



UNIVERSITY
OF WOLLONGONG
AUSTRALIA

University of Wollongong
Research Online

Faculty of Engineering and Information Sciences -
Papers: Part A

Faculty of Engineering and Information Sciences

2013

An adaptive bilateral negotiation model based on Bayesian learning

Chao Yu

University of Wollongong, cy496@uowmail.edu.au

Fenghui Ren

University of Wollongong, fren@uow.edu.au

Minjie Zhang

University of Wollongong, minjie@uow.edu.au

Publication Details

Yu, C., Ren, F. & Zhang, M. (2013). An adaptive bilateral negotiation model based on Bayesian learning. *Studies in Computational Intelligence*, 435 75-93.

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library:
research-pubs@uow.edu.au

An adaptive bilateral negotiation model based on Bayesian learning

Abstract

Endowing the negotiation agent with a learning ability such that a more beneficial agreement might be obtained is increasingly gaining attention in agent negotiation research community. In this paper, we propose a novel bilateral negotiation model based on Bayesian learning to enable self-interested agents to adapt negotiation strategies dynamically during the negotiation process. Specifically, we assume that two agents negotiate over a single issue based on time-dependent tactic. The learning agent has a belief about the probability distribution of its opponent's negotiation parameters (i.e., the deadline and reservation offer). By observing opponent's historical offers and comparing them with the fitted offers derived from a regression analysis, the agent can revise its belief using the Bayesian updating rule and can correspondingly adapt its concession strategy to benefit itself. By being evaluated empirically, this model shows its effectiveness for the agent to learn the possible range of its opponent's private information and alter its concession strategy adaptively, as a result a better negotiation outcome can be achieved. © Springer-Verlag Berlin Heidelberg 2013.

Keywords

model, bilateral, bayesian, learning, negotiation, adaptive

Disciplines

Engineering | Science and Technology Studies

Publication Details

Yu, C., Ren, F. & Zhang, M. (2013). An adaptive bilateral negotiation model based on Bayesian learning. *Studies in Computational Intelligence*, 435 75-93.

An Adaptive Bilateral Negotiation Model Based on Bayesian Learning

Chao Yu, Fenghui Ren, and Minjie Zhang

Abstract. Endowing the negotiation agent with a learning ability such that a more beneficial agreement might be obtained is increasingly gaining attention in agent negotiation research community. In this paper, we propose a novel bilateral negotiation model based on Bayesian learning to enable self-interested agents to adapt negotiation strategies dynamically during the negotiation process. Specifically, we assume that two agents negotiate over a single issue based on time-dependent tactic. The learning agent has a belief about the probability distribution of its opponent's negotiation parameters (i.e., the deadline and reservation offer). By observing opponent's historical offers and comparing them with the fitted offers derived from a regression analysis, the agent can revise its belief using the Bayesian updating rule and can correspondingly adapt its concession strategy to benefit itself. By being evaluated empirically, this model shows its effectiveness for the agent to learn the possible range of its opponent's private information and alter its concession strategy adaptively, as a result a better negotiation outcome can be achieved.

1 Introduction

Negotiation is a fundamental topic in multi-agent systems because it allows self-interested agents to achieve mutually beneficial agreements and partition resources efficiently and effectively [9]. In recent years, researchers have paid their increasing attention to the integration of learning techniques into agent negotiation [1] [2] [3] [4] [5] [8] [11] [13]. In this type of learning circumstances, agents need adapt themselves to the changes of opponents and/or the environment through learning in order to achieve a satisfactory result. One promising paradigm of involving learning

Chao Yu · Fenghui Ren · Minjie Zhang
School of Computer Science and Software Engineering,
University of Wollongong, Wollongong, NSW, 2522, Australia
e-mail: cy496@uowmail.edu.au,
fren,minjie@uow.edu.au

in negotiation is through opponent modeling to let agents learn the model of their opponent/teammates (such as preferences, private information and capabilities etc.) in the environment so as to make good predications for expected behaviors and to act accordingly to cooperate with the teammates more effectively or take the best advantage of the opponents. However, due to the essence of competition, privacy and uncertainty in real-life negotiation [10], negotiators are always unwilling to reveal their private information (e.g., parameters such as the deadline, reserve price, or strategy profiles) to their opponents in case of being forced to a worse outcome, thus making learning in negotiation a challenging problem.

In current literature, a number of approaches have been developed by employing agents learning methods into negotiation process. Zeng and Sycara proposed an approach based on Bayesian learning to learn the opponent's reserve price [16]. Their approach assumes that agents have priori knowledge about the opponent's bidding strategy. This assumption may not always be true in real-world negotiations. Hindriks and Tykhonov also proposed an approach to discover opponent's information [5] by using Bayesian learning based on the assumptions that 1) agents know the opponent's weights ranking on negotiation issues and 2) all agents' preferences can be modeled by three proposed functions, which may impact the use of this approach in a wide range when these assumptions conflict with the real world situations. Ren and Zhang introduced approaches based on regression analysis to predict the opponent's concession strategy by using the historical offers only [6] [7]. However, their approaches did not give further advice on how to adapt agent self's concession strategy based on the learning results. Brzostowski and Kowalczyk also presented a way to estimate partners' behaviors in different types of agents, based only on the historical offers in the current negotiation [10]. However, the accuracy of classification on partners' types may impact the accuracy of prediction results. The current challenging issues in agent learning during negotiation include (1) how to design a learning method without priori knowledge of the opponent's private information, (2) how to develop an effective learning strategy only based on the historical offers of current negotiation, and (3) how to produce a constructive guidance from learning to adapt agent's negotiation behaviors so as to achieve a better negotiation outcome.

This research attempts to address the above three challenging issues. In this paper, we propose a novel model by combining Bayesian learning and a regression analysis approach to dynamically learn the opponent's negotiation deadline and reservation offer for an adaptive negotiation process. To be more specific, firstly, a negotiation agent defines some regions and evenly initialize the probability of each region. The probability here indicates how likely that the opponent's deadline and reservation offer are located in the corresponding region. By using the predefined regions, the agent can have some estimations on the opponent's negotiation behaviors. Secondly, by using the regression analysis, the differences between the opponent's real negotiation behavior and the agent's estimated results are calculated. The more similar between the opponent's real behavior and an estimated behavior, the more likely that the opponent's real deadline and reservation offer are located in the corresponding region. Thirdly, based on the similarities between the opponent's real behavior and the estimated behaviors, the probabilities assigned to each region will be updated

dynamically through Bayesian learning updating rule. Lastly, the agent will propose a countermeasure for each estimated behavior of the opponent, and all countermeasures will be combined based on the likelihood of each estimated behavior. The combined result will be employed by the agent to perform a reasonable reaction. During the negotiation, each region's probability will be dynamically updated and gradually close to the real situation. Thus, the agent will also gradually adapt its negotiation strategy to reach a better negotiation outcome. Our model only uses historical offers in the current negotiation, without requesting prior knowledge about the environment and the opponent.

