VeNAS: Versatile Negotiating Agent Strategy via Deep Reinforcement Learning (Student Abstract)

Toki Takahashi,[1,2] **Ryota Higa,**[2,3] **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
tokit@st.go.tuat.ac.jp, r-higaryouta@nec.com, katfuji@cc.tuat.ac.jp, nakadai@nec.com

Abstract

Existing research in the field of automated negotiation considers a negotiation architecture in which some of the negotiation components are designed separately by reinforcement learning (RL), but comprehensive negotiation strategy design has not been achieved. In this study, we formulated an RL model based on a Markov decision process (MDP) for bilateral multi-issue negotiations. We propose a versatile negotiating agent that can effectively learn various negotiation strategies and domains through comprehensive strategies using deep RL. We show that the proposed method can achieve the same or better utility than existing negotiation agents.

Introduction

Negotiation has always been an important element in establishing cooperation and collaboration in multi-agent systems. In the field of automated negotiation, the topic of negotiation strategies is being actively studied, and various strategies are being discussed in competitions such as Automated Negotiating Agents Competition (ANAC) [1]. Recently, agent strategies using reinforcement learning (RL) have attracted much attention because they can be adapted to various scenarios and opponents (Bakker et al. 2019; Razeghi, Yavus, and Aydoğan 2020). However, the existing studies considered a negotiation architecture in which some of the negotiation components, such as bidding, opponent modeling, and acceptance, are designed separately. It therefore remains an open and interesting challenge to identify approaches that use RL to design comprehensive negotiation strategies without heuristic negotiation components that rely on expert knowledge and experiments. In this study, we propose a versatile negotiating agent strategy (VeNAS) that comprehensively considers the key components of the negotiation strategy. We demonstrate that the proposed method can achieve comparable or higher utility than the existing baseline negotiation agents based on heuristic strategies, including champions of previous competitions.

VeNAS: Versatile Negotiating Agent Strategy

We assume a bilateral multi-issue negotiation in which two agents negotiate a domain D. A negotiation domain D defines a set of all possible outcomes Ω that can be proposed during the negotiation. Every agent has a unique preference profile that represents its own preferences for the outcome $\omega \in \Omega$, and they are not shared with other agents. The utility of an outcome is defined by a utility function $U(\cdot)$, which is normalized to the range $[0, 1]$. The interaction between negotiating agents is regulated by a negotiation protocol. Here, we consider the alternating offers protocol (AOP). A negotiation session has the timeline $t \in [0, T]$, where T is a deadline. The outcome proposed in the negotiation is called a bid, and the bid at time t is denoted as ω_t.

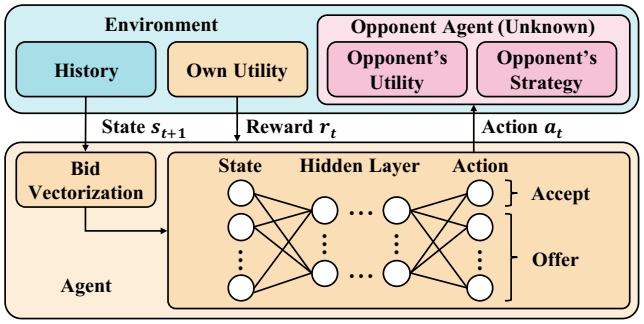

Figure 1: VeNAS architecture. The top box is the environment and the bottom box is the body of the VeNAS model.

VeNAS Architecture Figure 1 illustrates the proposed Versatile Negotiating Agent Strategy (VeNAS) architecture. The environment includes the history of exchanging bids between the agent and the opponent and their own utility functions. The opponent agent has the opponent's utility function and strategy, but they are not included as input, reward, or body in the learning architecture of VeNAS because they are unknown information in negotiations. Compared with a few existing studies ((Bagga et al. 2020), etc.), the action of the VeNAS architecture covers all negotiation actions, not just what to offer. Its input is the bid, not the utility of the bid, and output is the negotiation action, including the offer and its bid and accept.

Markov Decision Process for Negotiation To achieve VeNAS, it is necessary to formulate a Markov decision process for bilateral multi-issue negotiations. A finite MDP is

