

Artificial Intelligence Techniques

Negotiation Agent Design - Analysis Report

Group 4

4004868 Tung Phan

4409159 Francesco Corsini

1369326 Dirk Meijer

February 15, 2015

Contents

1	Introduction	2
2	Testing approach	2
3	Test results	2
4	Improvements made to the agent	3
5	Final results	5

1 Introduction

To prepare the competing negotiation agents for the upcoming tournament, all compiled agents were publically made available. This enabled us to run tests against these other agents, generating data such as the utility of the agent per session, the distances of the solutions to Pareto and Nash, social welfare and whether the agents reached an agreement or not. These results are invaluable to tweaking and optimizing our agent, which helps us prepare for the tournament.

Our testing approach will be elaborated in section 2, followed by the test results in section 3. Using these test results, we made improvements to our agent, which will be discussed in section 4, along with our reasons for those changes. Finally, we will do a quick test to verify that the changes we made actually improved the performance of our agent in section 5.

2 Testing approach

In order to produce meaningful test results, we first have to prepare a testing approach. Putting it into context, we have to choose the right negotiation profiles, the number of parties per negotiation and whether we use all the available parties.

We chose to use a different domain for our test, since we wanted even to see how our agent performed in a different domain. We picked the Tram Domain. This domain has 3 different scenarios, we therefore tested them all. We also discovered that using the other domains did not provide enough diversity in choices, resulting in almost the same results across all scenarios.

The amount of parties involved per session is set to 3, as more will only increase the amount of test results without providing actual additional useful data, and it further complicates the negotiation sessions due to the amount of parties needed to agree. We opted to leave the deadline at 180, as we see it as a reasonable deadline. For our own agent, the deadline will have minimal impact, as our behaviour scales along with the deadline. The only difference it will have is that our agent will concede quicker, because the deadline is nearer.

We omitted group 9 from our testing, because group 9 was the only group producing `NullPointerException`s.

3 Test results

We therefore tested our agent on a tournament. Following here are the results with the parameter $p = 5$ (see next section for details). Tram Domain:

- Profiles: 1,2,3
 - Total Negotiations: 216

- Average Utility: 0.8776
- Max Utility: 1.0
- Min Utility: 0.6454
- Average welfare Utility: 2.5036
- Pareto counting: 154
- Nash counting: 9
- Profiles: 4,5,6
 - Total Negotiations: 216
 - Average Utility: 0.8287
 - Max Utility: 1.0
 - Min Utility: 0.5123
 - Average welfare Utility: 2.3010
 - Pareto counting: 160
 - Nash counting: 2
- Profiles: 7,8,9
 - Total Negotiations: 216
 - Average Utility: 0.7780
 - Max Utility: 1.0
 - Min Utility: 0.5992
 - Average welfare Utility: 2.1345
 - Pareto counting: 122
 - Nash counting: 2

From the above we can clearly see that the first implementation of our agents already provide great results.

For every set of profiles(collaborative, competitive, moderate competitive) the following can be seen:

- utility always over the 77%
- more than half of the sessions have a Pareto outcome
- the welfare utility is always above 2
- the maximum utility reached is always 1
- the minimum utility is never lower than 0.5

However we were not satisfied with our utility performance, believing that a better grade could be achieved. For that we decided to change our agent, as described in the following section.

4 Improvements made to the agent

Given how the agent was programmed in first place, no big major code rework were made. This is because we already implemented the agent with a set of parameters, ready to be tweaked during this second iteration of the project. The process was to see the test results, change a little bit the parameters, run the test again and try to improve the overall performance.

The parameter we mainly worked on was the parameter that changed the shape

of the threshold function for the acceptance of the bid. For easy reference, the definition of the function is shown below:

$$T = T_0 - (T_0 - T_\infty) \times \left(\frac{r}{R}\right)^p \quad (1)$$

- T is the current threshold.
- T_0 is the starting threshold.
- T_∞ is our reservation value.
- r is the current round.
- R is the total number of rounds.
- p is a power that scales the speed of the descent. $p = 1$ is a linear descent, higher values of p make the agent wait before compromising too much.

The first one was the parameter that changed the shape of the threshold function for the acceptance of the bid. We started with $p = 5$, that made our acceptance function a good compromise between our own utility and the overall welfare. However, after intense testing, we discovered that $p = 25$ is the best for our own utility, making the function flat in the beginning of the negotiation(early stage, keep position) and steep in the end, making more concessions as the deadline approaches. This is shown in figure 1.

