Accepting Optimally in Automated Negotiation with Incomplete Information

Tim Baarslag
Interactive Intelligence Group
Delft University of Technology
Mekelweg 4, Delft, The Netherlands
T.Baarslag@tudelft.nl

Koen Hindriks
Interactive Intelligence Group
Delft University of Technology
Mekelweg 4, Delft, The Netherlands
K.V.Hindriks@tudelft.nl

ABSTRACT

When a negotiating agent is presented with an offer by its opponent, it is faced with a decision: it can accept the offer that is currently on the table, or it can reject it and continue the negotiation. Both options involve an inherent risk: continuing the negotiation carries the risk of forgoing a possibly optimal offer, whereas accepting runs the risk of missing out on an even better future offer. We approach the decision of whether to accept as a sequential decision problem, by modeling the bids received as a stochastic process. We argue that this is a natural choice in the context of a negotiation with incomplete information, where the future behavior of the opponent is uncertain. We determine the optimal acceptance policies for particular opponent classes and we present an approach to estimate the expected range of offers when the type of opponent is unknown. We apply our method against a wide range of opponents, and compare its performance with acceptance mechanisms of state-of-the-art negotiation strategies. The experiments show that the proposed approach is able to find the optimal time to accept, and improves upon widely used existing acceptance mechanisms.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*intelligent agents, multi-agent systems*

General Terms

Algorithms, Experimentation, Performance, Theory

Keywords

Negotiation; Acceptance strategy; Optimal stopping

1. INTRODUCTION

Suppose two parties A and B are conducting a negotiation, and B has just proposed an offer to A. A is now faced with a decision: she must decide whether to continue, or to accept the offer that is currently on the table. On the one hand, accepting the offer and ending the negotiation means running the risk of missing out on a better deal in the future.

On the other hand, carrying on with the negotiation involves a risk as well, as this gives up the possibility of accepting one of the previous offers. How then, should A decide whether to end or to continue the negotiation?

Of course, A's decision making process will depend on the current offer, as well as the offers that A can expect to receive from B in the future. However, in most realistic cases, agents have only incomplete information about each other [5, 9, 15]. In this paper, we explore in particular the setting where the opponent has only limited or no knowledge of A's preferences, and the proposals that A receives will therefore be necessarily uncertain. This makes A's task of predicting B's future offers by no means an easy one.

Moreover, predicting B's future offers is only part of the solution: even when A can predict B's moves reasonably well, A still has to decide how to put this information to good use. In other words, even when a probability distribution over the opponent's actions is known, it is not straightforward to translate this into effective negotiation behavior. As an extreme example, consider an opponent R (for Random) who will make random offers with utility uniformly distributed in $[0, 1]$. Suppose furthermore that we can expect to receive two more bids from R until the deadline occurs. R currently makes an offer of utility $x \in [0, 1]$; for what x should we accept? Of course, an even better bid than x might come up in one of the two remaining rounds; on the other hand, it might be safer to settle for this bid if x is large enough. For this particular case, we will prove that there is an optimal acceptance strategy, and we show exactly for what x to accept (see Section 3).

The main contribution of this paper is that we address both of A's problems: first, at every stage of the negotiation, we provide a technique to estimate the bidding behavior of various opponent classes by modeling A's dilemma as a stochastic decision problem. For particular opponent classes we are able to provide precise models, and to formulate exact mathematical solutions to our problem. For the second step, using the ranges found earlier, we borrow techniques from optimal stopping theory to find generic, optimal rules for when to accept against a variety of opponents in a bilateral negotiation setting with incomplete information. The solutions proposed are optimal in the sense that there can be no better strategy in terms of utility.

We begin by introducing our approach in Section 2, and we apply our methods to find optimal rules in the specific case of opponents that bid randomly in Section 3. We then build upon these cases and subsequently work out more realistic scenarios in the following sections. In Section 4, we

explore opponents that change their behavior over time, and we determine optimal stopping rules when good estimates of their bidding behavior are known. We extend these results by combining our approach with a state-of-the-art prediction mechanism, and we demonstrate that our approach outperforms existing accepting mechanisms, even when the opponent's behavior is unknown.

2. DECISION MAKING IN NEGOTIATION UNDER UNCERTAINTY

In this work we focus on a bilateral automated negotiation, wherein two agents try to reach an agreement while maximizing their own utility. Agents use the widely-employed *alternating-offers protocol* for bilateral negotiations [21], in which the negotiating parties take turns in exchanging offers for a fixed number of rounds N. In case the deadline is reached before both parties come to an agreement, both receive zero utility. A negotiation scenario consists of the negotiation domain, which specifies the setting and all possible bids, together with a privately-known preference profile for each party. A preference profile is described by a utility function $u(x)$, which maps each possible outcome x in the negotiation domain to a utility in the range $[0, 1]$. If an agent receives an offer x, its acceptance mechanism has to decide whether $u(x)$ is high enough to be acceptable.

2.1 Stochastic Behavior in Negotiation

Suppose player B is involved in a bilateral negotiation with private preference information, and at some point in time, he has decided that he is satisfied with a utility around u, called his *target utility*. However, there are many possible bids with approximately this utility for player B. As usual, we call this set of bids X the *iso-level* bids with utility u. As player B is indifferent between these bids, B may attempt to optimize player A's utility in order to maximize the chance of an agreement. But this is difficult for B to achieve, as he does not know A's preferences. When using the alternating offers protocol, player B cannot simply send out all considered offers as one bundle, but instead, he can only offer them sequentially. Player B typically continues to select different bids from X until his target utility changes; then a new set of iso-level bids is generated, and the process starts again. The order in which player B picks bids from the set of equally preferred bids X will differ per player, but due to incomplete information, he can only select a bid with a particular opponent utility with limited certainty. Therefore, we can reasonably model the offers that are presented to A as a *stochastic process*.

