Acting while negotiating

Yi Luo and Ladislau Bölöni
School of Electrical Engineering and Computer Science
University of Central Florida
Orlando, Florida
yiluo@mail.ucf.edu, lboloni@eecs.ucf.edu

ABSTRACT

In the convoy formation problem, two embodied agents are negotiating the synchronization of their movement for a portion of their respective paths from source to destinations. In this paper, we consider a setting in which the negotiation happens in physical time, thus the agents have the opportunity to perform actions, such as movement, while negotiating. In these settings, the agent's behavior is controlled by the pair of the negotiation and action strategies. After considering the challenges of acting while negotiating for the general convoy formation problem, we propose three static and one learning based strategies for the specific case where convoys can traverse a rectangular obstacle which is unaccessible to individual agents. Through a series of experiments we study the interaction between the action and negotiation strategies and the performance advantage of learning based approaches in incomplete information scenarios.

1. INTRODUCTION

1.1 Convoy formation in spatio-temporal domain

Let us start by defining the convoy formation problem for embodied agents. Two agents A and B move from their source positions S_A and S_B to their destinations D_A and D_B. We assume that the agents move along the paths given by the function $P_a(t) \rightarrow L$, which we read by saying that agent a is at the location L at time t.

At the initial timepoint t_0 we have $P_A(t_o) = S_A$ and we define the *arrival time* of A as the smallest time t_{arr} for which $P_A(t_{arr}) = D_A$. For every path we define the *unit cost* $c_P(t)$, and the cost of a time segment $C(t_1, t_2) = \int_{t_1}^{t_2} c_P(t)dt$. Most of the time, we are interested in the cost of the path defined as $C_P(t_0, t_{arr})$. In the simplest case we are only interested in the time to reach the destination. This corresponds to a unit cost $c_P(t) = 1$, and the cost of the path $C_P(t_0, t_{arr}) = t_{arr} - t_0$. Many environmental factors can be modeled by the appropriate setting of the unit costs. For instance, the unit cost might be dependent on the location $c_P(t) = f(P_A(t))$ or on the speed of the agent $c_P(t) = f(P'_A(t))$. Locations or speeds which are unfeasible to the agent can be set to have an infinite unit cost.

Two agents form a *convoy* if they are following the same path $P_{A+B}(t)$ over the period of time $[t_{join}, t_{split}]$. Agents join into a convoy because of the *convoy advantage*: the unit cost for the convoy is smaller than for the individual agent over the same path. One example is the case when convoys can traverse areas which are not accessible to individual agents: $\exists t \in [t_{join}, t_{split}] \;\; \exists l \; P_{A+B} = l$ with $c_{P,A}(t) = \infty$ and $c_{P,A+B}(t) = c \in \mathbb{R}$. Naturally, convoy and non-convoy segments of the path need to be continuous in space: $P_A(t_{join}) = P_B(t_{join}) = P_{A+B}(t_{join}) = L_{join}$ and $P_A(t_{split}) = P_B(t_{split}) = P_{A+B}(t_{split}) = L_{join}$. We call L_{join} and t_{join} the join locations and time, and L_{split} and t_{split} the split locations and time, respectively.

We are considering self-interested agents which are searching for the path with the smallest cost from source to destination. This path might or might not include segments traversed as a convoy. In the following we assume that the agents are using negotiation to agree on the segment traversed as a convoy. The negotiation succeeds if an agreement is reached over a quadruplet $(L_{join}, t_{join}, L_{split}, t_{split})$.

In [4, 5] we have considered a simplified convoy formation problem called Children in the Rectangular Forest (CRF), where the convoy advantage is represented by the convoys ability to traverse a rectangular obstacle which is not accessible to the individual agents. The CRF problem presents many challenges of the general problem such as the difficulty of establishing whether an offer is feasible to the opponent, whether it represents a concession or not, and the difficulty of simultaneously negotiating temporal and spatial issues. At the same time, the CRF problem simplifies away the path planning problem, as all the Pareto-optimal deals correspond to paths formed of at most three linear segments.

The work described in this paper represents a step towards bringing convoy negotiation closer to a more realistic setting. Rather than assuming that the agents are negotiating instantaneously, we assume that the negotiation process is happening in physical time, during which the agents can take real world actions, such as moving towards their destination, their expected meeting point or other locations. The immediate consequence is that in addition to the negotiation strategy, the agents also need to consider the *action strategy*. The relationship between the two is complex. A good action strategy will consider the current status of negotiation; in its turn, the actions taken by the agent will change the value

of the exchanged offers.

The remainder of this paper is organized as follows. Section 2 discusses related work. Section 3 introduces some general considerations about the acting while negotiating for the convoy formation problem. In Section 4 we define the acting while negotiating problem in the context of the CRF canonical problem extended to allow movement while negotiating. Section 5 describes the Lambda-Gamma meta-strategy for the AWN problem and three non-learning implementations: monotonic concession in space (MCS), exhaustive try (ET) and uniform concession (UC). We also describe the problem of balancing selfishness and optimism and illustrates the reduction of the offer space during the acting while negotiating process. Section 6 describes a negotiation which extends the Lambda-Gamma model with the learning of the evolving opponent model. The beliefs about the opponent model are tracked through a Sampling-Importance-Resampling particle filter. A clustering method generates concrete models which are used by the learning agents to generate offers. In Section 7 we describe several experiments studying the interaction between the action and negotiation strategies, as well as the performance advantage of the learning based approach in incomplete information settings. We conclude in Section 8.

