Chapter 11
CUHKAgent: An Adaptive Negotiation Strategy for Bilateral Negotiations over Multiple Items

Jianye Hao and Ho-fung Leung

Abstract Automated negotiation techniques can greatly improve the negotiation efficiency and quality of our human being, and a lot of automated negotiation strategies and mechanisms have been proposed in different negotiation scenarios until now. To achieve efficient negotiation, there are two major challenges we are usually faced with: how to model and predict the strategy and preference of the opponent. To this end we propose an adaptive negotiating strategy (CUHKAgent) to predict the opponent's strategy and preference at a high level, and make informed decision accordingly.

Keywords Adaption • Negotiation • Reinforcement learning

11.1 Introduction

Negotiation is a commonly used approach to resolve conflicts and reach agreements between different parties in our daily life. Automated negotiation techniques can, to a large extent, alleviate the efforts of human, and also facilitate human in reaching better negotiation outcomes by compensating for the limited computational abilities of humans when they are faced with complex negotiations.

Until now, a lot of automated negotiation strategies and mechanisms have been proposed in different negotiation scenarios [1–5]. The major difficulty in designing automated negotiation agent is how to achieve optimal negotiation results given incomplete information on the negotiating partner. The negotiation partner usually keeps its negotiation strategy and its preference as its private information to avoid exploitations. To achieve efficient negotiation, a lot of research efforts have been

J. Hao (✉) • H.-f. Leung
Department of Computer Science and Engineering,
The Chinese University of Hong Kong, Hong Kong
e-mail: jyhao@cse.cuhk.edu.hk; lhf@cse.cuhk.edu.hk

I. Marsa-Maestre et al. (eds.), *Novel Insights in Agent-based Complex Automated Negotiation*, Studies in Computational Intelligence 535,
DOI 10.1007/978-4-431-54758-7_11,

devoted into the following two directions: learning the opponent's negotiation decision function [4, 6] and estimating the opponent's preference profile [3, 7, 8]. Previous work usually assumes that the opponent's strategy or preference profile follow certain predefined patterns which can be accurately modeled as certain classes of mathematical functions. For example, one may assume that the opponent makes negotiation following certain probability function [2], and the task is how to accurately estimate the corresponding coefficients of the probability function based on the negotiation histories. For utility function, one commonly adopted assumption is that the utility function is additive in which the contribution of each issue to the overall utility is independent. However, in practice, the agents may not strictly follow any function to make decisions, and also they may not determine their preferences over different combinations of items following a fixed type of utility function. The consequence is that it may not be feasible to learn the opponent's decision function or utility function since such kind of functions may not exist at all. Even if the opponent indeed makes decisions following certain forms of mathematical functions, it is highly likely that it has already changed its decision function to another form, which thus makes what we have learned useless.

Due to the aforementioned issues, we propose that, to make efficient negotiation, an agent should focus on making timely and effective adaption to the opponent's past behaviors rather than learning the exact forms of the opponent's decision function or utility function. Considering the high diversity of the available negotiation strategies that an agent can choose, it is usually very difficult (or even impossible) to predict which specific strategy (or combination of strategies) the negotiating partner is using based on this limited information. To effectively cope with different types of opponents, we introduce the concept of non-exploitation point to adaptively adjust the degree that an agent exploits its negotiating opponent. The value of the non-exploitation point is determined by the characteristics of the negotiation scenario and the concessive degree of the negotiating partner, which is estimated based on the negotiation history. Besides, to maximize the possibility that the offer our agent proposes will be accepted by its negotiating partner, it can be useful to make predictions on the preference profile of the negotiating partner. Instead of explicitly modeling the negotiation partner's utility function, we propose a reinforcement-learning based approach to determine the optimal proposal for the negotiating partner based on the current negotiation history.

The structure of this paper is organized as follows. In Sect. 11.2, we discuss a number of key issues related with negotiation strategy design. Following that, we introduce our negotiation agent CUHKAgent in details in Sect. 11.3. Finally we make conclusions in Sect. 11.4.