The remainder of this paper is structured as follows. In Section 2, we recap the general negotiation model, especially the basic principles of the time dependent tactic. The proposed learning model is introduced in detail in Section 3, and in Section 4 empirical evaluation and analysis are presented. The discussion and related work are given in Section 5. Finally the paper is concluded and some directions for future work are outlined in Section 6.

2 A General Negotiation Model

Before laying out our learning model, we give a brief description of a time dependent, bilateral single-issue negotiation model, which is widely used in many applications. Let i ($i \in \{b, s\}$) represent a negotiator, i.e., b for a buyer agent and s for a seller agent. Both agents have an initial price IP_i and reserve price RP_i for the negotiating issue. The interval $[IP_i, RP_i]$ indicates the range of all the possible agreements, and can be normalized in-between $[0, 1]$ using a utility function. In this paper, we choose the widely accepted linear utility function [15] shown in Equation 1:

$$u_i(p_i) = \frac{p_i - RP_i}{IP_i - RP_i} \quad i \in \{b, s\} \quad (1)$$

where p_i is the value of an offer in the range of $[IP_i, RP_i]$.

In time dependent tactic, agent i concedes its utility $u_i(t)$ under the time constraint. At the beginning of negotiation, agent i has its highest utility of 1 for the *initial price*. As the negotiation proceeds on, the utility $u_i(t)$ decreases according to the decision functions [15], which are a family of polynomial functions given by Equation 2.

$$u_i(t) = 1 - \left(\frac{t}{T_i}\right)^\beta \quad i \in \{b, s\} \quad (2)$$

where T_i is the deadline of agent i and β is the concession parameter. Figure 1 shows three different concession strategies called *Conceder*, *Boulware*, *Linear*, respectively, signifying different concession rates in the negotiation process.

- *Conceder*: When $0 < \beta < 1$, the agent decreases its utility quickly at the early stage of negotiation and slowly when the deadline is approaching.
- *Linear*: When $\beta = 1$, the agent's utility decreases at a constant rate throughout the negotiation process.

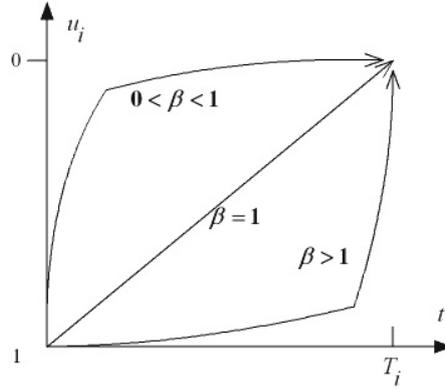


Fig. 1 Three different kinds of concession strategies [15]

- *Boulware*: When $\beta > 1$, it is the contrary of Conceder, which means slow concession at the beginning but quick concession at the late stage of negotiation.

When β is settled, the utility $u_i(t)$ can be computed during the negotiation. As a result, the agent can give a counter offer at time t according to the following offer generating equation.

$$Offer_i(t) = RP_i + u_i(t)(IP_i - RP_i) \quad i \in \{b, s\} \quad (3)$$

Combining Equation 2 and Equation 3, the offer generating Equation 3 is rewritten as Equation 4.

$$Offer_i(t) = IP_i + (RP_i - IP_i)\left(\frac{t}{T_i}\right)^\beta \quad i \in \{b, s\} \quad (4)$$

In a non-learning negotiation setting, once an agent sets the value of concession parameter β , the agent will keep this value unchanged through the negotiation process, without any adaptation to the dynamic environment or the revelation of opponent's private information. However, if the agent can learn some useful information from the opponent during the negotiation, it will be able to adapt its original concession strategies and gain more benefits to produce good outcomes for negotiation. In the following section, we will present an adaptive negotiation model using regression analysis and Bayesian learning to enable an agent to alter its concession strategy dynamically, thereby a better outcome will be obtained.

3 An Adaptive Negotiation Model

In this section, an adaptive negotiation model is proposed. This model includes two parts: a learning mechanism and an adaptive concession strategy. Each part will be introduced in detail by Subsections 3.2 and 3.3, respectively. In this paper hereinafter, the discussion is taken from the perspective of the buyer agent unless

otherwise specified. However, such a discussion will not lose the generality of our model, i.e. a seller agent can also use our model to learn its opponent's behaviors.

3.1 Model Description

As we can see from Equation 4, the parameters of deadline and reserve price are two main factors dominating the negotiation process and outcomes. If agents can obtain the information about these two parameters from the opponent, a better strategy can be employed to increase agents' benefits and/or the negotiation efficiency. Our aim is to model the opponent in terms of these two private information. Before going deep into our model, we firstly give the definitions needed for further illustration.

Definition 0.1. Let x-axis represent negotiation time and y-axis represent the negotiation price. A *detecting region* $DetReg$ is a rectangle in this two-dimensional area to present an estimation of the opponent's deadline and reserve price. This area is defined by a 4-tuple $DetReg = (T^l, T^h, P^l, P^h)$, where T^l, T^h are the estimated lower and upper boundary of the opponent's deadline, and P^l, P^h are the estimated lower and upper boundary of the opponent's reserve price.

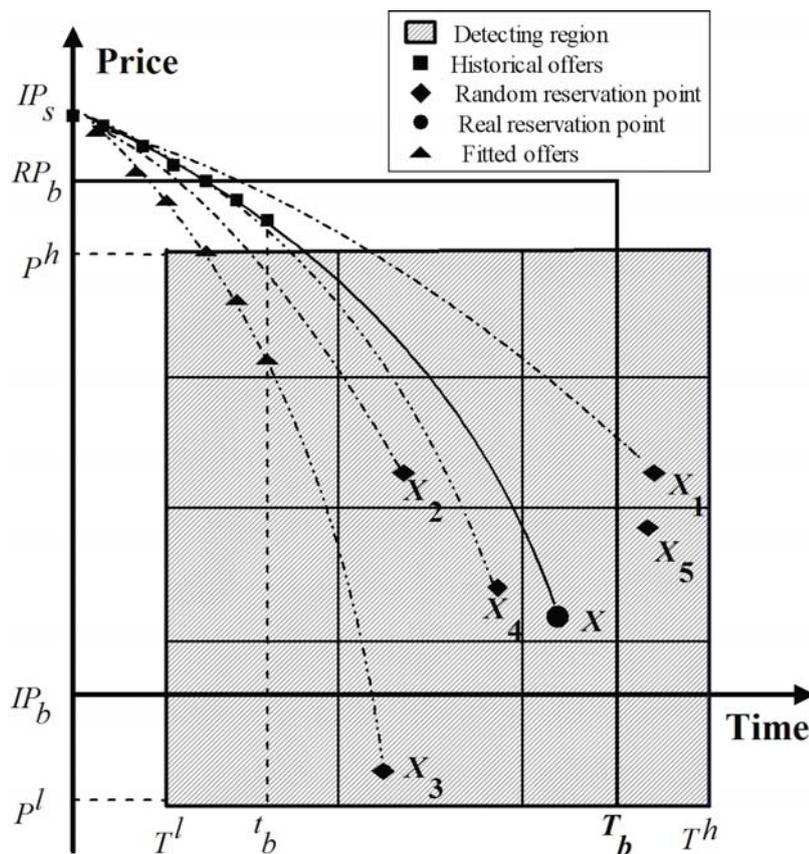


Fig. 2 An example of demonstrating our learning process

As shown in Figure 2, the shadowy area indicates the detecting region for a buyer agent during a learning process based on Definition 1. T_b is buyer's deadline and t_b is the current time in negotiation. IP_b and RP_b represent the buyer's initial price and reserve price, respectively, and IP_s is the seller's initial price. Points appeared in the detection region of the figure will be explained in Definition 3. The lines shown in the figure will be explained in Subsection 3.2 during introducing the learning mechanism.