2. RELATED WORK

Ito et al. [3] investigated the inter-dependent multi-issue problems which have nonlinear utility function. They introduced the Bidding-based Negotiation Protocol, letting the agents sample and adjust possible offers base on its own utility and then identify deals relying on a mediator. In this paper, we are more focusing on the spatio-temporal negotiation which also have non-linear utility functions, and constraints between issues. But we let the agent learn the opponent independently and select the next counter offer.

Fatima et al. [2] used a shrinking pies model to research the multi-issue negotiation problem with deadlines. They apply the back-reasoning method to concede the amount of utility which will shrink next round and the greedy strategy to concede the pie the opponent like more. In the spatio-temporal negotiations, however, a lot of possible offers in the agent's point of view are not feasible to the opponent. In addition, those number of possible solutions are also shrinking along the negotiation, but the agent can decrease this trend by moving itself to the potential deals (if it estimates correctly).

Hindriks et al. [1] used Bayesian learning to study the opponent's preference for a specific issue. They apply the probabilistic guess over a set of hypothesis of the opponent's type. These probabilities are updated based on Bayes' rule and the distance between the opponent's expected utility and the utility of actual bid. We also use the same idea to update belief for learning agent, but we use a particle filter model to realize the dynamic reasoning over time. Moreover, we don't assume that the opponent will propose the offer with linear concession in the utility. Instead, we model the expected offer by the opponent at a specific time and we update the probabilities according to the similarity between the expected offer and the opponent's actual offer.

3. GENERAL CONSIDERATIONS ABOUT ACTING WHILE NEGOTIATING

The idea that a negotiation is a process which is happening in time is not new, but many applications it is considered under strong simplifying assumptions. For instance, the split the pie game, used to model worth oriented negotiations, frequently assumes that the pie shrinks a fixed fraction at every negotiation round. Although this is a good model for motivating the agents to reach a deal as soon as possible, it does not capture the ability of the agents to take actions, and the relationship between the elapsed time and the value of offers is unrealistically simple.

In the case of the convoy formation problem, allowing acting while negotiating means that we consider every negotiation turn to take a time t_i, during which the agents can move on any feasible trajectory, naturally incurring the corresponding costs. For the remainder of this paper, we will make the assumption that t_i is a constant value. We have seen that the agents participating in a negotiation under these conditions need to have both a negotiation and an action strategy.

Let us now consider several extreme examples of action strategies. The simplest action strategy would be for the agent to stand still during the negotiation. The disadvantage of such an approach is that the value of all possible deals will become lower with the amount of time wasted during negotiating. For instance, an agent which spends 100 seconds negotiating, finding out that no deal is possible, then moving on the conflict deal trajectory, would arrive 100 seconds later than an agent which did not even negotiate. This scenario is very similar to the "shrinking pie" scenarios in worth-oriented negotiations, which also assume that the agents do not act while negotiating.

The second strategy would be to continue moving on the originally established trajectory, that of the conflict deal. This corresponds to a pessimistic agent, which up to the moment when a deal is agreed upon will assume that no deal is possible. The advantage of this choice is that the agent has a guarantee that it will not fare worst than the conflict deal. Unfortunately, moving on the path of conflict deal will reduce the value of every offer, and it can make some offers unfeasible in the sense that the agent can not reach the proposed join location L_{join} in time t_{join}.

At the other extreme, the agent might act optimistically: it can move on the shortest trajectory to the location of its own latest offer. Provided that the offer is accepted, this is the action which would provide the agent with the lowest possible cost. On the other hand, it requires a risky commitment from the agent: if no deal will be reached, or if the deal reached will be relatively far from the predicted one, the cost to destination will be actually higher than if the agent has not participated at all in the negotiation.

3.1 Baseline and pragmatic rationality

One of the consequences of the situation is that we need to refine our definition of rationality of a deal. At the beginning of the negotiation, at time t_0, the agent has a conflict deal path of cost $C_{conflict}$. According to the *baseline rationality* definition, any offer which has a higher cost than $C_{conflict}$ is not rational and it will not be accepted by the agent. If the agent is taking risks by acting optimistically, at some point in time t_x it might find itself in the position that it has already incurred costs C_x, and the best path from current location L_x to the destination will have a cost $C^x_{conflict}$. If at this moment an offer with cost C_{offer} is received, it will be

called *pragmatically rational* if $C_{offer} + C_x < C^x_{conflict}$ and *baseline rational* if $C_{offer} + C_x < C_{conflict}$. A rational agent will need to act based on the pragmatic rationality, as the original conflict deal alternative is not available any more at this moment in time. Occasionally, the agent might find it necessary to accept deals which are not baseline rational.

However, when we are measuring the overall performance of the negotiation strategy / action strategy pairs, the term of comparison should be the original conflict deal. In order for a strategy pair to be acceptable, it needs to provide at least a statistical improvement over the conflict deal.