11.2 Designing Issues

In this section, we discuss a number of key issues when designing an efficient negotiation strategy.

11.2.1 Learning the Opponent's Decision Function or Not?

Much effort has been given to predict the opponent's exact decision function in previous work. This is usually based on the assumption that the opponent makes decisions following certain predefined patterns which can be accurately modeled as certain classes of mathematical functions. For example, one may assume that the opponent decides whether to accept an offer following certain probability function [2], and the task is how to accurately estimate the corresponding coefficients of the probability function based on the negotiation histories. Another example is that in [4] the authors propose a way of predicting the opponent's next round offer based on the assumption that the opponent make decisions based on the combination of time-dependent and behavior-dependent decision functions. Based on the prediction results, the optimal offer(s) to be proposed to the opponent can be determined by modeling the negotiation as a multi-stage control process and calculating the sequence of optimal controls (offers) accordingly.

However, in practice, this kind of assumption is usually not valid considering the high diversity of the possible strategies that an agent may adopt. An agent usually can exhibit highly dynamic behaviors which cannot be modeled as certain types of mathematical functions. Even if the opponent indeed makes decisions by strictly following certain forms of mathematical functions, it is highly likely that its decision function is changed in a dynamic way, which thus may make what we have learned useless. Instead of predicting the opponent's decision function, an alternative approach is that we can model the opponent's behavior at a high level based on certain high-level characteristics such as its concession degree, and make adaptive response accordingly.

11.2.2 How to Make Concessions to the Opponent?

There are a number of factors to be considered when determining the concession degree to the opponent. The first factor is the amount of negotiation time left. The more the negotiation time has passed, the less utility an agent may obtain due to discounting effect. Therefore we need to carefully balance between the possible utility gain by being tough and the utility loss due to discounting effect. The second factor is the discounting degree. This factor is closely related with the first factor—the negotiation time left. The larger the discounting factor is, the more cautious we need to be to avoid possible utility loss due to discounting effect. The last factor is the concession degree of the opponent. The more concessive the opponent is, the more we can exploit the opponent by being tough to the opponent, and vice versa.

11.2.3 How to Guess the Opponent's Preference?

In the current setting of the ANAC competition [9], the agents' preference functions are assumed to be additive, and thus it may be possible for an agent to learn its opponent's preference function through past negotiations. For example, in [3], the authors propose a Bayesian learning based approach to learn the opponent's preference function, i.e., the issue preference and the issues priorities of the opponent.

However, in general, an agent's preference function can be in any form and may not be known to other agents. Thus it is infeasible for us to learn the exact preference function within limited negotiation time considering the high diversity of possible utility functions that an agent can choose. Instead of learning the exact preference function of the opponent, an alternative approach is to directly learn the relative importance of each proposal (combination of items) of the opponent based on the opponent's past proposals.

11.3 Strategy Description

In this section, we describe the key components of CUHKAgent, which is a specific implementation of the *ABiNeS* strategy [10]. Before describing the details, we introduce a few mathematical notions which will be used in the following descriptions. For each negotiation scenario, both agents can negotiate over multiple issues (items), and each item can have a number of different values. Let us denote the set of items as $\mathscr{M}$, and the set of values for each item $m_i \in \mathscr{M}$ as $\mathscr{V}_i$. For each negotiation outcome ω, we use $\omega(m_i)$ to denote the corresponding value of the item m_i in the negotiation outcome ω.

11.3.1 How to Determine the Acceptance Threshold

The determination of the acceptance threshold is the key issue in designing a negotiation strategy. For CUHKAgent, the principle is that it accepts a proposal from its opponent if its utility over this proposal is higher than its current acceptance threshold, and also any proposal offered by CUHKAgent should also exceed its acceptance threshold. The value of the acceptance threshold reflects the agent's current concession degree and should be adaptively adjusted based on the opponent's concession degree and the characteristic of the negotiation environment.