This kind of stochastic behavior can be observed in practice in the Automated Negotiating Agents Competition [1, 2]. ANAC is a yearly international competition in which negotiation agents compete in an incomplete information setting. Half of the participants [1, 3, 7, 13, 24] were not designed to explicitly optimize opponent utility and therefore, with the limited information available, simply selected a random element from X; others used opponent modeling techniques that estimate the opponent's preferences in order to select bids closer to the Pareto optimal frontier. However, opponent modeling is seldom capable of making perfect estimates [20]. Consequently, even when player B employs an opponent modeling technique, A will still receive bids of varying utility. Moreover, the agents usually already antici-

pated the limitations of their opponent model, and therefore randomly chose among the estimated top bids for the opponent [11, 16], adding even further to the random appearance of the utility of their bids. As a result, the negotiation traces of ANAC 2011 showed to a very large extent the stochastic behavior discussed above (see also Figure 1). Only 25% of the negotiation moves were an improvement for the opponent over the previous bid; the other 75% of the moves could be classified as selfish, unfortunate, or silent [4].
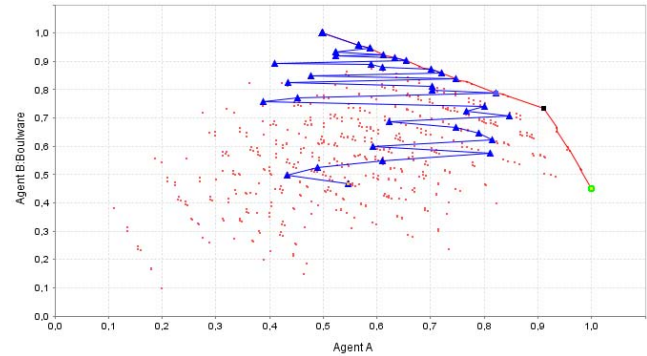

Figure 1: Despite the fact that Side B concedes predictably over time, the utility of the offers seem to be randomly distributed around the $[0.4, 0.8]$ interval for Side A, and as a result, the best bids for A occur during the middle of the negotiation.

2.2 Optimal Stopping in Negotiation

We can frame the problem of accepting a bid as an optimal stopping problem [6], in which an agent is faced with the dilemma of choosing when to take a particular action, in order to maximize an expected reward or minimize an expected cost. In such problems, observations are taken sequentially, and at each stage, one either chooses to stop to collect, or to continue and take the next observation (usually at some specified sampling cost).

The model of bid reception is as follows: at each of a total of N rounds, we receive a bid, which has an associated utility, or value, drawn from a random variable over the unit interval. At this point, we must decide whether to accept the bid, or not. Once we accept, the deal is settled and the negotiation ends. If we continue, then there is no possibility of recalling passed-up offers; i.e., previous offers are unavailable unless they are presented to us again. Hence, at each round, we must decide to either continue or to stop participating in the negotiation, and we wish to act so as to maximize the expected net gain. Once an offer is turned down, and we decide to wait for another bid (at a cost C), the total number of remaining observations decreases by one. We will develop the theory here for arbitrary sampling cost C, but in the remainder of the paper, we will assume the cost to be zero.

At every stage, the current situation may be described by a *state* (j, x), which is characterized by two parameters: the number of remaining observations $j \in \mathbb{N}$, and the latest received offer $x \in [0, 1]$. Let the utility distribution with j rounds remaining be given by a random variable X_j, with associated distribution function F_j. We can think of X_j as the possible utilities we receive when the opponent makes

iso-level bids, and $F_j(u)$ represents the probability of receiving a bid with utility less than or equal to u. The expected payoff is then given by

$$V(j, x) = \max(x, E(V(j-1, X_{j-1})) - C), \qquad (1)$$

where we abbreviate the second term $E(V(j-1, X_{j-1})) - C$ as v_j. This represents the expected value of rejecting the offer at (j, x), and going on for (at least) one more period. Note that v_j does not depend on x. Thus, using the substitution, we get

$$v_j = E(\max(X_{j-1}, E(V(j-2, X_{j-2})) - C) - C,$$

which leads to the following recurrence relation:

$$\begin{cases} v_0 = 0, \\ v_j = E(\max(X_{j-1}, v_{j-1})) - C. \end{cases} \qquad (2)$$

In [6, 17] it is proven that for any $s \in \mathbb{R}$, and for any random variable X with distribution function F for which $E(X)$ is finite, the following holds:

$$E(\max(X, s)) = s + T_F(s),$$

with

$$T_F(s) = \int_s^\infty (1 - F(t))\, \mathrm{d}t.$$

And therefore the recurrence relation describing v_j can be written as

$$v_j = v_{j-1} + \int_{v_{j-1}}^\infty (1 - F_{j-1}(t))\, \mathrm{d}t - C. \qquad (3)$$

Thus, if we know the distribution F_j for every j, we can compute the values v_j using the above recurrence relation. Then, deciding whether to accept an offer x is simple: if $x \geq v_j$ we accept, otherwise we reject the offer (see Algorithm 1).

There is however, a serious impediment to using our stochastic decision model in practice: we do not *know* the distributions of the utility that the opponent will present to us in the upcoming rounds; furthermore, the distributions are highly influenced by the specifics of the negotiation scenario.

However, against specific classes of opponents, we are able to establish these probabilities, and in an exact way. We will first focus our attention on theoretical cases that resemble the relevant cases encountered in practice. In order to develop the theory, we will first take on the extreme case of random opponent behavior, and gradually add complexity as we proceed.