4. ACTING WHILE NEGOTIATING IN THE CRF MODEL

In the following we shall study the issue of acting while negotiating in the Children in the Rectangular Forest (CRF) problem, an instance of the convoy formation problem where the "convoy advantage" is the ability of the convoy to traverse a rectangular region inaccessible to the individual agents. We assume the negotiation protocol to be Simple Exchange of Binding Offers (no argumentation). We also a zero-knowledge environment; the only source of information of the agents is through the offers of the opponent. We will consider the cost of a path to be the time to destination along that path.

When an agent receives an offer from its negotiation partner, it first checks it for feasibility. An offer is not feasible if the agent can not reach the designated locations on time, and we will consider these offers to have a cost of $+\infty$. For an offer $\mathbf{O} = (L_m, t_m, L_s, t_s)$ made at time t_{crt}, the agent A with source location at L^A_{src}, current location at L^A_{crt} and destination at L^A_{dest} the cost of the offer will be:

$$C^A(\mathbf{O}) = \begin{cases} +\infty \text{ if } t_{crt} + \frac{dist(L^A_{crt}, L_m)}{v_A} > t_m \\ +\infty \text{ if } \frac{dist(L_m, L_s)}{v_A} > t_s - t_m \\ t_s + \frac{dist(L_s, L^A_{dest})}{v_A} \text{ otherwise} \end{cases} \quad (1)$$

Similarly we define the cost of the conflict deal as the time spent in the negotiation until the current moment t_{crt}, plus the time necessary to reach the destination from the current location L_{crt} by going around the forest. Note that the cost of both the collaboration and the conflict deal depend on the state (the current time and location of the agent). As we discussed in the general convoy formation case, the pragmatic rationality of the offer is also state dependent. An offer might be pragmatically rational for an agent at a certain moment in the negotiation, even if its cost is higher than the original conflict deal cost. The opposite case is also possible: an offer which would have been favorable at the beginning of the negotiation might not be rational for the agent in the current state (for instance, if the agent is already well on its way towards the conflict deal).

At the other extreme from the conflict deal is the "ideal offer" with the cost C^A_{best}, which corresponds to the earliest time the agent can reach its destination, assuming an opponent which is ideally collaborative and has ideal capabilities. For a real opponent, this ideal offer might not be rational, or even feasible. We define the *utility* of an offer by the fraction of how much it can save from the cost of the conflict deal in comparison to the ideal offer.

$$U^A(\mathbf{O}) = \frac{C^A_{conflict} - C^A(\mathbf{O})}{C^A_{conflict} - C^A_{best}} \quad (2)$$

With this definition, the utility of non-rational offers is negative and the utility of non-feasible offers is minus infinity. Naturally, the utility of an offer depends on the current state. What is then, the role of the past in the agents' behavior? As it can not go back in time to change previous decisions, the agent should consider its current location and time as the starting point of the negotiation. The history of the negotiation is only relevant in the information it provides the agent about the opponent (its location, utility function, capabilities and strategy).

5. STRATEGIES FOR THE AWN PROBLEM

5.1 The Selfishness-Optimism meta-strategy

We have seen that an AWN agent requires a pair of interacting strategies for negotiating and acting. To capture the relationship between the two into an easy-to-understand framework, we propose a technique which integrates the offer acceptance decision and the action strategy into a single meta-strategy. This Selfishness-Optimism meta-strategy (see Algorithm 1) does not define the offer formation mechanism; this needs to be provided separately, and is normally inherited from non-AWN strategies.

The *selfishness* λ is the lowest utility of the offer, as defined by equation 2, which the agent is ready to accept. A fully selfish agent ($\lambda = 1$) will only accept its ideal offer, a fully benevolent agent ($\lambda = 0$) will accept any rational offer.

The *optimism* γ governs the agent's movement and represents the amount of hedging between moving towards its own latest offer versus the conflict deal location. A fully pessimistic agent ($gamma = 0$) assumes that there will be no deal and move on the conflict deal trajectory.

The reader might notice that this meta strategy can be immediately generalized by making the λ and γ parameters variable over the course of the negotiation. An agent, seeing that the opponent conceded too readily, might decide to drive a hard bargain by increasing its selfishness. An agent might make its optimism dependent on an external machine learning system which predicts the likelihood of a deal. A particularly Machiavellian agent might even make offers only to confuse the opponent and move to a predicted deal location which is far from its current offer.

For the remainder of this paper, we will assume agents with the λ and γ parameters fixed and determined at the beginning of the negotiation.

Algorithm 1 Generic behavior of agent A at time t

```
1: receive(O_B^{t-1})
2: B(t) ← B_update(B(t-1), O_B^{t-1})
3: if isFeasible(O_B^{t-1}) and U(O_B^{t-1}) ≥ λ then
4:    send(O_B^{t-1}) // form agreement
5: else
6:    O_A^t = S(B(t), λ)
7:    if not isFeasible(O_A^t) then
8:       send(∅) // conflict deal
9:    else
10:      send(O_A^t)
11:   end if
12: end if
13: L(t) = moving(L(t), B(t), γ)
```

5.2 Inferring information from offers

Let us first discuss the information available to an agent participating in an AWN-CRF negotiation. We assume a zero-knowledge environment: the only information the agents have about each other is extracted from the offers. An offer does not immediately identify the agent's source and destination, even if the agent offers its own ideal trajectory. The factor which is relatively easy to identify is the speed capability of the agent. As every offer is binding, the first offer made by an agent will identify a minimum value on the agent's speed capability based on the speed on the common trajectory portion. Unless the agent is engaged in deceptive practices, this first offer will be based on its maximum possible speed.