A buyer agent can initialize the value of each component of *DetReg* according to its estimation about seller's private information. The more precise the estimation is, the smaller the detecting region will be and the buyer can strive for a better result because more errors can be avoided when the buyer agent adapts its concession strategy based on this estimation.

After confirming the detecting region by the buyer agent, this region will be further divided into smaller areas according to $N = (N^t, N^p)$ in which N^t denotes that the detecting region is evenly divided into N^t columns on the x-axis (i.e. time values), and N^p stands for the row number on the y-axis (i.e. price values) in the detecting region. In this way, the detecting region can be divided into a number of smaller blocks, called *detecting cells*. The total number of detecting cells in a detecting region is represented by N_{all} and can be calculated by the formula $N_{all} = N^t \times N^p$. Fig. 2 exemplifies a scenario with $N = (3, 4)$ and there are totally 12 detecting cells in the whole detecting region.

Definition 0.2. A *detecting cell* C_i ($i \in 1, 2, \dots, N_{all}$) is a divided block in the detecting region, which can be denoted by a 4-tuple $C_i = (t_i^l, t_i^h, p_i^l, p_i^h)$ where t_i^l, t_i^h are the lower and higher boundaries of time in the cell and p_i^l, p_i^h are the lower and higher boundaries of price in the cell, respectively.

Definition 0.3. A *random reservation point* $X_i(t_i^x, p_i^x)$ is a randomly selected point in each cell C_i , where $t_i^l < t_i^x < t_i^h$ and $p_i^l < p_i^x < p_i^h$.

In Figure 2, points X_1, X_2, X_3, X_4 are several random reservation points in the detecting region and point X is the real reservation point of the opponent. The detecting cell is a region where seller's real reservation point X might be located. That means the real reservation point X might be out of the detecting region in real case. The buyer agent has some belief about the probability distribution of all the detecting cells. The probability of each cell signifies the likelihood that the opponent's real reservation point X might be located in this cell. This belief can be revised more precisely through learning from opponent's historical offers (see Subsection 3.2). Based on this learning result, the agent can adjust its concession strategy adaptively (see Subsection 3.3) to gain more profit over its opponent.

3.2 The Learning Mechanism

The purpose of this leaning mechanism is to let the agent revise its belief about the probability distribution of the cells in the detecting region. Because the agent has no

knowledge about the opponent, it is hard to determine the precise location of the real reservation point. However the agent can observe its opponent's historical offers to renew the belief about the approximate range of the reservation point. This mechanism consists of two parts, a regress analysis and a Bayesian learning. In regression analysis, (1) an agent chooses a random reservation point in every detecting cell first, based on the belief that this point is the reservation point of the opponent; (2) the agent conducts the regression analysis for all random reservation points corresponding to all detecting cells, respectively; (3) the agent compares the fitted offers on each regression line with opponent's historical offers by the non-linear correlation. By this way, resemblance between the selected random reservation point and the opponent's real reservation point can be calculated. The bigger the non-linear correlation between two lines is, the more alike they will be. This also means that the randomly chosen reservation point has a bigger possibility to be the real reservation point. Then by using Bayesian learning, the agent's belief on the probability distribution will be dynamically updated at every step of the negotiation. The regression analysis and our Bayesian learning method are introduced in the following two subsections, respectively.

3.2.1 Regression Analysis

Before the leaning process, the buyer should initialize $DetReg$, N as well as the probability distribution in each detecting cell, which presents the likelihood that the seller's reservation point is in this cell. When the learning begins, the buyer can do the following steps sequentially.

Step 1: At round t_b , the buyer selects a random reservation point $X_i(t_i^x, p_i^x)$ in each cell C_i of the detecting region;

Step 2: Using each point $X_i(t_i^x, p_i^x)$ chosen in Step 1, the buyer calculates the regression line l_i based on the seller's historical offers $O_{t_b} = \{p_0, p_1, \dots, p_{t_b}\}$ until round t_b . Based on Equation 4, the following power regression function is generated to calculate the regression curve.

$$Offer_i(t) = p_0 + (p_i^x - p_0) \left(\frac{t}{t_i^x}\right)^b \quad (5)$$

where p_0 is the *initial price* of seller. The regression coefficient b is the concession parameter β in the utility function in Equation 4. Then we can calculate coefficient b based on seller's historical offers O_{t_b} by Equation 6 as proposed in [6].

$$b = \frac{\sum_{i=1}^{t_b} t_i^* p_i^*}{\sum_{i=1}^{t_b} t_i^{*2}} \quad (6)$$

where $p_i^* = \ln \frac{p_0 - p_i}{p_0 - p_i^x}$, $t_i^* = \ln \frac{t}{t_i^x}$. In Figure 2, the solid line is the curve of the seller's historical offers while the dashed line is the regression curve based on each random reservation point.

Step 3: Based on the calculated regression line l_i given by Equation 5 and 6, the buyer can calculate the fitted offers $\hat{O}_{t_b} = \{\hat{p}_0, \hat{p}_1, \dots, \hat{p}_{t_b}\}$ at each round.

Step 4: The buyer calculates the non-linear correlation between seller's historical offers O_{t_b} and the fitted offers \hat{O}_{t_b} . The coefficient of nonlinear correlation γ can be calculated by Equation 7.

$$\gamma = \frac{\sum_{i=1}^{t_b} (p_i - \bar{p})(\hat{p}_i - \bar{\hat{p}})}{\sqrt{\sum_{i=1}^{t_b} (p_i - \bar{p})^2 \sum_{i=1}^n (\hat{p}_i - \bar{\hat{p}})^2}} \quad (7)$$

where $\bar{\hat{p}}$ is the average value of all the fitted offers till time t_b and \bar{p} represents the average value of all the historical offers of the seller. The non-linear correlation γ , where $(0 \leq \gamma \leq 1)$, is a parameter reflecting the non-linear similarity between the fitted offers and the historical offers, which can be used as a criterion to evaluate the resemblance between the random reservation point X_i and seller's real reservation point X . This is an important parameter to be used in Bayesian learning for the belief updating as described in the following section.