Of course, in a general setting we do not know the opponent's behavior, and in that case we require a method to determine the distributions X_j for every remaining round j. This means that for every round, we need to estimate the probability of receiving certain utility in *our* utility space. This is the most difficult case, which we will cover at the end of the paper.

All in all, we consider three different opponent classes:

1. Random behavior: *fixed* and *known* uniform X_j in every round; this is solved mathematically in Section 3.

2. Known time-dependent behavior: *changing*, but *known* uniform X_j; this is optimally solved in Section 4.1.

3. Unknown time-dependent behavior: *changing* and *unknown* arbitrary X_j; covered in Section 4.2.

We will start with the first case, where we consider random opponent behavior.

3. ACCEPTING RANDOM OFFERS

Suppose an agent A is negotiating with its opponent, and the deadline is approaching, so both agents have only a few more offers to exchange. As argued above, the opponent will often offer bids with varying utility for A, due to its incomplete information of what A exactly wants. This means that from A's point of view, the utility of the presented offers will have a particular stochastic distribution. The aim of A is then to pick the best one given the limited time that is left.

We start by studying the extreme case of a maximally unpredictable opponent, or *Random Walker* (also known as the *Zero Intelligence* strategy [12]), who makes random bids at every stage of the negotiation. We first solve this case analytically, before moving on to more complicated settings.

3.1 Uniformly Random Behavior

Using equation (2), we can determine the optimal solution against *Random Walker*, using the added knowledge that every X_j does not depend on the number of rounds left, and assuming every X_j is uniformly distributed:

PROPOSITION 3.1. *Against an opponent who makes random bids of utility uniformly distributed in $[0, 1]$, and with j offers still to be observed, one should accept an offer of utility x exactly when $x \geq v_j$, where v_j satisfies the following equation:*

$$\begin{cases} v_0 = 0, \\ v_j = \frac{1}{2} + \frac{1}{2} v_{j-1}^2, \end{cases} \qquad (4)$$

This recurrence relation has the following properties: v_j is monotonically increasing, and

$$\lim_{j \to \infty} v_j = 1.$$

PROOF. Let X be the uniform distribution over $[0, 1]$ with distribution function F. Playing against *Random Walker*, all X_j's are uniform distributions over $[0, 1]$ and hence equal to X. This yields:

$$T_F(s) = \int_s^\infty (1 - F(t))\, \mathrm{d}t \qquad (5)$$

$$= \begin{cases} \frac{1}{2} - s, & s < 0. \\ \frac{1}{2}(1 - s)^2, & 0 \leq s \leq 1. \\ 0, & s > 1. \end{cases} \qquad (6)$$

Since we are in the case $0 \leq s \leq 1$, we get:

$$v_j = v_{j-1} + \frac{1}{2}(1 - v_{j-1})^2 = \frac{1}{2} + \frac{1}{2} v_{j-1}^2.$$

It is easy to show with recursion that

$$1 - \frac{2}{j} < v_{j-1} < v_j < 1,$$

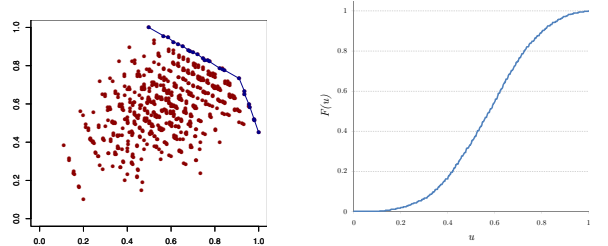and therefore,

$$\lim_{j \to \infty} v_j = 1. \quad \square$$

When we substitute $v_j = 1 - 2x_j$ in equation (4), we get the equivalent relation of the logistic map $x_j = x_{j-1}(1 - x_{j-1})$ at $r = 1$, which due to its chaotic behavior does not in general have an analytical solution. However, we have visualized its behavior for $j \in [0, 200]$ in Figure 3 (uniform case). From this, we see that the answer to the question posed in the introduction is as follows: with two rounds to go, one should accept an offer x exactly when $x \geq v_2 = 0.625$.

Note that $\lim_{j \to \infty} v_j = 1$ means we can expect to receive utility arbitrarily close to the maximum given enough time, and that this limit $v = 1$ is also the fixpoint of recurrence relation (4).

3.2 Non-Uniform Random Behavior

In Proposition 3.1, we consider random behavior by the opponent in a uniform way; i.e., a scenario where every received utility is equally likely. However, in practice such situations rarely occur. Negotiation scenarios usually enable agents to make trade-offs between multiple issues, resulting in clustering of potential outcomes. Hence, in a typical scenario, even when the opponent chooses bids randomly, the utilities of those bids are *not* distributed uniformly.

A typical example of such a multi-issue negotiation scenario is depicted in Figure 2a, and involves a case where England and Zimbabwe negotiate an agreement on tobacco control [2, 19]. The leaders of both countries have contradictory preferences for two issues, but three other issues have options that are jointly preferred by both sides. We will use it as a running example, but the outlined technique can be applied to any negotiation scenario.

(a) The outcome space of the England–Zimbabwe negotiation scenario.

(b) The cumulative distribution function $F(x)$.

Figure 2: The outcome space for player A and B, and the resulting cumulative distribution function for player A.

For such an outcome space, we cannot simplify equation (5) any further, so instead, we need to integrate the cumulative distribution function $F(x)$ directly (see Figure 2b). Note that $F(x)$ can be computed by the agent, simply by considering the distribution of the utilities of all possible outcomes. Using equation (3), we can now compute the values v_j for a scenario such as England–Zimbabwe; see Figure 3.