We assume that the negotiating partner is self-interested, and it will accept any proposal when the deadline is approaching ($t \sim 1$). Therefore the acceptance threshold of CUHKAgent is always higher than the highest utility it can obtain when $t = 1$. Specifically, at any time t, the acceptance threshold l_t of CUHKAgent should not be lower than $u^{max}\delta^{1-t}$, where u^{max} is its maximum utility over the negotiation domain without discounting. Since the negotiating goal is to reach an agreement which maximize the agent's own utility as much as possible, its negotiating partner should be exploited as much as possible by setting its acceptance threshold as high as possible. One the other hand, due to the discounting effect, the actual utility the agent receives can become extremely low though its original utility over the mutually-agreed negotiation outcome is high, if it takes too long for the agents to reach the agreement. In the worst case the negotiation may end up with a break-off and each agent obtains zero utility. Thus we also need to make certain compromises to the negotiating partner, i.e., lower the acceptance threshold, depending on the type of the partner we are negotiating with. Therefore, the key problem is how to balance the trade-off between exploiting and making compromise to the negotiating partner. Towards this end, we introduce the adaptive non-exploitation point λ, which represents the specific time when we should stop exploitations on the negotiating partner. This value is adaptively adjusted based on the behavior of the negotiating partner. Specifically we propose that for any time $t < \lambda$, CUHKAgent always exploits its negotiating partner (agent B) by setting its acceptance threshold to a value higher than $u^{max}\delta^{1-\lambda}$ and approaching this value until time λ according to certain pattern of behavior. After time λ, its acceptance threshold is set to be equal to $u^{max}\delta^{1-t}$ forever, and any proposal over which its utility is higher than this value will be accepted. Formally, the acceptance threshold l_A^t of CUHKAgent at time t is determined as follows,

$$I = \begin{cases} u^{max} - (u^{max} - u^{max}\delta^{1-\lambda})(\frac{t}{\lambda})^{\alpha} & if t < \lambda \\ u^{max}\delta^{1-t} & \text{otherwise} \end{cases} \tag{11.1}$$

where the variable α controls the way the acceptance threshold approaches $u^{max}\delta^{1-t}$ (boulware ($\alpha > 1$), conceder ($\alpha < 1$) or linear ($\alpha = 1$)). One example showing the dynamics of the acceptance threshold with time t with different value of λ is given in Fig. 11.1.

The remaining question is how to calculate the value of non-exploitation point λ. The value of λ is determined by the characteristics of the negotiation scenario (i.e., discounting factor δ) and the concession degree of the negotiating partner. The smaller the discounting factor δ is, the less actual utility we will receive as time goes by, which means more risk we are facing when we continue exploiting the negotiating partner. Therefore the value of λ should be decreased with the decreasing of the discounting factor δ. The concession degree of the negotiating partner is estimated based on its past behaviors. Intuitively, the more number of new negotiation outcomes that the negotiating partner has recently proposed, the more it is willing to make concession to end the negotiation. Specifically, the negotiation partner's concessive degree σ^t is defined as the ratio of new negotiation outcomes it proposed within the most recent finite negotiation history. If we predict that the negotiating partner is becoming more concessive, we can take advantage of this prediction by postponing the time we stop exploitations, i.e., increasing the