3.2.2 Bayesian Learning

In general, Bayesian learning can be used when an agent has a set of hypotheses about its opponent's information. The belief about the probability distribution of these hypotheses can be revised through a posterior probability by observing the outcome of its opponent. In our model, we define the hypothesis space as $H_i, (i \in 1, 2, 3, \dots, N_{all})$, where N_{all} is the total cell number in the detecting region. Each hypothesis H_i stands for the assumption that seller's reservation point X is in cell C_i . The prior probability distribution, denoted by $P(H_i), (i \in 1, 2, 3 \dots N_{all})$, signifies the agent's belief about the hypothesis, that is, how likely the hypothesis fits the real situation. At first, the agent can initialize the probability distribution of the hypotheses based on some public information if available, otherwise a uniform distribution $P(H_i) = 1/N_{all}$ is assigned.

During each round of negotiation t_b , the probability of each hypothesis can be altered by the Bayesian updating rule given in Equation 8.

$$P(H_i|O) = \frac{P(H_i)P(O|H_i)}{\sum_{k=1}^{N_{all}} P(O|H_k)P(H_k)} \quad (8)$$

where the conditional probability $P(O|H_i)$ represents the likelihood that outcome O might happen based on hypothesis H_i . In our learning model, the agent has no information about its opponent, thus the observed outcome O is opponent's historical offers $O_{t_b} = \{p_0, p_1, \dots, p_{t_b}\}$. The conditional probability $P(O|H_i)$ thereby means how likely seller's historical offer O_{t_b} can happen based on the hypothesis H_i that seller's real reservation point X is in cell C_i . The posterior probability $P(H_i|O)$ is a renewed belief based on the observed outcome O and at next round, the agent will update the prior probability $P(H_i)$ using the posterior probability $P(H_i|O)$, thus a more precise estimation is achieved by using Equation 8.

To let the Bayesian learning rule work, the most critical problem is how to obtain the conditional probability $P(O|H_i)$. Most approaches using Bayesian learning method usually require a priori knowledge as the conditional probability, such as the one in [16]. However, our learning model does not require any priori knowledge about the opponent and works based only on the historical offers received until t_b from the opponent. By comparing the fitted points \hat{O}_{t_b} on the regression line based on each random reservation point X_i with the historical offers O_{t_b} , the conditional probability $P(O|H_i)$ is obtained. The more consistent the fitted offers are with opponent's historical offers, the higher the conditional probability $P(O|H_i)$ will be. As showed at Step 3 in Subsection 3.2.1, the difference between the regression curve and opponent's bidding sequence can be indicated by the non-linear correlation coefficient γ . Thus, we can use the value of γ as the conditional probability.

The learning approach will increase the probability of a hypothesis when the random reservation point selected in the detecting cell is most consistent with the real reservation point of the opponent. However, in some cases, it is possible that the learning may have errors. As seen in Figure 2, compared with point X_5 , point X_4 has a higher non-linear correlation with the real reservation point X , but point X_4 and X are not in the same detecting cell. As a result, the hypothesis that the real reservation point X belongs to the cell where point X_4 is located has a higher probability. Nevertheless, we claim that this situation does not affect the learning effectiveness based on the following two considerations. Firstly, although in certain circumstances, using the non-linear correlation to calculate the difference between the regression line and the real bidding sequence does not necessarily reveal the real situation, the error will be eased through Bayesian learning from a probabilistic point of view. Secondly, even the error exists, the learning approach still works because we only need to find an approximate range of the reservation point, not the precise value of opponent's reservation point. In some cases, the real reservation point X might not be located in the whole detecting region, but those cells which are closer to point X will still have a higher probability compared with other cells.

Another issue that should be taken into account is the learning rate and efficiency. At the early stage of leaning, the hypotheses space can be quite large depending on the value of $DetReg$ and N (recall Subsection 3.1). It is time consuming to keep all the hypotheses in the searching space. Some hypotheses can be precluded from the hypotheses space when the current time and opponent's bidding value have surpassed the detecting cell boundary. For example, for a cell $C_i = [t_i^l, t_i^h, p_i^l, p_i^h]$, if current negotiation time $t_b > t_i^h$, the hypothesis based on this cell is meaningless because the negotiation process has already proved it false.

3.3 The Adaptive Concession Strategy

Through regression analysis and Bayesian learning stated above, a more precise estimation of the opponent's reservation point is derived, represented by the renewed belief of the probability distribution of the hypothesis H_i . Now, the agent needs to take an action to give a counter offer based on this new belief, i.e. which concession

strategy to take and how strong it should be in terms of a value of the concession parameter β . Our adaptive concession strategy includes two parts: the optimal concession strategy described in Subsection 3.3.1 and the combining mechanism described in Subsection 3.3.2.

3.3.1 The Optimal Concession Strategy

There are four scenarios according to different location of the random reservation point. As we believe that the agent is rational, it always strives for a highest utility of its own regardless of its opponent fully. Therefore, in each scenario, the buyer needs to adopt different concession strategies to maximize its expected utility as depicted in Figure 3. In Figure 3, point $b_0(t_0, p_0)$ is buyer's current offer at time t_0 , point $b_T(T_b, RP_b)$ is the buyer's reservation offer at deadline T_b , and point $X_i(t_i^x, p_i^x)$ is the random reservation point of seller. Then the buyer needs to find another point $P(t_p, p_p)$, which is called a concession point, in its negotiation region to set the concession strategy and the value of β .

- **Scenario 1:** $(t_i^x < T_b)$ and $(p_i^x > p_0)$.

In this scenario, the random reservation point X_i is in the buyer's negotiation region. Because the buyer agent is rational, it will always try to gain the maximal utility itself. If the buyer knows that the seller will quit the negotiation at point X_i (i.e., the deadline of the seller t_i^x is shorter than its deadline T_b), the optimal concession strategy for the buyer is to set his bidding price to p_i^x at time t_i^x . Otherwise, if the buyer gives more concession, it cannot achieve the maximal utility after finishing negotiation. On the contrary, a less concession may result in a failure of the negotiation. As illustrated in Figure 3(a), the random reservation point X_i is set to be the concession point P in this case and the dashed line crossing point X_i is the concession line of the buyer.

- **Scenario 2:** $(t_i^x \geq T_b)$ and $(p_i^x \geq p_0)$.