The agent making the second offer can find itself in one of two possible situations. It can find that the opponent's speed is larger than its own. Then it needs to structure its counter-offer based on its own, lower speed. On the other hand, if it finds the opponent's speed to be smaller than its own capability, it will make an offer assuming the opponent's speed for the common part of the trajectory, without disclosing its own higher capabilities. In either way, by the end of the first offer exchange, the agents will know their maximum common speed, and will use this in all subsequent offers. Thus, the remainder of the offers will always be feasible for the common portion of the trajectory, the one traversing the forest. It is, however, much harder to determine the current location of the opponent agent. There is thus no guarantee that the offers are feasible from the point of view of the opponent being able to reach the meeting point in time.

5.3 Three "simple" offer formation strategies

Let us now introduce three offers formation strategies for the Selfishness-Optimism meta strategy. These strategies are using learning only in the limited sense of basic information inference described in the previous section. We will describe a more complex learning-based strategy in Section 6.

Monotonic Concession in Space (MCS) calculates the next offer by conceding the location fields of its own offer, to the opponent's last offer. It is parameterized by the conceding pace at each side of the forest (C_m, C_s). The meeting time is tightly calculated based on its own ability (the physical time it will arrive the meeting location from its current location). The splitting time is calculated based on the opponent's inferred speed (the physical time both agents will arrive the splitting location in the speed of the slower agent). If the utility of the next conceding offer is below the selfishness, or no concession is possible (e.g. the opponent's last offer and the agent's last offer met together in location), the negotiation stops with no agreement.

The MCS strategy resembles the monotonic concession strategy from single-issue worth-oriented domains. There are, however, some important differences. Conceding in the meeting and splitting location does not necessarily mean any concession in terms of utility. By exploring only specific combinations of meeting and joining points, the strategy excludes a large part of the solution space.

Exhaustive Try (ET) generates a pool of all possible offers, described as combinations of meeting and splitting location with a certain resolution, as well as possible time buffers for the meeting time. The splitting time is calculated based on the maximum common speed. Only the offers which are rational, feasible and have an utility higher than the selfishness λ are included in the pool. At every round, the ET selects the offer which is the most similar to the opponent's last offer. The similarity between two offers is defined by the sum of squared difference of each issue (see Equation 3). If the offer pool is empty, the

$$\mathbf{O^t_{agent}} = \arg\min_{\mathbf{O}}(||\mathbf{O} - \mathbf{O^{t-1}_{opponent}}||^2) \quad (3)$$

Uniform Concession (UC) modifies the ET strategy by defining a conceding rate α and a current utility range (with the span of α) for each round. When calculating the next offer, the agent only searches the offers in the current utility range for the one most similar to the opponent's offer. The utility range starts at 1 and decreases with α each round until the selfishness level is reached (see Algorithm 2). Thus every offer made will be a concession of about α, in terms of the offering agent's utility.

Algorithm 2 The function to calculate next offer in the UC agent

```
Create Set⟨offer⟩ to hold all possible offers;
while Set⟨offer⟩ is empty do
    lower = lower − α;
    if lower ≤ λ then
        return O^t_next ← null;
    end if
    find all Offer that Utility(Offer) ∈ (lower, lower + α);
    add all Offer in Set⟨offer⟩
end while
find most similar Offer to O^{t−1}_opponent in Set⟨offer⟩;
return O^t_me ← offer;
```

6. A PARTICLE FILTER LEARNING STRATEGY FOR THE AWN PROBLEM

It is natural that the more an agent knows about its opponent, the more effective its negotiation will be. For instance in Figure 2 (a) and (b) we can see the difference between the agent's pool of feasible offers and the pool of possible deals. An agent which can evaluate the opponent's utility function can guarantee that all its offers are from the deal pool, thus improving the likelihood that they get accepted. The agent can also notice immediately the moment when the deal pool becomes empty, thus can interrupt the negotiation without further waste of time and utility.

As for the AWN problem the opponent is moving while negotiating as well, the problem is not only one of learning the initial parameters, but one of maintaining a dynamically evolving model of the opponent, a problem of *probabilistic reasoning over time*. In this section we describe a strategy which uses a Sampling-Importance-Resampling (SIR) particle filter to update its beliefs about the opponent, then uses a K-Means clustering technique to extract a likely hypothesis on which the offer formation is based. The resulting PF strategy is still in the Lambda-Gamma family, thus the only components we need to specify are the belief update and the offer formation mechanisms.

6.1 Update current knowledge based on opponent's offer

The PF strategy represents its knowledge about the opponent as a cloud of weighted particles. In the following we discuss (1) the particle representation, (2) the prediction model, describing how the particles evolve in time and (3) the sensor model, which describes how observations (which in our case are offers made by the opponent) affect the weight of the particle.