Note that the value of v_j for England-Zimbabwe increases faster than in the uniform case, but at the same time it also tends to 1 more slowly. This can be explained by the fact that this outcome space is more sparse in both extremes: since there are less bids of very low utility, it should aim higher at the end of the negotiation, and as there are also less bids of high utility, it should be satisfied more easily at the start of the negotiation.

3.3 Experiments

In order to test the efficacy of the optimal stopping condition, we first integrated it into a functional negotiating agent. This requires care, as normally, the behavior (and thus the performance) of a negotiating agent is determined by many factors outside of the acceptance mechanism, par-

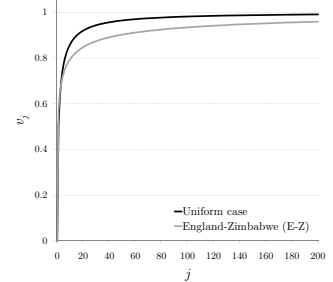| j	Unif. v_j	E-Z v_j
0	0	0
1	0.5	0.5734
2	0.625	0.6449
3	0.6953	0.6855
4	0.7417	0.7134
5	0.7751	0.7344
10	0.8611	0.7953
100	0.9812	0.9338
200	0.9903	0.9586

Figure 3: The optimal stopping values v_j for different rounds j versus uniformly random (*Uniform case*) and non-uniformly random (*E-Z*) behavior.

ticularly its bidding strategy. Note however, that against *Random Walker*, bidding strategies with the same acceptance policy perform equally, as it does not matter which offers are sent out. This holds because of three properties:

1. *Random Walker*'s offers do not depend on the opponent's behavior; hence, it is not sensitive to the other's bidding strategy;

2. *Random Walker* does not accept any offers; in our experiments, opponents are not allowed to accept, as this could prematurely end the negotiation, without revealing anything about the performance of the acceptance strategies.

3. The optimal stopping condition works independently of bids that are sent out.

Taking this into account in our experiments, we opted for an accompanying bidding strategy that is as simple as possible, namely *Hardliner* (also known as 'sit and wait agent'). This strategy simply makes a bid of maximum utility for itself and never concedes. Clearly, in a real negotiation setting, this is not a viable bidding tactic as it generally negatively influences the opponent's behavior, but this is of no concern against a non-behavior-based opponent.

We then compared its performance with the strategies of other state-of-the-art agents currently available for our setting. We selected all agents from the ANAC 2010 and 2011 editions. We also included the time dependent tactics (TDT's) such as *Hardliner* (with concession factor $e = 0$), *Boulware* ($e = \frac{1}{2}$), *Conceder Linear* ($e = 1$), and *Conceder* ($e = 2$) taken from [8]. To analyze the performance of different agents, we employed GENIUS [18], which is an environment that facilitates the design and evaluation of automated negotiators' strategies.

For our negotiation scenarios, we opted for the England-Zimbabwe domain, and a discretized version of Split the Pie [21, 23], where two players have to reach an agreement on the partition of a pie of size 1. The pie will be partitioned only after the players reach an agreement, in which case one gets $x \in [0, 1]$, and the other gets $1 - x$. In this scenario, *Random Walker* makes bids of utility uniformly distributed in $[0, 1]$, since it proposes random partitions of the pie.

The results of our experiment in the uniform case and on England-Zimbabwe are plotted in Figure 4 and Figure 5 respectively, both for $N = 10$ and $N = 100$ negotiation rounds. The optimal stopping condition significantly outperforms all agents (*one-tailed t-test*, $p < 0.01$) in all cases.

Algorithm 1: Optimal Stopping main decision body

Input: The number of remaining rounds j, and the last received bid x by the opponent.
Output: Acceptance or rejection of x.
begin
    offeredUtility $\longleftarrow$ getUtility(x);
    target $\longleftarrow$ determine(v_j);
    if offeredUtility $\geq$ target **then**
        $\llcorner$ **return** ACCEPT
    else
        $\llcorner$ **return** REJECT // And send a counter-offer

In the uniform case (cf. Figure 4), it obtains the highest score possible both in 10 and in 100 rounds, getting respectively 86% and 98% of the pie on average. Note that this is exactly equal to the theoretical values v_{10} and v_{100} shown in the uniform case of Figure 3.
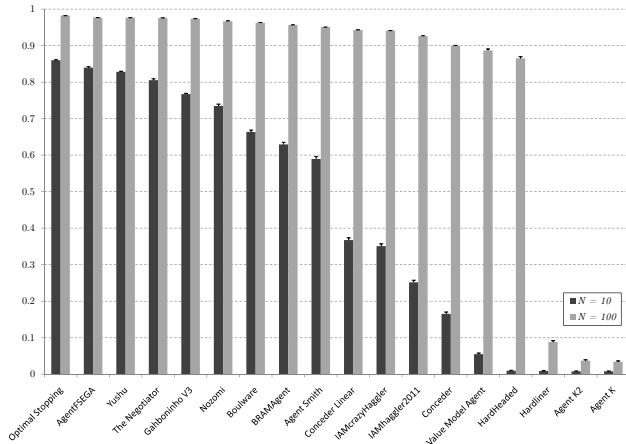
Figure 4: The optimal stopping condition outperforms all ANAC agents against uniform *Random Walker* for 10 and 100 rounds. Average utility plotted over 5000 runs; the errors bars indicate one standard deviation to the mean.

On England-Zimbabwe (cf. Figure 5), the optimal stopper obtains less utility for $N = 10$ in an absolute sense compared to the uniform case, but the results are even more pronounced relative to the other agents. A moment of reflection makes clear why: given the clustering of bids of medium utility (see Figure 2a), there is less chance for *Random Walker* to propose a very fortunate bid for the opponent. This explains why the acceptance criteria of the other agents perform relatively worse. Note that the end result obtained by the optimal stopper is approximately 0.79 and 0.93 for $N = 10$ and $N = 100$ respectively, which is again equal to the optimal values v_{10} and v_{100} shown in the E-Z case of Figure 3.