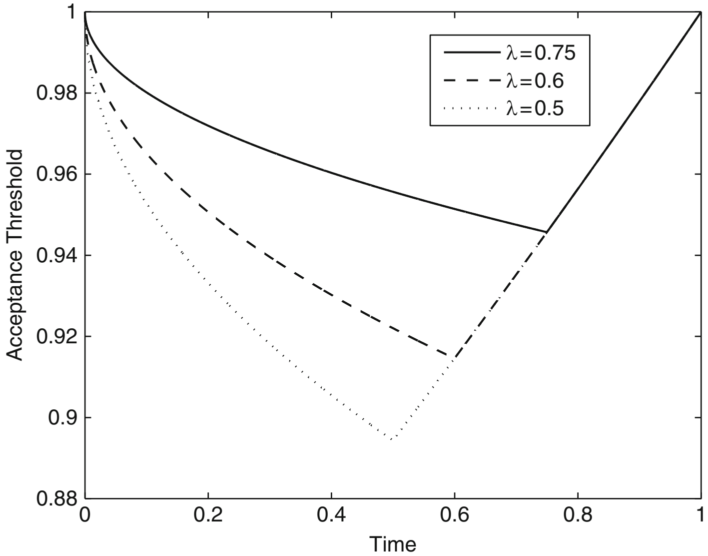

Fig. 11.1 The dynamics of the acceptance threshold ($u^{max} = 1, \alpha = 0.5$ and $\delta = 0.8$)

value of λ. Initially the value of λ is determined by the discounting factor δ only since we do not have any information on the negotiating partner yet. After that, it is adaptively adjusted based on the estimation of the concession degree of the negotiating partner. The overall adjustment rule of λ during negotiation is shown in Fig. 11.2.

11.3.2 How to Propose Bids to the Opponent

In previous section, we have described how CUHKagent determines its acceptance threshold. When the proposed offered by its opponent is not satisfactory, it needs to propose a counter offer higher than its current acceptance threshold to its opponent. Given the current acceptance threshold, any negotiation outcome over which CUHKAgent's utility is higher than the acceptance threshold can be a reasonable outcome to propose. To maximize the likelihood that the offer will be accepted by the opponent, we need to predict the negotiation outcome ω_{max} which can maximize the opponent's utility among the set $\mathscr{C}$ of candidate negotiation outcomes.

To obtain ω_{max}, we need to estimate the opponent's private preference based on its past negotiation moves. Different approaches [1, 3, 7, 8] have been proposed to explicitly estimate the negotiating partner's utility function in bilateral negotiation scenarios. To make the estimation feasible with the limited information available, we

---

Initial values

- λ_0 - the minimum value of λ,
- β - the controlling variable determining the way the value of λ changes with respect to the discounting factor δ, i.e., boulware ($\beta < 1$), conceder ($\beta > 1$) or linear ($\beta = 1$),
- σ^t - the estimation of the negotiating partner's concessive degree at time t,
- γ- the controlling variable determining the way the value of λ changes with respect to σ^t , i.e., boulware ($\gamma < 1$), conceder ($\gamma > 1$) or linear ($\gamma = 1$),
- w - the weighting factor adjusting the relative effect of σ^t on the non-exploitation point λ.

if $t = 0$ **then**
 $\lambda = \lambda_0 + (1-\lambda_0)\delta^{\beta}$
end if
if $0 < t \leq 1$ **then**
 $\lambda = \lambda + w(1-\lambda)\sigma^{t\gamma}$
end if

Fig. 11.2 Adjustment rule of λ at time t

usually need to put some restrictions on the possible structures that the negotiation partner's utility function can have [3] or assume that the preference profile of the negotiation partner is chosen from a fixed set of profiles [7]. Due to the previous concerns mentioned in Sect. 11.2, instead of estimating the opponent's utility function directly, here we adopt a more general way to predict the current best negotiation outcome for the opponent based on model-free reinforcement learning approach. The only assumption we need here is that the negotiating opponent is individually rational and follows some kind of concession-based strategy when proposing bids, which is the most commonly used assumption in both game-theoretic approaches and negotiations [3, 11].

Based on the above assumption, it is natural to assume that the sequence of past negotiation outcomes proposed by the opponent should be in accordance with the decreasing order of its preference over those outcomes. Intuitively, for a value v_i of an item m_i, the earlier and the more frequent it appears in the negotiation outcomes of the past history, the more likely that it weights more in contributing to the negotiation partner's overall utility. Therefore, for each value of each item m_i in the negotiation domain, we keep record of the number of times that it appears in the negotiating partner's past negotiation outcomes and update its value each time a new negotiation outcome ω' is proposed by the opponent as follows,

$$n(\omega'(m_i)) = n(\omega'(m_i)) + \eta^k \quad \forall m_i \in \mathscr{M} \tag{11.2}$$

where ω' is the most recent negotiation outcome proposed by the opponent, η is the discounting factor reflecting the decreasing speed of the relative importance of the negotiation outcomes as time increases, and k is the number of times that the value $\omega'(m_i)$ of item m_i has appeared in the history.