The particle representation

A particle should contain all the information the learning agent needs to know the opponent. We decide a particle $\mathbf{X_t}$ inside the learning agent at a specific offering time t [1] as a vector of its opponent's current state:

$$\mathbf{X_t} = \langle L_{src}, L_{crt}, L_{dest}, S_{id} \rangle$$

where L_{src} is the source of its opponent, L_{crt} is the current location of its opponent, L_{dest} is the destination, and S_{id} is the strategy its opponent uses.

The prediction model

In AWN, negotiation is proceeding in an evolving world. So the learning agent should evolve its particles along the negotiation round. Specifically, at each time when the learning agent calls, it should update its particle $\mathbf{X_t}$ from the previous one $\mathbf{X_{t-1}}$.

$$\mathbf{X_t} = \begin{cases} L_{src}(t) = L_{src}(t-1) + \xi_{src} \\ L_{dest}(t) = L_{dest}(t-1) + \xi_{dest} \\ L_{crt}(t) = f(S_{id}, L_{crt}(t-1)) + \xi_{current} \\ S_{id}(t) = S_{id}(t-1) \end{cases}$$

$\xi_.$ is a random variable generated from the two-dimensional normal distribution. $f(.)$ is a function to calculate the next location according to the opponent's strategy S_{id} and its former location $L_{crt}(t-1)$. The Gaussian noise added to the particles accounts for the uncertainty of the estimation.

The sensor model

The weights of the particle are updated according to the new observation, in our case, it is the opponent's last offer. Specifically, for each particle i, the learning agent calculates the probability to propose that offer $Pr(\mathbf{O_t}|\mathbf{X_t^i})$, where $\mathbf{O_t}$ is the opponent's last offer, and $\mathbf{X_t^i}$ is the current state of particle i that the agent assumes the opponent is in. To do this, we first calculate the offer which would have been made by the agent described by the particle $O_{exp}(X_t^i)$ and then calculate the probability based on the difference of the real offer from the expected offer:

$$\begin{aligned} Pr(\mathbf{O_t}|\mathbf{X_t^i}) &= Pr(O_t|O_{exp}(X_t^i)) \\ &= g_4(y_m, t_m, y_s, t_s | y_m^{exp}, t_m^{exp}, y_s^{exp}, t_s^{exp}) \\ &= g(y_m|y_m^{exp}) g(t_m|t_m^{exp}) g(y_s|y_s^{exp}) g(t_s|t_s^{exp}) \end{aligned}$$

In the formula, $(y_{meet}, t_{meet}, y_{split}, t_{split})$ is the actual values in opponent's last offer $\mathbf{O_t}$. $g_4(.)$ is the four-dimensional Gaussian p.d.f with centers at expected offer $O_{exp}(X_t^i)$ and with specific coefficient matrix. In our case, we simplify such matrix into diagonal matrix with specific coefficient factor for each issue. So the value equals to the product of one-dimensional Gaussian p.d.f $g(.)$ with the center at expected value and specific coefficient factor for each issue. At last, the learning agent uses this product to update the current weight of the particle at this round

$$w_i(t) = Pr(\mathbf{O_t}|\mathbf{X_t^i}) w_i(t-1)$$

All the weights of particles are normalized after the update, and if the estimate of effective number of particles in which the total number of particles is P

$$\hat{N}_{eff} = \frac{1}{\sum_{i=1}^{P} (w_i)^2}$$

is less than a given the threshold $N_{threshold}$, a resampling is performed using the stratified resampling algorithm. Then, the particles and their weights are used as the current knowledge the learning agent uses to proposes the counter-offer.

Readers familiar with particle filter representations will notice that our representation, which includes the strategy in the particle representation is unusual. The strategy does not have an obvious distance metric, and it does not evolve through the life of the particle. On the other hand, it participates in the evolution of the other components and it plays an important role in the weight update. Thus, the strategy component can also considered as a particle coloring mechanism.

A related problem is the initialization of the strategy field of the particle. While the source and destination location can be initialized through random sampling of their respective domains, the strategy can not be randomly generated from the space of all possible strategies. For our current implementation, we consider the strategy to represent a discrete choice among a small number of possible strategies, from which we choose according to some a priori probabilities.

6.2 Calculate the next offer based on the current knowledge

Algorithm 3 depicts the calculation of the next offer by the PF agent. The first main step of the algorithm is to associate each particle an offer. Specifically, for each particle i, the learning agent calculates all potential deals which are feasible for both agents and acceptable for both selfishness λ and λ_i. If there is no such deal, it assigns no offer to the particle. If there are more than one deals can be found, it assigns the offer which provides opponent best utility, assuming the opponent is indeed in the particle i.

The next step is to decide whether to propose counter-offer or not: if none of the particles have an associated offer, or the accumulated weights of particles who have assigned offer is less than a threshold, the learning agent concludes that there is no deal possible for the current state of the world, so it reports no further offer and stop negotiation.

In the third step, we perform a K-Means clustering on all the particles which have assigned offers. The distance metric used is the sum of squared difference between the issues.