3.4 When Optimal Stopping Is Most Effective

As is evident from the results, optimal stopping performs better than the other agents against *Random Walker*. This is to be expected given the fact that no agent could possibly do better; however, the difference with the current state-of-the-art is surprisingly big in some cases, for example compared to the ANAC 2010 winner, *Agent K* [13], and even
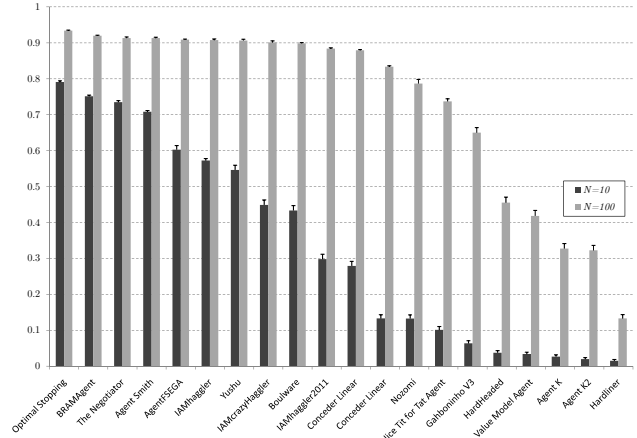
Figure 5: The optimal stopping condition outperforms all ANAC agents against *Random Walker* on England-Zimbabwe for $N = 10$ and $N = 100$ rounds; utility averaged over 5000 runs.

more so when the number of rounds is limited. The reason is that many of the currently used acceptance mechanisms are rather straightforward, and only become successful against *Random Walker* when enough time is available.

This can be illustrated by considering the baseline acceptance condition where an agent accepts if and only if the offered utility is above a fixed threshold α, as is done by various agents [13, 24, 25]. In the uniform case, its expected utility obtained over N rounds equals the probability it will obtain an offer above α multiplied by the expected utility above α:

$$(1 - \alpha^N) \cdot \frac{\alpha + 1}{2} \qquad (7)$$

This is not a very efficient acceptance condition for small N; for example, for $N = 2$, the optimal value of α (i.e., the value of α that maximizes formula (7)) is $\frac{1}{3}$, with expected utility of 0.593, while the optimal value that can be obtained is $v_2 = 0.625$. However, for large N, choosing α close to one is already surprisingly efficient. For example, for $N = 100$, the optimal value of α is 0.948, with expected utility of 0.969. This is already quite close to the optimal value of $v_{100} = 0.981$; this indicates that in case bids are randomly distributed, the added value of optimal stopping lies primarily in negotiations with a limited amount of total rounds, or when only limited time is left in a negotiation.

4. TIME DEPENDENT OFFERS

One of the most restrictive assumptions so far was to assume the opponent plays completely randomly. As we argued, this is a sensible assumption when modeling an opponent that is extremely unpredictable due to imperfect information, but the general case is more complicated. Almost all negotiation agents change their range of offers over time; i.e., they are *time dependent strategies*.[1] Hence, we require optimal stopping policies for these cases as well.

The challenge of the more general case is that we have to account for the fact that not only the presented utilities

---

[1]Note that these should not be confused with the well-known time dependent *tactics* from [8], which are particular kinds of time dependent strategies.

may fluctuate, but also the *range* of future offers may be different at different times. Establishing this range is not easy, because the strategy used by the opponent is of course unknown to us. The offers of any time dependent opponent with incomplete information can again be modeled by a stochastic distribution, but this time the distribution will change over time. In terms of optimal stopping, this means that the bid distribution X_j can be different for every j.

4.1 Uniformly Unpredictable Offers

If we assume that the opponent's offers are uniformly distributed, we only need to know the *interval* of utilities we can expect in every round. If this is the case, then we are able to compute the optimal time to accept, as is stated in the following proposition.

PROPOSITION 4.1. *Against a time-dependent opponent who, with j rounds still to be observed, makes bids uniformly distributed in $X_j = [a_j, b_j]$, the optimal stopping cut-off is v_j, where v_j satisfies the following equation:*

$$v_j = \begin{cases} 0, & \text{if } j = 0 \\ v_{j-1}, & \text{if } v_{j-1} \geq b_{j-1} \\ \frac{b_{j-1} + a_{j-1}}{2}, & \text{if } v_{j-1} \leq a_{j-1} \\ v_{j-1} + \frac{1}{2} \cdot \frac{(b_{j-1} - v_{j-1})^2}{b_{j-1} - a_{j-1}}, & \text{if } a_{j-1} < v_{j-1} < b_{j-1}. \end{cases}$$

PROOF. From equation (2), we have

$$v_j = E(\max(X_{j-1}, v_{j-1})),$$

so immediately, if $v_{j-1} \geq b_{j-1}$, then $v_{j-1} \geq X_{j-1}$, and thus $v_j = v_{j-1}$. On the other hand, if $v_{j-1} \leq a_{j-1}$, then $v_j = E(X) = \frac{b_{j-1} + a_{j-1}}{2}$. So therefore, the only case left is $a_{j-1} < v_{j-1} < b_{j-1}$, in which case we derive the following:

$$\begin{aligned} v_j &= P(X_{j-1} \leq v_{j-1}) \cdot v_{j-1} + \\ & \quad P(X_{j-1} > v_{j-1}) \cdot P(X_{j-1} \mid X_{j-1} > v_{j-1}) \\ &= \frac{v_{j-1} - a_{j-1}}{b_{j-1} - a_{j-1}} \cdot v_{j-1} + \frac{1}{2} \cdot \frac{b_{j-1} - v_{j-1}}{b_{j-1} - a_{j-1}} \cdot (v_{j-1} + b_{j-1}) \\ &= v_{j-1} + \frac{1}{2} \cdot \frac{(b_{j-1} - v_{j-1})^2}{b_{j-1} - a_{j-1}}. \quad \square \end{aligned}$$

Note how this proposition is an extension of Proposition 3.1: if we set $X_j = [0, 1]$ for every j, the equation simplifies to $v_j = \frac{1}{2} + \frac{1}{2} v_{j-1}^2$ again.