For each negotiation outcome ω, we define its accumulated frequency $f(\omega)$ as the criterion for evaluating the relative preference of the opponent over it. The value of $f(\omega)$ is determined by the value of $n(\omega(m_i))$ for each item $m_i \in \mathscr{M}$ based on the current negotiation history. Formally, for any negotiation outcome ω, its accumulated frequency $f(\omega)$ is calculated as follows,

$$f(\omega) = \sum_{m_i} n(\omega(m_i)) \quad \forall m_i \in \mathscr{M} \tag{11.3}$$

The negotiation outcome ω_{max} is selected based on the ϵ-greedy exploration mechanism. With probability $1 - \epsilon$, it chooses the negotiation outcome with the highest f-value from the set $\mathscr{C}$ of candidate negotiation outcomes, and chooses one negotiation outcome randomly from $\mathscr{C}$ with probability ϵ.

11.4 Conclusion

In this paper, we propose an adaptive negotiating agent—CUHKAgent—to perform automatic negotiation in bilateral multi-issue negotiation scenarios. We introduce the concept of non-exploitation point λ to adaptively adjust the agent's concession degree to its negotiating opponent, and propose a reinforcement-learning based approach to determine the optimal proposal for the opponent to maximize the possibility that the offer will be accepted by the opponent. As future work, one worthwhile direction is to further refine the estimation of the negotiating partner's concessive degree to make more effective exploitation on the negotiating opponent, by taking into consideration the magnitude of the utility that the opponent proposes.

References

1. Faratin, P., Sierra, C., Jennings, N.R.: Using similarity criteria to make negotiation trade-offs. Artif. Intell. **142**(2), 205–237 (2003)
2. Saha, S., Biswas, A., Sen, S.: Modeling opponent decision in repeated one-shot negotiations. In: AAMAS'05, pp. 397–403 (2005)
3. Hindriks, K., Tykhonov, D.: Opponent modeling in automated multi-issue negotiation using Bayesian learning. In: AAMAS'08, 331–338 (2008)
4. Brzostowski, J., Kowalczyk, R.: Predicting partner's behaviour in agent negotiation. In: AAMAS '06, 355–361 (2006)
5. Hao, J.Y., Leung, H.F.: An efficient negotiation protocol to achieve socially optimal allocation. In: PRIMA'12, 46–60 (2012)
6. Zeng, D., Sycara, K.: Bayesian learning in negotiation. In: AAAI Symposium on Adaptation, Co-evolution and Learning in Multiagent Systems, pp. 99–104 (1996)

7. Zeng, D., Sycara, K.: Bayesian learning in negotiation. Int. J. Hum. Comput. Syst. **48**, 125–141 (1998)
8. Coehoorn, R.M., Jennings, N.R.: Learning an opponent's preferences to make effective multi-issue negotiation trade-offs. In: Proceedings of ICEC'04, ACM Press, 59–68 (2004)
9. Baarslag, T., Fujita, K., Gerding, E.H., Hindriks, K., Ito, T., Jennings, N.R., Jonker, C., Kraus, S., Lin, R., Robu, V., Williams, C.R.: Evaluating practical negotiating agents: results and analysis of the 2011 international competition. Artif. Intell. **198**, 73–103 (2013)
10. Hao, J.Y., Leung, H.F.: Abines: an adaptive bilateral negotiating strategy over multiple items. In: Proceedings of IAT'12, vol. 2, pp. 95–102 (2012)
11. Osborne, M.J., Rubinstein, A.: A Course in Game Theory. MIT Press, Cambridge (1994)