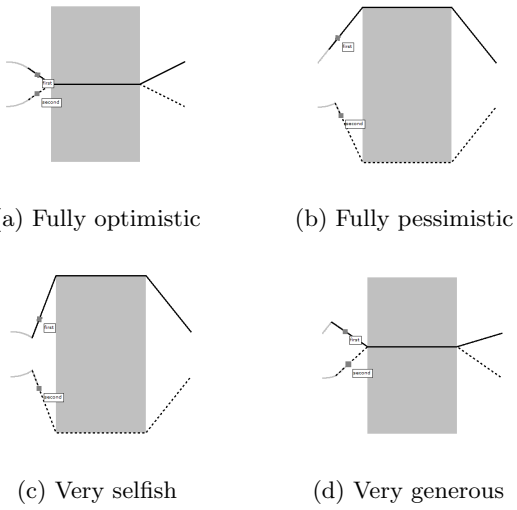
[1] This time should be the proposing order for the learning agent, because it doesn't update belief if it is not its turn to call

After clustering, the learning agent calculates the weight of cluster as the accumulated weight for all particles belongs to it and selects the cluster with highest weight. Then, it will propose the averaged offer of that cluster as the next counter-offer to the opponent.

What this process effectively does it to discover the natural grouping of the particles in several discrete hypothesis (well visible in the visualization). If we would average over the complete set of particles, the resulting estimate might fall in the low probability zone between hypotheses.

Algorithm 3 The function to calculate next offer in the PF agent

```
1: for all particle i in belief do
2:    search all O^i where U_agent(O^i) ≥ λ and
      U_opponent(O^i) ≥ λ_i;
3:    if no any O^i then
4:       O^i_best ← null;
5:    else
6:       O^i_best ← arg max U_opponent(O^i);
7:    end if
8: end for
9: if no particle has O^i_best or Σ w_i ≤ threshold then
10:   return O_next ← null;
11: else
12:   cluster all particles whose O_best ! = null;
13:   calculate weights of all clusters;
14:   find the most weighted cluster j;
15:   return O_next ← O_ave(j);
16: end if
```



(a) Fully optimistic (b) Fully pessimistic

(c) Very selfish (d) Very generous

Figure 1: The influence of the selfishness and optimism to the course and the outcome of the negotiation. The meta-strategy of agent B is fixed to $\lambda_B = 0.6$ and $\gamma_B = 1$. The values for agent A are: (a) $\lambda_A = 0.6$, $\gamma_A = 1$ - average selfishness, fully optimistic, (b) $\lambda_A = 0.6$, $\gamma_A = 0$ average selfishness, fully pessimistic, (c) $\lambda = 0.8$, $\gamma = 1$ high selfishness, fully optimistic and (d) $\lambda = 0.2$, $\gamma = 1$ low selfishness, fully pessimistic.

7. EXPERIMENTAL STUDY

7.1 The influence of the selfishness and optimism on the agent trajectories

To understand the impact of the selfishness and optimism settings on the behavior of agents, we have run a series of experiments. We considered a scenario where a mutually advantageous deal is possible. The size of the map is 600×400, with the forest located at (200,25) with the size of 200 ×350. Agent A moves from (100,150) to (500,150) with the speed of 1.0, agent B with the fixed values of $\lambda = 0.6$ and $\gamma = 1$ moves from (100,250) to (500,250) with the speed of 1.0. Both agents use the MCS strategy($C_m = 2, C_s = 2$) to calculate the next offer. This is a "hard" scenario, because the social deal is only marginally better than the conflict deal.

Figure 1 shows the path of the agents for four different settings of the selfishness and optimism for agent A. As the MCS strategy does not depend on the current location, the actual offers exchanged are identical. Interestingly, however, in cases (a) and (d) the agents agreed to collaborate, while for (b) and (c) they did not. Figure 1-a shows an agent with $\lambda_A = 0.6$ and $\gamma_A = 1$, that is, of average selfishness but fully optimistic. The agent moves towards its own offer at every step which results in a curving trajectory as the offer evolves. As the agents are getting closer and closer together, the utility of their respective offers keeps increasing, thus a deal is eventually reached.

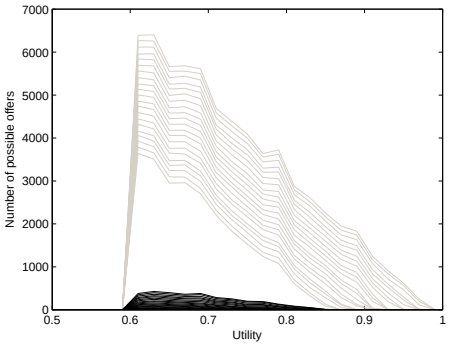
In Figure 1-b agent A is fully pessimistic and of average selfishness ($\lambda_A = 0.6$, $\gamma_A = 0$). Agent A moves in a straight line towards the conflict deal, making the offers of the opponent and its own offers less and less valuable, despite the opponents' concession. Finally, the offer which the agent needs to make according to its strategy becomes of lower utility than the conflict deal, the negotiation is broken off, and the opponents move on the conflict deal trajectory. Note that agent B actually ended up on a trajectory which is worse than the original conflict deal.

Figure 1-c shows a run with A being fully optimistic but of high selfishness ($\lambda_A = 0.8$, $\gamma_A = 1$). The trajectories are initially similar to case (a), however, A will reach a point in which its next offer will have an utility smaller than its selfishness. At this point A breaks of the negotiation and moves to the conflict deal. In this case *both* agents end up on trajectories which are worse than the original conflict deal.