Also, we observe that in the special case of perfect information, the distributions would be singletons of the form $X_j = \{x_j\}$, with probability 1 for the outcome x_j. The equation of Proposition 4.1 then simplifies to

$$\begin{aligned} v_j &= \begin{cases} 0, & \text{if } j = 0 \\ \max(x_{j-1}, v_{j-1}), & \text{otherwise.} \end{cases} \\ &= \max_{0 \leq k < j} x_k. \end{aligned}$$

This means that the optimal stopping procedure has the desirable property that when it gets perfect estimates as input, it will also produce perfect output.

4.2 Arbitrarily Unpredictable Offers

Proposition 4.1 is useful to gain insight into the optimal acceptance policy, but in practice, the distributions X_j are neither known, nor uniformly distributed, and therefore an estimation method is required against arbitrary opponents. Of course, the success of the optimal stopping rules

will greatly depend on the fidelity of the estimating technique used to predict the opponent's behavior. Therefore, we first examine the case of a perfect estimator, to see how our method performs in the ideal case. After that, we will move our focus to an estimator that can be used in practice.

4.2.1 Opponent Prediction Using Perfect Estimates

The perfect estimation method that we employ divides the number of rounds N into a number of time slots S. Then, by momentarily using perfect information, it gets the minimum and maximum utility that will be offered by the opponent during that time slot. This allows us to control exactly the precision of the estimate, where using more slots emulates having more information about the opponent's behavior. If we set the number of slots equal to the total number of rounds, we are in a full information state and the performance should be theoretically optimal. If we use only one slot, we have less information, knowing only the opponent's utility range over the entire N rounds.

4.2.2 Opponent Prediction Using Gaussian Process Regression

Finally, we consider an estimation method that uses as input only the information that can be observed during the negotiation, namely the utility of the offers made by the opponent. For this, we opted for a Gaussian process regression (GPR) technique as described in [24]. We selected the GPR technique because it can be computed in real time during the negotiation, and it is specifically designed to be robust with respect to significantly varying observations. It works as follows: for each offer made by the opponent, the round at which the offer was made is recorded, along with the offered utility. From this, the future concessions of the opponent are estimated using regression with a Gaussian process. To reduce the effect of noise, the offers received are aggregated in a number of time windows, and only the maximum value that is received in each time window is used as input for the Gaussian process.

The output of the Gaussian process regression is a normal distribution for every upcoming round k, with mean μ_k and standard deviation σ_k. The mean μ_k gives a prediction of the most likely offered utility value in round k, whilst the standard deviation σ_k gives an indication of how accurate the prediction is. When using GPR, the opponent bid distribution is estimated in real-time by a *normal* distribution, truncated to fit in the range $[0, 1]$.

Algorithm 2: Determining v_j

Input: The number of remaining rounds j, and all negotiation outcomes Ω.
Output: v_j.
begin
  if $j = 0$ **then**
    └ **return** 0
  // Use either perfect estimation, or GPR
  Y_{j-1} ⟵ estimated utility distribution at $j-1$;
  // Use either uniform, or Gaussian distribution for X_{j-1}
  X_{j-1} ⟵ utility distribution of Y_{j-1} over Ω;
  // Recursively determine v_{j-1}
  return $E(\max(X_{j-1}, v_{j-1}))$

4.3 Experiments

To analyze the performance of optimal stopping (OS) against time dependent negotiation strategies, we adopted the same experimental setup as before, this time testing it with both perfect estimates and the GPR technique. We set the number of Gaussian process regressions to 10, and we set the number of samples equal to the number of rounds (for details, see [24]).

We used two versions of perfect estimation: the full information state by setting $S = N$ (called "perfect estimation with full slots"), and a version with $S = 1$ (called "perfect estimation with one slot"). We also tested two variants of the GPR technique: in one, we simply set X_j equal to the truncated Gaussian distribution with mean μ_j and standard deviation σ_j as predicted by the GPR technique (called "Gaussian GPR prediction"). These predictions turned out to be overly optimistic in most cases, since the GPR technique uses as input the maximum utility received in each time slot. Therefore, we opted to include a simplified second version, which produces a uniform distribution between zero and the estimated maximum offered utility, which we set to $\mu + 2\sigma$ (called "uniform GPR prediction"). See also Algorithm 2.

As the specifics of the negotiation scenario influences the behavior of the opponent, we picked a total of six negotiation scenarios from ANAC, aiming for a large spread of negotiation characteristics (see Table 1).

Scenario	Size	Opposition
ADG [1]	15625	Low
Amsterdam [1]	3024	Medium
England–Zimbabwe [19]	527	Medium
Nice Or Die [3]	3	High
Itex–Cypress [14]	180	High
Travel [2]	188160	Medium

Table 1: Characteristics of the negotiation scenarios

For the opponents, we selected various TDT's from [8]. Our optimal stopping policy works against *any* type of time dependent negotiation strategy, but we selected TDT's because they are typical, well-known examples of strategies that change their range of bids over time. Additionally, as in the case of *Random Walker*, TDT's are non-adaptive and hence it is not important what counter-offers are sent out to them.

