotiation. The architecture of the proposed learning agent considers the characteristics of multi-issue negotiation domains and policy-based learning. We demonstrate that proposed method achieve equivalent or higher utility than existing negotiation agents in the large-sized domains.

Introduction

Negotiation is an essential element for establishing cooperation and collaborations in multi-agent systems. Automated negotiation strategies have attracted significant research attention, and competitions, such as Automated Negotiation Agents Competitions (ANAC) [1], have been organized to discuss various negotiation strategies. The first comprehensive negotiation strategy is the versatile negotiating agent strategy (VeNAS) via deep reinforcement learning (Takahashi et al. 2022). However, the limitation of VeNAS is that it has not demonstrated effective performance in the case of negotiation with a domain of large size. In this study, we propose a scalable negotiating agent strategy via a multi-issue policy network (MiPN) using a policy network. We demonstrate that the proposed method achieves comparable or higher utility in the large-sized domains than existing baseline negotiation agents based on heuristic strategy and VeNAS not considering issue-based representations.

End-to-end Scalable Negotiating Agent Strategy via Multi-Issue Policy Network

We assume a bilateral multi-issue negotiation, and employ the same negotiation environment as in the previous research (Takahashi et al. 2022). In addition, to apply machine learning to negotiation agents, it is necessary to formulate Markov Decision Process (MDP) for multi-issue negotiations. We also use the same formulation of alternating offers protocol (AOP) using finite MDP defined in the previous research (Takahashi et al. 2022)

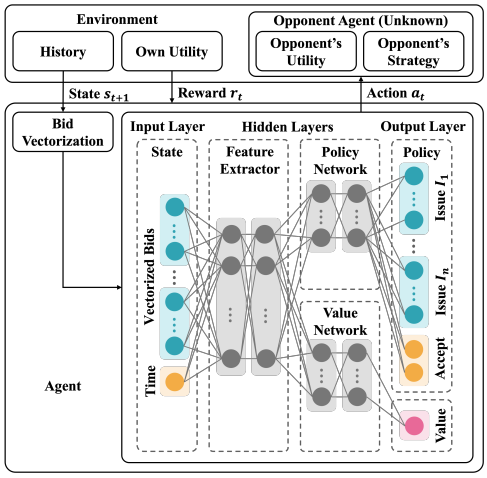

Figure 1: MiPN architecture. The top box is the environment and the bottom box is the body of the proposed agent architecture.

Proposed Architecture Figure 1 illustrates the proposed reinforcement learning architecture via a multi-issue policy network. The environment includes the history of the bids exchanged between the agent and the opponent, including their own utility functions. The opponent possesses its utility function and strategy; however, these are unknown to the agent. In the agent architecture, the nodes of the input layer are allocated as vectorized bids, and time is defined as states. The nodes of the output layer are allocated vectorized bids grouping as multi-issue for offer, acceptance, and value. A part of the hidden layer consists of an end-to-end learning model based on a policy network.

Multi-Issue Policy Network We propose a deep strategy network model for the combinatorial issue and accept shared policies as follows

$\pi_\theta(a|s) := \pi_{\theta_A}(a_A|h(s)) \prod_{i=I_1}^{I_n} \pi_{\theta_i}(a_i|h(s))$,

where $\eta \sim \pi_{\theta_A}(a_A|h(s))$ is the accept strategy policy function, which generates an accept signal $\eta \in \{0, 1\}$, and $v^i_{k_i} \sim \pi_{\theta_i}(a_i|h(s))$ is the issue I_i's policy function, which gener-

[1] http://web.tuat.ac.jp/~katfuji/ANAC2021/

	Grocery(10^3, L)			thompson(10^3, H)			Car(10^4, L)			EnergySmall_A(10^4, H)		
	Bas.	MiPN	VeN.	Bas.	MiPN	VeN.	Bas.	MiPN	VeN.	Bas.	MiPN	VeN.
Boulware	0.640	**0.992**	0.960	0.330	**0.985**	0.910	0.560	**0.974**	0.570	0.360	**0.951**	0.580
Linear	0.810	**0.960**	0.810	0.610	**0.855**	0.710	0.850	**0.972**	0.930	0.590	**0.932**	0.490
Concedder	0.930	**1.000**	**1.000**	0.820	**0.913**	0.650	0.980	**0.994**	0.970	0.830	**0.933**	0.920
TitForTat1	0.790	**1.000**	0.960	0.830	**0.879**	0.110	0.920	**0.998**	0.790	0.910	**0.993**	0.260
TitForTat2	0.870	**1.000**	0.960	0.850	**0.913**	0.140	0.800	**0.995**	0.310	0.790	**0.951**	0.290
AgentK	**0.720**	0.630	0.590	0.220	**0.240**	0.220	0.600	**0.883**	0.120	0.210	**0.232**	0.000
HardHeaded	**0.640**	0.630	0.600	0.200	0.290	**0.300**	0.550	**0.852**	0.069	**0.100**	0.066	0.000
Atlas3	**0.810**	0.668	0.970	0.610	**0.976**	0.940	0.730	**0.954**	0.820	0.550	**0.981**	0.690
AgentGG	0.600	**0.782**	0.350	0.370	**0.664**	0.180	0.690	**0.874**	0.010	0.240	**0.605**	0.000