In this scenario, random reservation point X_i is out of the buyer's negotiation region. There are two cases in this scenario according to the different regression lines of the seller. As can be seen in Figure 3(b), in the first case, regression line l_1 traverses the buyer's negotiation region while l_2 does not. In the same way of analyzing in *Scenario 1*, buyer's optimal concession line for l_1 is to pass through the intersection point of the line l_1 and the right boundary of the buyer's negotiation region. Considering that the buyer should give out its reserve price at deadline T_b , for simplicity, we let the buyer's concession line cross the concession point P_1 on the regression line one step ahead of the deadline T_b (i.e., $T_b - 1$) such that a concrete value of the concession parameter β can be computed. As for the second case, the regression curve l_2 has no intersection with the buyer's negotiation region, which means even the buyer concedes, the negotiation based on this random reservation point is doomed to fail. Nevertheless, the buyer will spare no efforts to reverse this unfavorable situation. With this aim, it will give the reserve price at next round $(t_0 + 1)$. To compute a value of β , we choose a variable ϕ_{max} ($0 < \phi_{max} < 1$) which is quite close to 1. The concession point P_2 in

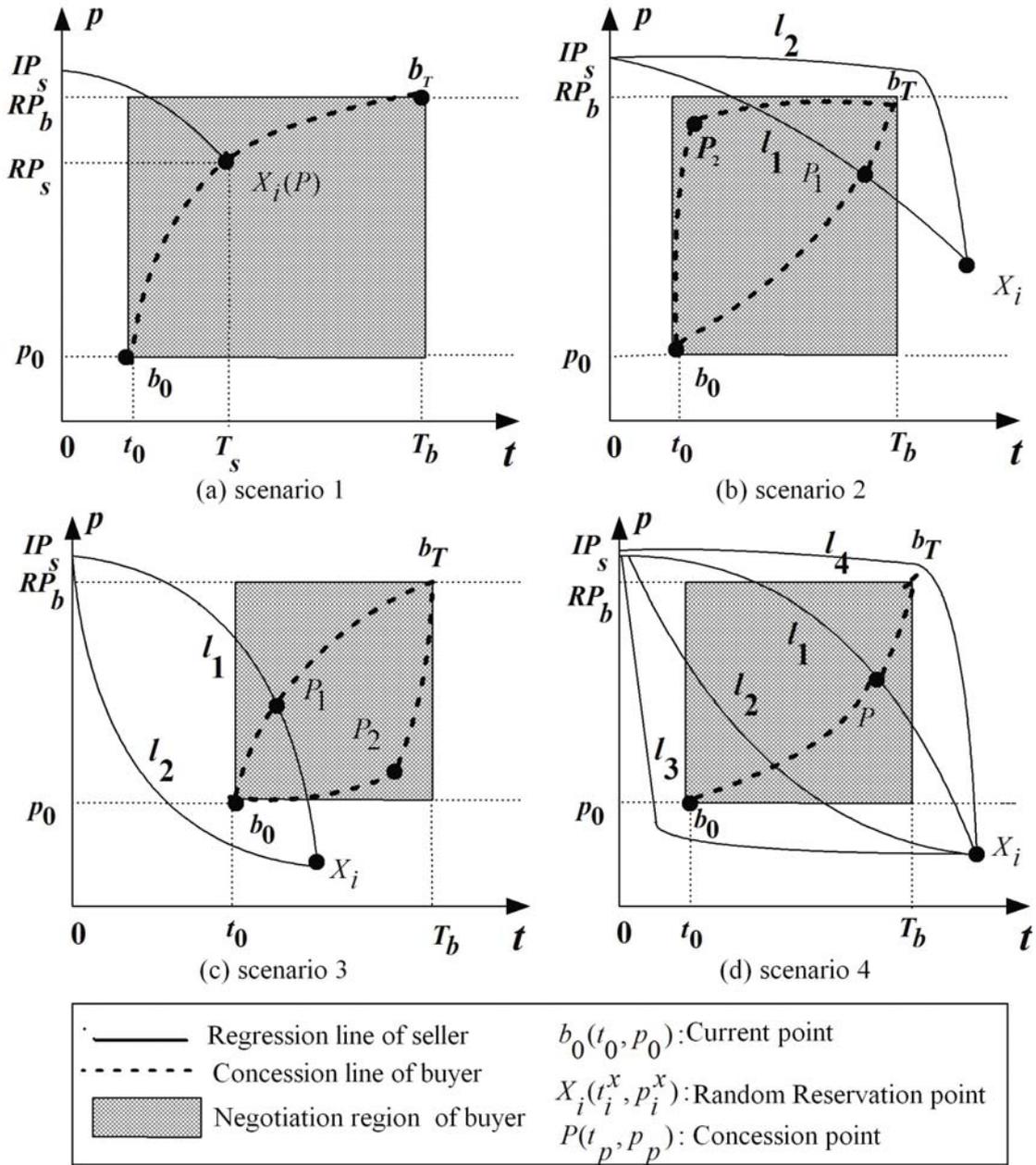


Fig. 3 Four Scenarios of Concession Strategy

this case is set to be $P_2(b_0 + 1, \phi_{max} \cdot RP_b)$ so as to make the price at next round close to the reserve price of RP_b and finally to give out the reserve price RP_b at the next round, i.e., its deadline.

- **Scenario 3:** $(t_i^x < T_b)$ and $(p_i^x < p_0)$.

There are also two cases in this scenario, which can be signified by l_1 and l_2 shown in Figure 3(c). As for case 1, the optimal strategy of the buyer is to cross the intersection of l_1 and bottom line of the buyer's negotiation region. To compute a value of β , we set the concession point P_1 be the point one step earlier than the intersection point on the regression line l_1 . As for case 2, the line

l_2 does not go through the buyer's negotiation region. In this case, the optimal strategy for the buyer is to keep its price unchanged until $T_b - 1$ and then gives its reserve price at the deadline. To compute the value of β , we can set the price at concession point P_2 very close to current price p_0 . Similarly, a variable ϕ_{min} ($0 < \phi_{min} < 1$), which is quite close to 0, can be chosen to set the price at next round to $(1 + \phi_{min}) \cdot p_0$ such that this price will keep almost the same as the current price p_0 .

- **Scenario 4:** ($t_i^x \geq T_b$) and ($p_i^x \leq p_0$).

This scenario, which is a combination of the former Scenarios 2 and 3, is the most complicated case of all. Each line of l_1 , l_2 , l_3 and l_4 can be analyzed in the same way as stated in the previous scenarios. In Figure 3(d), we depict the concession line based on l_1 as an example.

3.3.2 The Combining Mechanism

We have given out all possible situations of the random reservation points and the corresponding optimal concession strategies that the buyer can adopt to increase its utility as well as to avoid the failure of negotiation to its best. Because the buyer still uses the family of polynomial functions to concede, the counteroffer from point $b_0(t_0, p_0)$ can be generated by Equation 9 based on Equation 4.

$$Offer_b(t) = p_0 + (RP_b - p_0) \left(\frac{t - t_0}{T_b - t_0} \right)^\beta \quad (t > t_0) \quad (9)$$

Using this equation, we can guarantee that at its deadline T_b , the buyer will give the reserve price RP_b . At current time t_0 , the buyer's offer is p_0 and the buyer concedes in the form of polynomial function. Then given the concession point $P(t_p, p_p)$ in its negotiation region, a new value of parameter $\hat{\beta}$ can be calculated as follows.

$$\hat{\beta} = \log_{\frac{t_p - t_0}{T_b - t_0}} \frac{p_0 - p_p}{p_0 - RP_b} \quad (t_0 < t_p < T_b) \quad (10)$$