We selected the same TDT's used earlier, namely: *Hardliner*, *Boulware*, *Conceder Linear*, and *Conceder*. We generated many variants of each opponent by choosing the values 0.7, 0.8, and 0.9 for the *min* parameter (which controls the utility threshold up to where the agent will concede [8]). Note that this creates quite a competitive opponent pool, as the opponents will never fully concede. This leads to only small utility differences between the different acceptance strategies, but this should be regarded an artifact of our competitive setup. The average score of all agents is shown in Figure 6.

Optimal stopping with perfect estimation with full slots should be considered the theoretical upper bound here; and indeed, it outperforms all other methods. Among the agents that act with *incomplete* information, the *Boulware* agent obtains a surprisingly good score. Its strategy turns out to be particularly successful against TDT's since it waits for a long time to let the opponent concede as much as possible, until it quickly concedes in the end to obtain an agreement.
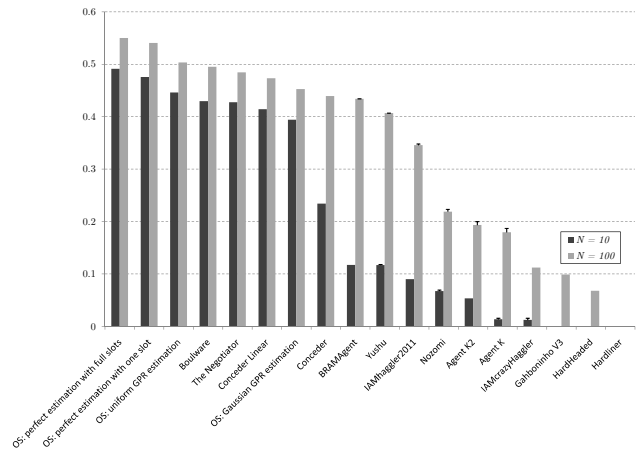

Figure 6: The utilities obtained by ANAC agents and optimal stopping conditions with different estimation methods against time dependent opponents for 10 and 100 rounds.

However, because it waits for so long, it misses out on good offers that are offered earlier.

Gaussian GPR prediction is not as successful, mainly because it was found to overestimate the opponent's willingness to concede, and hence it aimed for too much during the negotiation. It is optimal stopping with uniform GPR that performs significantly best (*one-tailed t-test*, $p < 0.01$), which shows that the optimal stopping policy is indeed a robust mechanism that can still perform well in an incomplete information setting.

5. RELATED WORK

As far as the authors are aware, this is the first work that deals with the optimal decision on the acceptance of an offer in a negotiation setting of incomplete information. In many settings of *complete* information ([21] is a typical example) the deal is usually formed right away and as such, sequential decisions whether to accept do not come into play.

In [26], a sequential decision making framework is also employed, using similar arguments for using it as we do. Furthermore, they also choose actions that maximize the expected payoff using a recursive formula; however, their approach uses Bayesian learning techniques and does not provide solutions specifically aimed at acceptance strategies.

The work by Fatima et al. [10] also treats optimal strategies in an incomplete information setting, but it primarily focuses on bidding strategies in the context of unknown deadlines and reservation values, and does not deal with acceptance strategies.

A work that comes closest to ours is [22], where optimal stopping is employed to decide when a party should reach an agreement in the context of conflict resolution. In contrast to our work, the scope of the paper is limited to simple bargaining games, and deals with one-sided incomplete information only.

6. DISCUSSION AND FUTURE WORK

This paper deals with the question of when to accept in a bilateral negotiation with incomplete information. Our approach has been to model the opponent's bids as a stochastic

process, and to regard the decision of when to accept as a sequential decision problem. We first determined the optimal acceptance policies for particular opponent classes of which we were able to predict the behavior well. Of course, in the general case of unknown opponents, the solutions are only as good as the estimation of the opponent's behavior. We have shown however, that our techniques are robust, in the sense that they also perform well in practice. This demonstrates that our optimal stopping mechanism is a valuable element of a negotiating agent's strategy, whether in a complete or incomplete information setting.

Our work also opens up various lines of possible future work. First, there is an important aspect of negotiation that is not included in our model: negotiation is a dynamic two-party process. The opponent's behavior is influenced by whether or not we accept, and by what counter offers we make. In this paper we focused on the decision of when to accept, rather than on what counter offer should be generated in case the offer is unacceptable. Clearly, in practice, the bidding strategy is also very important in conducting a successful negotiation, and its effects on the acceptance strategy are yet to be determined.

Second, our model already incorporates the concept of negotiation costs, but in this paper we assumed them to be zero; however, it would be interesting to see the effects of costs on optimal acceptance behavior. Similarly, we also plan to study negotiation scenarios that have discounted payoffs. Both extensions will incentivize agents to employ more permissive acceptance conditions. On the other hand, adding reservation values to the agent's preferences would make an agent *less* inclined to accept. Combining this with discounted scenarios could induce agents to fall back on their reservation value by *ending* the negotiation prematurely. This 'outside option' gives rise to a new variety of optimal acceptance strategies that have to make the optimal choice between continuing, accepting, or walking away.

7. REFERENCES

[1] Tim Baarslag, Katsuhide Fujita, Enrico H. Gerding, Koen Hindriks, Takayuki Ito, Nicholas R. Jennings, Catholijn Jonker, Sarit Kraus, Raz Lin, Valentin Robu, and Colin R. Williams. Evaluating practical negotiating agents: Results and analysis of the 2011 international competition. *Artificial Intelligence Journal*, 2012.

[2] Tim Baarslag, Koen Hindriks, Catholijn M. Jonker, Sarit Kraus, and Raz Lin. The first automated negotiating agents competition (ANAC 2010). In *New Trends in Agent-based Complex Automated Negotiations*, pages 113–135, Berlin, Heidelberg, 2012. Springer-Verlag.