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

	Laptop (S, Lo)			It.vsCy. (M, Hi)			IS_BT_Ac. (M, Lo)			Grocery (L, Lo)			thompson (L, Hi)		
	Bas.	VeN.	DRB.	Bas.	VeN.	DRB.	Bas.	VeN.	DRB.	Bas.	VeN.	DRB.	Bas.	VeN.	DRB.
Boulware	0.841	**1.000**	**1.000**	0.487	**1.000**	0.904	0.853	**1.000**	0.963	0.669	**1.000**	0.960	0.431	**0.985**	0.940
Linear	0.924	**1.000**	**1.000**	0.660	**1.000**	0.904	0.901	**0.993**	0.940	0.838	**1.000**	0.960	0.645	0.870	**0.913**
Conceder	**1.000**	**1.000**	**1.000**	0.852	**0.904**	**0.904**	0.951	**0.993**	0.963	0.977	**1.000**	0.925	0.824	**0.942**	0.915
TitForTat1	0.910	**1.000**	**1.000**	0.615	**1.000**	0.760	0.697	**0.940**	**0.940**	0.840	**1.000**	0.960	0.801	**0.980**	0.940
TitForTat2	0.889	**1.000**	**1.000**	0.556	**1.000**	0.808	0.758	**0.948**	0.940	0.900	**1.000**	0.960	0.822	**1.000**	0.940
AgentK	0.705	0.863	**0.874**	0.308	0.292	**0.550**	**0.859**	0.725	0.853	0.588	0.536	**0.812**	0.257	0.224	**0.495**
HardHeaded	0.760	**1.000**	**1.000**	0.160	**0.355**	0.353	0.729	**0.873**	**0.873**	0.536	0.628	**0.686**	0.179	0.153	**0.323**
Atlas3	0.911	**1.000**	**1.000**	0.602	0.709	**0.905**	0.916	0.937	**0.964**	0.809	0.958	**0.960**	0.615	0.716	**0.940**
AgentGG	0.836	0.851	**0.852**	0.351	0.313	**0.395**	0.743	**0.875**	0.846	0.618	0.621	**0.690**	0.399	0.409	**0.462**

Table 1: Utility for each domain and opponent. Laptop (S, Lo) means that the domain name is Laptop, the domain size is small, and the opposition is low. Bas. is baseline, VeN. is VeNAS, and DRB. is DRBOA. Bold entries indicate the highest utility in each negotiation setting.

provided as $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{T} \rangle$, where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, $\mathcal{R}$ is the set of rewards, and $\mathcal{T}$ is the transition function. We define an AOP using a finite MDP as follows: Set of state $s_t \in \mathcal{S}$: The agent's offer ω_t, the opponent's offer ω'_t, the accept signal η'_t, and the normalized time t/T. Set of action $a_t \in \mathcal{A}$: The agent's selected offer ω_t and accept signal η_t. Reward function $r(s, a)$: When the agent accepts, $r(\{..., \omega'_t\}, \eta_{t+1}) = U(\omega'_t)$ is rewarded. When the opponent accepts $r(\{..., \omega_t, \eta'_{t+1}\}, \omega_t) = U(\omega_t)$ is rewarded. Penalty K is given when the negotiation ends without reaching an agreement. Otherwise, the reward is 0.

Experiments and Evaluations

The negotiation deadline T was set to 40 rounds, and the negotiation ends when both agents have each acted 40 times. To demonstrate adaptability of VeNAS to various domains and opponents, we used five domains and nine agents. Considering the size of the outcome space $|\Omega|$ and the opposition that represents the difficulty of reaching a better agreement, five domains were selected: Laptop, ItexvsCypress, IS_BT_Acquisition, Grocery, and thompson. We used 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 the negotiation platform GENIUS[2]. We used double deep Q-learning (DDQN) to the evaluate of our architecture. The training period was 2000 episodes, and we also set a penalty of $K = -1$ for failure to reach an agreement. To stabilize the learning, we trained with 300 different initial values. The performance of the agents was scored by their obtained utility and evaluated based on the highest utility among the 300 agents. For comparison, we used the baseline, which was the average score of the same nine agents used for training negotiated with eight other agents besides themselves, and DRBOA, which is RLBOA-agent (Bakker et al. 2019) trained by DDQN.

Experimental Results It is clear from Table 1 that VeNAS is able to achieve the same or better utility than the baseline, which indicates that the policy obtained by RL is more adaptive to the environment than the heuristic strategy. In particular, VeNAS can adapt to various negotiation strategies without designing an effective strategy that considers the opponent's strategies and domains. In some cases, VeNAS could not obtain a higher utility than the baseline, where the size was large and the opposition was high, such as in the thompson domain. This is owing to the fact that as the domain size increases, the size of the state space and action space of VeNAS increases. VeNAS also could not obtain a higher utility than DRBOA for competition champions. This may be because in DRBOA, unlike VeNAS, the size of the state and action space remains constant even when the domain size increases, taking into account the utility function.

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

Conclusion and Future Work

In this study, we propose a versatile negotiating agent strategy (VeNAS) via deep RL. Our model comprehensively processed the opponent's offer to determine its own next action, such as offering bids or accepting them. We formulated the bilateral multi-issue negotiation as MDP to apply RL. We demonstrated that VeNAS can achieve comparable or higher utility than existing baseline negotiation agents, including champions of previous competitions. One possible direction for future research is to improve our learning agent model to obtain an effective negotiation strategy, by reducing the state and action spaces.

References

Bagga, P.; Paoletti, N.; Alrayes, B.; and Stathis, K. 2020. A Deep Reinforcement Learning Approach to Concurrent Bilateral Negotiation. In *IJCAI*, 297–303.

Bakker, J.; Hammond, A.; Bloembergen, D.; and Baarslag, T. 2019. RLBOA: A Modular Reinforcement Learning Framework for Autonomous Negotiating Agents. In *AAMAS*, 260–268.

Razeghi, Y.; Yavus, C. O. B.; and Aydoğan, R. 2020. Deep reinforcement learning for acceptance strategy in bilateral negotiations. *Turkish Journal of Electrical Engineering & Computer Sciences*, 28: 1824–1840.