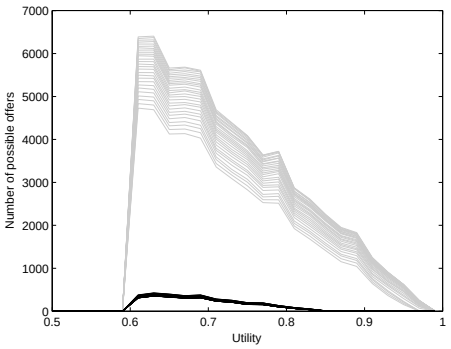
Finally, Figure 1-c shows a case when A is fully pessimistic but of low selfishness - very generous ($\lambda_A = 0.2$, $\gamma_A = 0$). Despite the fact that it starts to move towards the direction of the conflict deal, A and B successfully form a deal A will accept a relatively low utility rational offer. Thus A will reverse its course and move towards the collaborative deal. Note that A had lost some utility by making the "detour" towards the conflict deal.

7.2 The influence of the action strategy on the offer pool

Let us consider a negotiation turn where agent A needs to make an offer. We call the agent A's *offer pool*, the set of offers which are rational and feasible for A. The *supervisor's pool* is the set of offers which are feasible and rational for both A and B. Some strategies, such as ET generate the agent's pool explicitly. The supervisor's pool can not be computed by the agents in partial knowledge negotiations.



(a) Fully pessimistic



(b) Fully optimistic

Figure 2: Evolution of the histograms of offer pool (gray lines) and the supervisor's pool (black lines) function of the utility. (a) $\gamma = 0$ (fully pessimistic) and (b) $\gamma = 1$ (fully optimistic). For both cases, the $\lambda = 0.6$.

In the acting while negotiating problem, both the agent pool and the supervizor's pool decreases at every negotiation round, as some offers become unfeasible. However, which offers become unfeasible depends on the action strategy.

One way to characterize the agent and the supervisor pools is to consider the histogram of the offers in function of their utility. Figure 2 plots the evolution of these histograms over the negotiation scenario described in the previous section. Series of gray lines show the agent's offer pool, and black lines the supervisor's pool. Figure 2-a considers a fully pessimistic agent. As expected, the agent offer pool shrinks at every iteration. Furthermore, maximum utility from the agent's offer pool also becomes lower at every iteration, reflecting the fact that by moving on the conflict deal trajectory, the agent is reducing its own choices. The supervisor's pool is shrinking on its own as well, and eventually becomes empty.

Figure 2-b considers a fully optimistic agent. We note that the offer pool is still shrinking at every iteration, but the amount of decrease is smaller. Furthermore, the maximum possible utility remains very close to 1.0 during in the negotiation, because the agent optimistically moves towards these high utility offers. We also notice that the rate of shrinking of the supervisor's pool is much slower than in the pessimistic case.

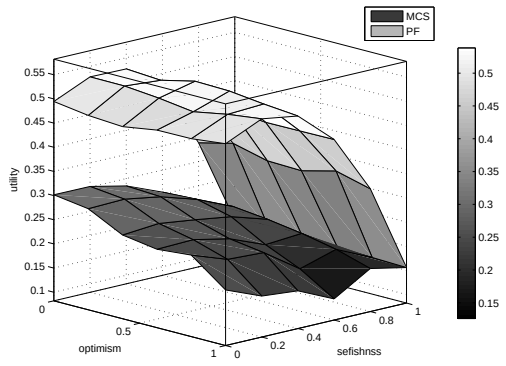


Figure 4: The statistical study of negotiation result when a PF agent and a MCS agent negotiates with the same opponent (another MCS agent) in 50 preload scenarios.

7.3 The learning process in the particle filter agent

In the AWN case every rejected offer comes with a cost in terms of loss of utility of the final deal, or it can even lead to a conflict. An agent with perfect knowledge of both the opponents physical location and destination, as well as strategy, would be able to make in the first round the perfect offer which would be (just) acceptable to the opponent and would maximize the utility for the offeror. We expect that the particle filter based agent we described, by learning during the negotiation some parameters of the opponent, should be able to achieve a higher utility and, possibly, turn negotiations ending in conflict into negotiations ending in a deal.