We have calculated all the concession values $\hat{\beta}$ for each valid random reservation point in the detecting region, with a probability distribution $P(H_i) = \{p(H_1), p(H_2), \dots, p(H_n)\}$ over these values derived from the regression analysis and Bayesian learning. Now comes to the problem of how to combine all the estimated value of $\hat{\beta}$ to an overall value. Let $\hat{\beta}_i$ ($i \in \{1, 2, \dots, n\}$) be the estimated concession value calculated from the concession point based on the random reservation point in cell C_i . $P(H_i)$ is the probability of the $\hat{\beta}_i$, presenting the weighting proportion of the corresponding $\hat{\beta}_i$ in all the concession values. The value of $\hat{\beta}_i$ signifies the concession degree of the agent, which can be represented by the area between the concession line and the time axis, which is called *concession area*. As can be seen from Figure 4, the concession area of $\hat{\beta}_1$ is S_1 , which can be denoted by $S_{b_0 \hat{\beta}_1 b_{T_b}}$. Let S_i be the

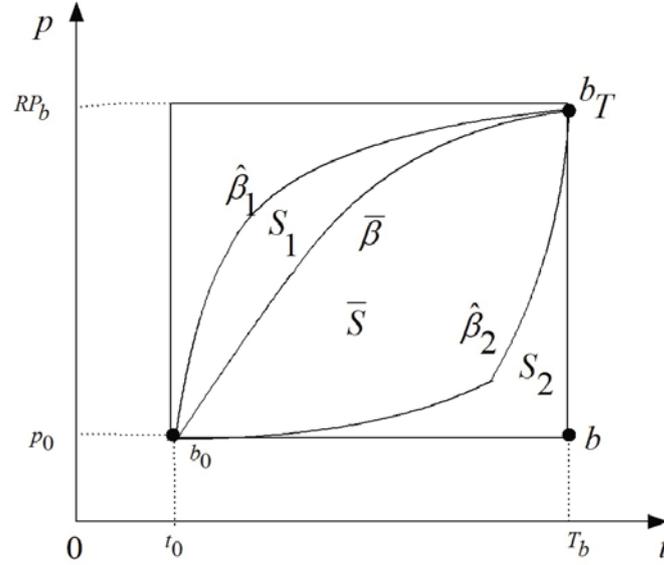


Fig. 4 Combination of the parameter β

concession area of $\hat{\beta}_i$ and let the concession area of the overall concession parameter $\bar{\beta}$ be \bar{S} . Based on Equation 9, we can have the following equations.

$$\bar{S} = \int_{t_0}^{T_b} [p_0 + (RP_b - P_0) \left(\frac{t - t_0}{T_b - t_0} \right)^{\bar{\beta}}] dt \quad (11)$$

$$\sum_{i=1}^n P(H_i) S_i = \sum_{i=1}^n P(H_i) \int_{t_0}^{T_b} [p_0 + (RP_b - P_0) \left(\frac{t - t_0}{T_b - t_0} \right)^{\hat{\beta}_i}] dt \quad (12)$$

because,

$$\bar{S} = \sum_{i=1}^n P(H_i) S_i \quad (13)$$

we can get the overall concession parameter $\bar{\beta}$ as follows:

$$\bar{\beta} = \frac{1}{\sum_{i=1}^n \frac{P(H_i)}{1 + \hat{\beta}_i}} - 1 \quad (14)$$

Then the buyer can set its concession parameter as $\bar{\beta}$ to give counter offer based on Equation 9 at every step of the negotiation. Each $\hat{\beta}_i$ is changing at each step according to the randomly selected reservation point and the corresponding $P(H_i)$ is revised by Bayesian learning throughout the negotiation process. Thus the concession parameter $\bar{\beta}$ adopted by the buyer at each step is totally different, making the negotiation an adaptive process in the point view of the learning agent buyer.

4 Experiment

4.1 Experimental Setting

In the experiment, a buyer and a seller negotiate over the price ranged in-between \$0 ~ \$100. In order to simplify the comparison process, we set the buyer agent's initial price to \$0 and the seller agent's initial price to \$100. The buyer's reserve price is randomly selected in-between \$50 ~ \$100 and seller's reserve price is randomly selected in-between \$0 ~ \$50. Such a setting ensures that the agreement zone between the two agents always exists. Our agents' deadlines are randomly selected in-between [20, 40], and the concession strategies are randomly selected in-between [0.5, 2]. The negotiation parameter initialization is showed in Table 1.

Table 1 Negotiation Parameters initialization

Agent	IP_i	RP_i	T_i	β_i
Buyer ($i=b$)	\$0	[\$50,\$100]	[20,40]	[0.5,2]
Seller ($i=s$)	\$100	[\$0,\$50]	[20,40]	[0.5,2]

To provide a benchmark we compare our negotiation model with the general NDF model. In the general NDF model, both agents randomly initialize their negotiation parameters according to Table 1, and keep these parameters unchanged during the negotiation process. On the contrary, in our model, the buyer agent will learn how to adjust its concession strategy adaptively to reach a better negotiation outcome. To use the learning mechanism, we set $\phi_{min} = 0.01$, $\phi_{max} = 0.99$. We choose two initialization of detecting region as $DetReg1 = (0, 1.5T_b, 0, RP_b)$, $DetReg2 = (T_b, 2T_b, 0.5RP_b, 1.5RP_b)$ to test its affect on the learning result. We outline four cases according to the different numbers of detecting cells as shown in Table 2.

Table 2 Four scenarios of different detecting cell numbers

Case	N^t	N^p	N_{all}
1	4	4	16
2	8	8	64
3	16	16	256
4	20	20	400

4.2 Results and Analysis

As our model depends on the regression analysis which may yield errors as stated before, we do not expect the learning result to be completely precise. Further more, many factors affect the learning process such as the number of detecting cells, the

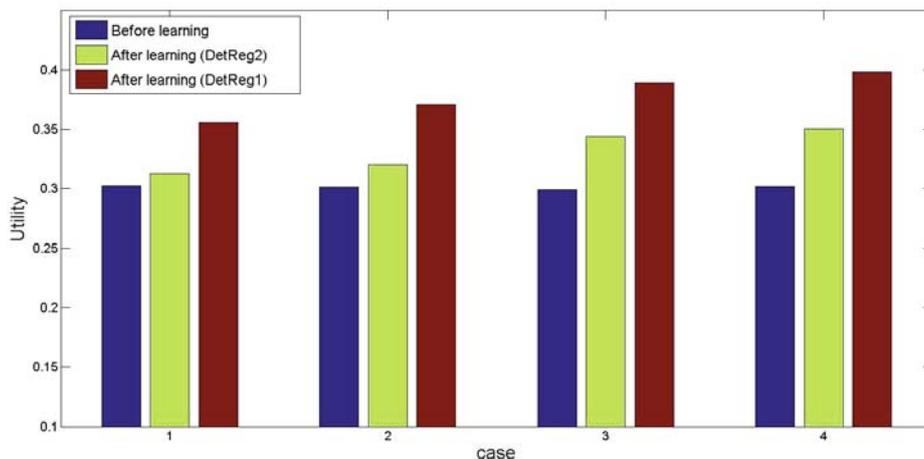


Fig. 5 The average utility in different cases before & after learning

initialization of the detecting region and different values of ϕ_{min} , ϕ_{max} . The objective of this experiment is therefore carried out to analyze the overall performance of this learning approach considering these uncertainties and potential errors.