[3] Mai Ben Adar, Nadav Sofy, and Avshalom Elimelech. Gahboninho: Strategy for balancing pressure and compromise in automated negotiation. In *Complex Automated Negotiations: Theories, Models, and Software Competitions*, pages 205–208. Springer Berlin Heidelberg, 2013.

[4] Tibor Bosse and Catholijn M. Jonker. Human vs. computer behaviour in multi-issue negotiation. In *Proceedings of the Rational, Robust, and Secure Negotiation Mechanisms in Multi-Agent Systems*, RRS '05, pages 11–, Washington, DC, USA, 2005. IEEE Computer Society.

[5] Robert M. Coehoorn and Nicholas R. Jennings. Learning on opponent's preferences to make effective multi-issue negotiation trade-offs. In *Proceedings of the 6th international conference on Electronic commerce*, ICEC '04, pages 59–68, New York, USA, 2004. ACM.

[6] Morris H. DeGroot. *Optimal statistical decisions*. McGraw-Hill, New York, 1970.

[7] A.S.Y. Dirkzwager, M.J.C. Hendrikx, and J.R. Ruiter. The negotiator: A dynamic strategy for bilateral negotiations with time-based discounts. In *Complex Automated Negotiations: Theories, Models, and Software Competitions*, pages 217–221. Springer Berlin Heidelberg, 2013.

[8] P. Faratin, C. Sierra, and N. R. Jennings. Negotiation decision functions for autonomous agents. *Int. Journal of Robotics and Autonomous Systems*, 24(3-4):159–182, 1998.

[9] S. S. Fatima, M. Wooldridge, and N. R. Jennings. An agenda-based framework for multi-issue negotiation. *Artificial Intelligence*, 152(1):1–45, 2004.

[10] S. Shaheen Fatima, Michael Wooldridge, and Nicholas R. Jennings. Optimal negotiation strategies for agents with incomplete information. In *Revised Papers from the 8th International Workshop on Intelligent Agents VIII*, ATAL '01, pages 377–392, London, UK, UK, 2002. Springer-Verlag.

[11] Asaf Frieder and Gal Miller. Value model agent: A novel preference profiler for negotiation with agents. In *Complex Automated Negotiations: Theories, Models, and Software Competitions*, pages 199–203. Springer Berlin Heidelberg, 2013.

[12] Dhananjay K. Gode and Shyam Sunder. Allocative efficiency in markets with zero intelligence (zi) traders: Market as a partial substitute for individual rationality. *Journal of Political Economy*, 101(1):119–137, 1993.

[13] Shogo Kawaguchi, Katsuhide Fujita, and Takayuki Ito. Agentk2: Compromising strategy based on estimated maximum utility for automated negotiating agents. In *Complex Automated Negotiations: Theories, Models, and Software Competitions*, pages 235–241. Springer Berlin Heidelberg, 2013.

[14] Gregory E. Kersten and Grant Zhang. Mining inspire data for the determinants of successful internet negotiations. *InterNeg Research Papers INR 04/01 Central European Journal of Operational Research*, 2003.

[15] Sarit Kraus, Jonathan Wilkenfeld, and Gilad Zlotkin. Multiagent negotiation under time constraints. *Artificial Intelligence*, 75(2):297 – 345, 1995.

[16] Thijs Krimpen, Daphne Looije, and Siamak Hajizadeh. Hardheaded. In *Complex Automated Negotiations: Theories, Models, and Software Competitions*, pages 223–227. Springer Berlin Heidelberg, 2013.

[17] Bjorn Leonardz. *To stop or not to stop. Some elementary optimal stopping problems with economic interpretations*. Almqvist & Wiksell, Stockholm,, 1973.

[18] Raz Lin, Sarit Kraus, Tim Baarslag, Dmytro Tykhonov, Koen Hindriks, and Catholijn M. Jonker. Genius: An integrated environment for supporting the design of generic automated negotiators. *Computational Intelligence*, 2012.

[19] Raz Lin, Sarit Kraus, Jonathan Wilkenfeld, and James Barry. Negotiating with bounded rational agents in environments with incomplete information using an automated agent. *Artificial Intelligence*, 172(6-7):823 – 851, 2008.

[20] Chhaya Mudgal and Julita Vassileva. Bilateral negotiation with incomplete and uncertain information: A decision-theoretic approach using a model of the opponent. In *Proceedings of the 4th International Workshop on Cooperative Information Agents IV, The Future of Information Agents in Cyberspace*, CIA '00, pages 107–118, London, UK, UK, 2000. Springer-Verlag.

[21] Ariel Rubinstein. Perfect equilibrium in a bargaining model. *Econometrica*, 50(1):97–109, 1982.

[22] Santiago Sánchez-Pagés. The use of conflict as a bargaining tool against unsophisticated opponents. ESE discussion papers, Edinburgh School of Economics, University of Edinburgh, 2004.

[23] Ingolf Stahl. *Bargaining theory*. Economic Research Institute, Stockholm, 1972.

[24] Colin R. Williams, Valentin Robu, Enrico H. Gerding, and Nicholas R. Jennings. Using gaussian processes to optimise concession in complex negotiations against unknown opponents. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*. AAAI Press, January 2011.

[25] Colin R. Williams, Valentin Robu, Enrico H. Gerding, and Nicholas R. Jennings. Iamhaggler: A negotiation agent for complex environments. In *New Trends in Agent-based Complex Automated Negotiations*, pages 151–158, Berlin, Heidelberg, 2012. Springer-Verlag.

[26] D. Zeng and K. Sycara. Bayesian learning in negotiation. *International Journal of Human Computer Systems*, 48:125–141, 1998.