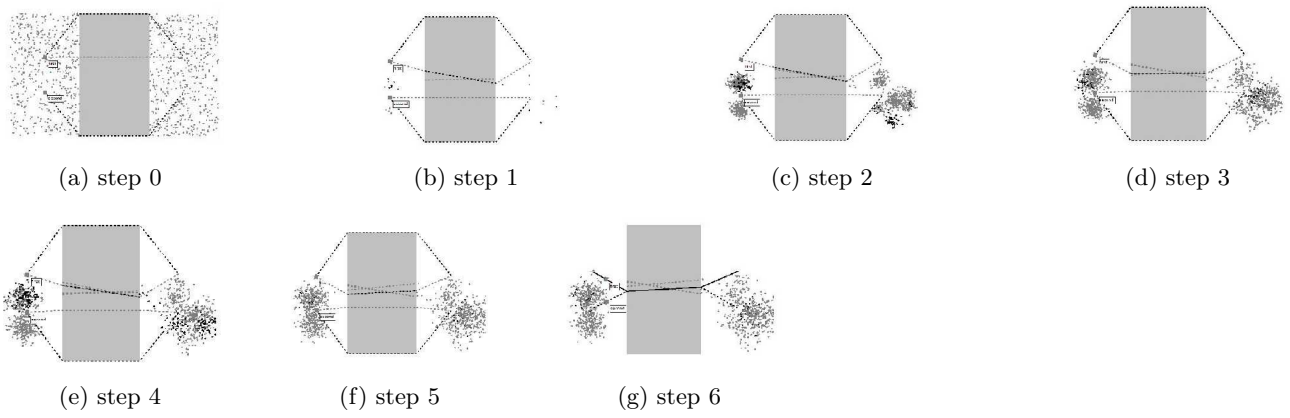
Let us consider the negotiation in Figure 1-c, with an MCS agent which is selfish ($\lambda = 0.8$), and optimistic ($\gamma = 1$), which ends in conflict. We repeat the experiment, replacing the MCS agent with a PF agent, with the same λ and γ values.

Figure

We note that the particles show a relatively large spread which changes from step to step. This is a result of the way in which the offers are formed based on the strongest cluster. If the opponent declines the offer, this represents a strong negative feedback to the selected cluster, which leads to a large variation in the particle cloud, which can be further amplified by the resampling step. Nevertheless, the particle clouds track relatively well the current location and destination of the opponent, which allows the PF agent to choose better offers from the offer pool. In our case, at negotiation round 10, the opponent accepts the PF agents offer, and they move together to their meeting location. Thus, the PF agent, under the same selfishness and optimism parameters, and starting from zero knowledge, could "save" a deal, which was lost for a MCS agent using the same parameters.

7.4 Statistical performance advantage of learning

The quality of a specific action strategy / negotiation strategy pair can be measured by the average utility of the deals it can reach over a set of randomly chosen represen-



(a) step 0 (b) step 1 (c) step 2 (d) step 3

(e) step 4 (f) step 5 (g) step 6

Figure 3: The learning progress for the PF agent. The black dots and the corresponding dashed line are the cluster and its center the learning agent selects at that round. The gray ones are those particles belongs to the other clusters.

tative scenarios against specific opponents. The statistical averaging is necessary because some strategies might be a better fit for certain scenarios: for instance, fully pessimistic action strategies will yield the best performance in scenarios where no deal is possible.

Figure 4 shows the relative utility obtained by the MCS and PF negotiation strategies for various set values of selfishness and optimism, when negotiating with a specific opponent (using the MCS strategy and the with the values of $\lambda = 0.6$ and $\gamma = 1$, used in the previous examples as well. The utility values were obtained as an average over 50 randomly generated negotiation scenarios.

As expected, the PF agent significantly outperforms the MCS strategy. The performance converges to the same value for the selfishness 1, where almost all the negotiations end up in conflict. We find that the performance of the PF agent is only moderately sensitive to the optimism value, while in the case of MCS, the performance in general decreases with the optimism. Note, however, that both agents are dealing with a fully optimistic opponent, pessimism from *both* sides would lead to a large increase in the number of conflict deals. For the PF agent, the maximum average utility is reached for a value of selfishness of approximately $\lambda = 0.4$. For values smaller than this, the performance suffers because the agent accepts deals of lower value, while for higher values, the performance decreases because a larger fragment of negotiations end in conflict.

8. CONCLUSIONS

In this paper we introduced the acting while negotiating variant of the convoy formation through negotiation problem. We have identified that the main challenge of this problem is the interaction between the negotiation strategy and the action strategy. We have introduced several negotiation strategies (three static and one learning based) for a specific case of convoy formation, the children in the rectangular forest game.

Our described work is just an initial investigation in a relatively major problem, with application both within the convoy formation problem and in other instances where agents are negotiating in physical time. Our future work involves both extending the proposed strategies to more general settings, as well as in developing more complex action strategies, such as strategies where the opponent model is used to adjust the optimism of the action strategy.

9. REFERENCES

[1] K. H. Dmytro Tykhonov. Opponent modelling in automated multi-issue negotiation using bayesian learning. In *Seventh International Conference on Autonomous Agents and Multiagent Systems (AAMAS 08)*, pages 331–338, 2008.

[2] S. S. Fatima, M. Wooldridge, and N. R. Jennings. Multi-issue negotiation with deadlines. *Journal of Artificial Intelligence Research*, 27:381–417, 2006.

[3] T. Ito, H. Hattori, and M. Klein. Multi-issue negotiation protocol for agents: Exploring nonlinear utility spaces. In *IJCAI*, pages 1347–1352, 2007.

[4] Y. Luo and L. Bölöni. Children in the forest: towards a canonical problem of spatio-temporal collaboration. In *The Sixth Intl. Joint Conf. on Autonomous Agents and Multi-Agent Systems (AAMAS 07)*, pages 986–993, 2007.

[5] Y. Luo and L. Bölöni. Collaborative and competitive scenarios in spatio-temporal negotiation with agents of bounded rationality. In *In the Proceedings of the 1st International Workshop on Agent-based Complex Automated Negotiations (ACAN 2008)*, pages 40–47, 2008.