We run 100 episodes for each case to show the generality and robustness of our model. The results of this experiment are presented in Figure 5. The x-axis indicates the four cases and the y-axis indicates the average utility of the buyer in each case. The blue bars represent the buyer's average utility gained by NDF model. The brown and green bars represent the buyer's average utility gained by using our model when the detecting region is initialized as *DetReg1* and *DetReg2*, respectively. We can see from Figure 5, the brown and green bars are higher than the blue bars in all cases, and gradually increase as the number of the detecting cells increase. Such experimental results indicate that: (1) using our learning mechanism and the adaptive concession strategy will result in a higher utility than the static concession strategy; and (2) as the total number of detecting cells increases, the agent has a more precise estimation of the opponent's reservation point, thus can result in a higher utility. From Figure 5, we can also see that there are some difference between the learning results when the detecting region is initialized differently. More specifically, the average utilities when the detecting region is initialized as *DetReg1* are higher than the utilities when the detecting region is initialized as *DetReg2*. This result can be explained by the fact that when the detecting region is *DetReg2*, the opponent's (seller's) real reservation point is out of the detecting region all the time. Therefore the cell with the highest probability in the detecting region (i.e., the cell that the opponent's real reservation point is most likely located in) cannot reflect the true situation, making the learning result in *DetReg2* be inferior to that in *DetReg1*.

In order to illustrate the dynamic adaptation of the concession parameter β , we give out the whole negotiation process to show how the buyer agent changes its

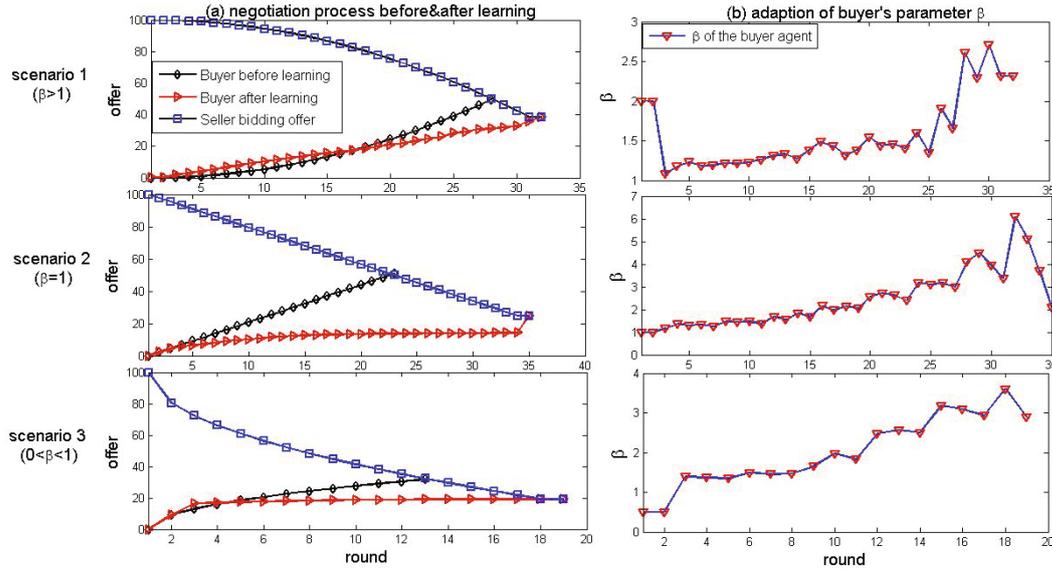


Fig. 6 The negotiation processes before & after learning

concession strategy adaptively. We select three scenarios when detecting region is *DetReg1* with the negotiation parameters as follows:

- Scenario 1 ($1 < \beta < 2$): $RP_b = \$69.58$, $RP_s = \$11.38$, $T_b = 32$, $T_s = 36$, $\beta_b = \beta_s = 2.0$
- Scenario 2 ($\beta = 1$): $RP_b = \$81.04$, $RP_s = \$17.82$, $T_b = 35$, $T_s = 36$, $\beta_b = \beta_s = 1.0$
- Scenario 3 ($0 < \beta < 1$): $RP_b = \$50.34$, $RP_s = \$11.38$, $T_b = 30$, $T_s = 22$, $\beta_b = \beta_s = 0.5$

Figures in 6(a) give the negotiation process between both agents before & after learning. Figures in 6(b) show the adaption of the buyer's concession parameter β . In Scenario 1, the seller adopts the *Boulware* concession strategy. Before learning, the negotiation ends at \$49.53 and both agents' concession strategies keep unchanged through the negotiation process. After learning, the buyer agent adjusts its concession strategy adaptively in terms of parameter β and the agreement price is reduced to \$38.46, which is a better result than that of before learning for the buyer agent. In Scenario 2, the seller uses the *Linear* concession strategy. Before learning, the negotiation ends at \$48.57 and after learning, the buyer can have a better agreement at \$22.34. In Scenario 3, the seller uses the *Conceder* concession strategy. Before learning, the final agreement is \$33.38 and after learning this value decreases to \$20.58. According to these experimental results from the three scenarios, we can conclude that, through learning of the opponent's historical offers, the agent employing our negotiation model can effectively adapt its concession strategy so as to increase its negotiation outcome. Our negotiation model is robust when the opponent employs different concession strategies.

In this section, we illustrate the experimental results of our negotiation model and compare the results with the NDF. The experimental results indicate that our negotiation model can dynamically adapt a negotiation agent's concession strategy and significantly increase a negotiation agent's utility through the learning of the opponent's historical offers.

5 Related Work

Although incorporating learning into agent negotiation is a relatively new research topic, many approaches, models and mechanisms have been developed in recent years to solve different issues in this topic [1] [3] [6] [7] [9] [10] [11] [12] [13] [16]. We discuss several related works and compare them with our work in this paper.

Bayesian learning technique has been widely applied in negotiation for a better negotiation outcome. Zeng and Sycara were probably the first to propose a Bayesian learning based negotiation model [16]. A sequential decision making model called *Bazaar* was introduced to model beliefs of the opponent's reservation point. Our model differs from their approach in two ways. (1) *Bazaar* can only learn the reserve price of the opponent while our model can learn both opponent's price and deadline, and (2) *Bazaar* requests priori knowledge about the potential distribution of of the opponent's reserve price while our model has no this request. In [11], the authors adopted the Markov chain framework to model bilateral single issue negotiations among agents in dynamic environments and use Bayesian learning to enable agents to learn an optimal strategy in incomplete information settings. However, as in [16], the approach still requires the learning agent to have a prior knowledge about the conditional probability when using the Bayesian updating rule. Furthermore, this approach is designed to model the strategies of the opponent, while in our approach the agent is learning by modeling the private negotiation parameters of the opponent.