Table 1: Utility for each domain and opponent. "Grocery (10^3, L)" means that the domain name is Grocery, the domain size is 10^3, and the opposition is low. "Bas." means baseline which was the average score of the same nine agents used for training negotiated with eight other agents apart from themselves, "MiPN" means the proposed architecture with PPO considering issue-based representation. "VeN." means VeNAS with DDQN not considering issue-based representation. Bold entries indicate the highest utility in each negotiation setting.

ates possible values. $v^i_{k_i}$ and $h(s)$ are feature and hidden parameters of the input policy functions, respectively. The policy gradient loss functions can be calculated independently as $\ln \pi_\theta(a|s) = \ln \pi_{\theta_A}(a_A|h(s)) + \sum_{i=I_1}^{I_n} \ln \pi_{\theta_i}(a_i|h(s))$, and this is scalable when applied to large issue spaces. To select an action ω, the proposed architecture computes $\omega_t = \arg\max_{a_t} Q(s_t, a_t)$, or samples by the policy-method $\omega_t \sim \pi_\theta(a_t|s_t)$.

Experiments and Evaluations

The negotiation deadline was set to 40 rounds, that is, the negotiation ends when both agents take action 40 times. To indicate that the proposed approach can be trained in various negotiation domains, we considered domains with comprehensive large sizes of outcome space and opposition, which represents the difficulty in reaching an agreement. Accordingly, we selected four domains: Grocery, Thompson, Car, and EnergySmall_A. We employed nine negotiating agents: three time-dependent (Boulware, Linear, and Conceder), two behavior-dependent (TitForTat1 and TitForTat2), and four past ANAC champions (AgentK, HardHeaded, Atlas3, and AgentGG). These negotiation domains and strategies were included in GENIUS[2].

We applied the proximal policy optimization (PPO) to the proposed architecture. The training period was 500,000 steps. When they reach an agreement, the utility of the agreement is rewarded. A penalty of $K = -1$ is given when the negotiation ends without reaching an agreement. Otherwise, the reward is 0. The performance of the agents was scored by their obtained utility and evaluated based on the average scores out of 100 negotiations.

Experimental Results It is clear from Table 1 that MiPN has the potential to achieve a comparable or better utility than the baseline, which indicates that the policy obtained by RL is more adaptive to the environment than the heuristic strategy.

[2]http://ii.tudelft.nl/genius/

However, VeNAS significantly decreased the individual utilities as the domain size was large, and it was comparable or worse than the baseline agents. This is because the several trained agents using the VeNAS architecture failed to achieve agreements in large-sized domains. Due to scalability issues, VeNAS was unable to learn the negotiation strategies to reach agreements using Q-learning based approach in large-sized domains in a realistic time frame.

Comparing RL-based agents, it can be observed that MiPN achieves comparable or higher utilities than VeNAS, which does not consider the multi-issue action representations. Therefore, MiPN can adapt to various negotiation strategies without designing an effective strategy that considers the opponent's strategies and domains. The results can be attributed to the proposed agent making its offer including the option of the issues with higher weights for its utility function by correctly predicting the weights of each issue (ω_i). Additionally, it does not stick to the issue that its weight is not high and the opponent's weight is high. It finds better options for both sides by making a concession effectively toward the issue that is important for both sides.

Conclusion and Future Work

This study proposes a scalable negotiating agent strategy via a multi-issue policy network (MiPN). MiPN improves the scalability of negotiation domains by considering the characteristic of multi-issue negotiation and policy-based learning. We demonstrated that MiPN achieved comparable or higher utility than existing baseline negotiation agents and VeNAS in large-sized domains. Further research is learning negotiation agent that can work effectively in a different domain and with a different opponent from the training.

References

Takahashi, T.; Higa, R.; Fujita, K.; and Nakadai, S. 2022. VeNAS: Versatile Negotiating Agent Strategy via Deep Reinforcement Learning. *Proceedings of AAAI-2022*, 36(11): 13065–13066.