Reinforcement learning [17] is another valid technique to be integrated into negotiation [1] [14]. Soo and Hung used Q-Learning algorithm in bilateral multi-issue negotiation [14]. However, in their work, the agent's reservation price was assumed as common knowledge. In [1], a bilateral price negotiation strategy based on Bayesian classification and Q-learning was proposed for a negotiation agent to make the best use of the opponent's negotiation history to make a decision of the opponent's classification based on Bayesian classification and then to create counter-offer efficiently by Q-learning. However, the approach in [1] is based on the classification of the opponent such that a learning agent can adjust its belief accordingly. This classification rule is set beforehand and is assumed as prior knowledge while our approach enables the agent to adapt its belief based on the learning results from the regression analysis, which are only determined by the historical offers.

Regression analysis was also employed by Ren and Zhang to predict the opponent's concession strategy by using the historical offers only [6] [7]. However, the approaches in [6] [7] did not give further advice on how to adapt agent self's

concession strategy based on the learning results. Our approach moves further by contrive an efficient adaptive concession strategy based on the learning results derived from the regression analysis and Bayesian learning, making the negotiation process totally dynamic and adaptive in the view of the learning agent.

Narayanan and Jennings proposed a novel adaptive negotiation model considering the dynamism in E-commerce settings [12]. Their model manages a negotiation process as a Markov Decision Process(MDP) and uses a value iteration algorithm to acquire optimal policies to adopt different concession strategies. However, their method can only determine the adaptive action to choose a concession strategy and cannot produce a precise concession value while our model can provide constructive guidance to the agent to dynamically adaptive its behaviors including both strategies and concession values. Bzostowski and Kowalczyk [10] presented an approach for modeling behaviors of negotiators and predictive decision-making. Both their approach and our work use the similar method in term of adaptive concession strategy based only on the historical offers. However, their approach focuses more on the analysis of the differences between adjacent offers from the opponents, and will become ineffective when these differences are not significant. Our approach employs the regression analysis and will not be affected by the variance of adjacent offers.

6 Conclusion and Future Work

In this paper, we proposed an adaptive bilateral negotiation model based on Bayesian learning. This model includes a learning mechanism and an adaptive concession strategy. Through Bayesian learning, an agent's belief about the opponent's reserve price can be revised dynamically during negotiation by comparing the fitted offers derived from a regression analysis with the opponent's historical offers. The agent then proposes a countermeasure based on an adaptive concession strategy. The proposed model can enable an agent to adapt its concession strategies dynamically according to the updated probability distribution in a predicting region, thus making the negotiation process dynamic and adaptive in the view of the learning agent. The experimental results demonstrate the good performance of our model by comparison with non-learning NDF model.

There are several direction for future research. Firstly, it is imperative to extend our model to multi-issue negotiation. Considering another important factor (i.e., weighting among the issues), our model can be further extended to make a possible win-win outcome for both agents. Secondly, we will take the non-linear utility function into account to broaden its application in practice. At last but not least, it is necessary to suit our model for more complex scenarios when the opponent is changing its deadline, reserve price, or even has a dynamic mixture of different strategies. These are all feasible and worthwhile aspects for further research. We leave them for future work.

References

1. Sun, T., Zhu, Q., Xia, Y., Cao, F.: A Bilateral Price Negotiation Strategy Based on Bayesian Classification and Q-learning. *Journal of Information & Computational Science* 13, 2773–2780 (2011)
2. Ng, S., Sulaiman, M., Selamat, M.: Intelligent negotiation agents in electronic commerce applications. *Journal of Artificial Intelligence* 2, 29–39 (2009)
3. Ng, S., Sulaiman, M., Selamat, M.: Machine learning approach in optimizing negotiation agents for E-commerce. *Information Technology Journal* 8, 801–810 (2009)
4. Aydogan, R., Yolum, P.: Ontology-based learning for negotiation. In: *Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, vol. 02, pp. 177–184. IEEE Computer Society (2008)
5. Hindriks, K., Tykhonov, D.: Opponent Modelling in Automated Multi-Issue Negotiation Using Bayesian Learning. In: *Proceedings of the 7th International Conference on Autonomous Agents and Multi Agent Systems, AAMAS 2008*, pp. 331–338. IFAAMAS, Richland (2008)
6. Ren, F., Zhang, M.-J.: Predicting Partners' Behaviors in Negotiation by Using Regression Analysis. In: Zhang, Z., Siekmann, J.H. (eds.) *KSEM 2007. LNCS (LNAI)*, vol. 4798, pp. 165–176. Springer, Heidelberg (2007)
7. Ren, F., Zhang, M.: Prediction of Partners' Behaviors in Agent Negotiation under Open and Dynamic Environments. In: *2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology Workshops*, pp. 379–382. IEEE (2007)
8. Aydogan, R., Yolum, P.: Learning consumer preferences using semantic similarity. In: *Proceedings of Sixth International Joint Conference on Autonomous Agents and Multi-agent Systems, AAMAS 2007*, pp. 1293–1300. ACM, New York (2007)
9. Ramchurn, S.D., Sierra, C., Godo, L., Jennings, N.R.: Negotiating using rewards. *Artificial Intelligence* 171, 805–837 (2007)
10. Bzostowski, J., Kowalczyk, R.: Predicting partner's behavior in agent negotiation. In: *Proceedings of 5th International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS 2006*, pp. 355–361. IFAAMAS, Richland (2006)
11. Narayanan, V., Jennings, N.R.: Learning to Negotiate Optimally in Non-stationary Environments. In: Klusch, M., Rovatsos, M., Payne, T.R. (eds.) *CIA 2006. LNCS (LNAI)*, vol. 4149, pp. 288–300. Springer, Heidelberg (2006)
12. Narayanan, V., Jennings, N.R.: An adaptive bilateral negotiation model for e-commerce settings. In: *Seventh IEEE International Conference on E-Commerce Technology*, pp. 34–41. IEEE (2005)
13. Li, J., Cao, Y.D.: Bayesian learning in bilateral multi-issue negotiation and its application in MAS-based electronic commerce. In: *Proceedings of the IEEE/WIC/ACM International Conference on Intelligent Agent Technology, IAT 2004*, pp. 437–440 (2004)
14. Soo, V.W., Hung, C.A.: On-line incremental learning in bilateral multi-issue negotiation. In: *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems: Part 1*, pp. 314–315. ACM, New York (2002)
15. Faratin, P., Sierra, C., Jennings, N.R.: Negotiation decision functions for autonomous agents. *Robotics and Autonomous Systems* 24, 159–182 (1998)
16. Zeng, D., Sycara, K.: Bayesian learning in negotiation. *International Journal of Human-Computers Studies* 48, 125–141 (1998)
17. Sutton, R.S., Barto, A.G.: *Reinforcement learning: An introduction*. MIT Press, Cambridge (1998)