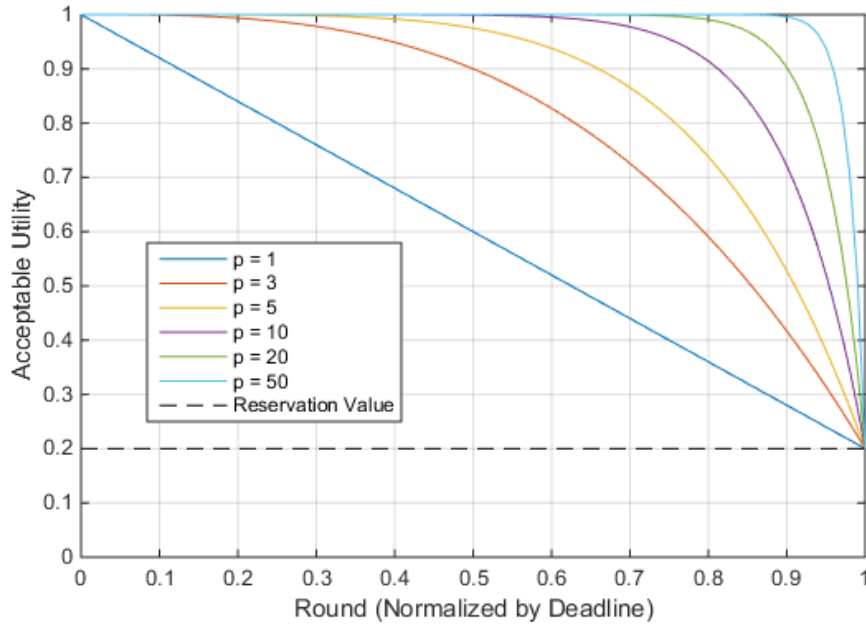


Figure 1: Our acceptance strategy for different values of p as the deadline is approached.

We made the choice to go to $p = 25$ empirically. It turned out that one agent in particular, Group7, was consistently getting a higher utility than we were by playing what we call the “Chicken game,” of waiting longer to compromise. At $p = 25$, we

have the advantage, while not harming the Social Welfare too much. (Like we would at $p > 50$.)

5 Final results

To benchmark our improvement, we performed several tournaments, with all the available agents (except for Group9, which kept throwing `NullPointerException`s) on different domains, with varying deadlines. In figure 2 we show the averaged utility over all the sessions in the tournament for each agent with $p = 25$. Our agent, Group4, and the agent Group7 outperform the rest by a small amount. And with $p = 25$, we outperform Group7 by a tiny amount, averaging at

$$E[U_4] = 0.9200$$

while Group7 averages at

$$E[U_7] = 0.918$$

This improved utility does harm the **Social Welfare** slightly, but not by a significant amount.

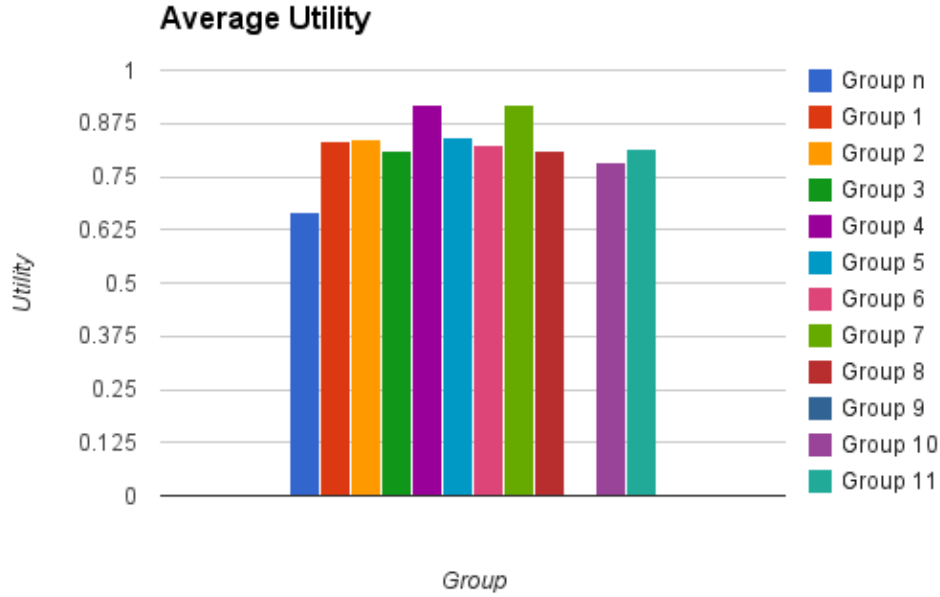


Figure 2: Mean utility, averaged over multiple tournaments with different deadline lengths.

We are overall really satisfied with our results. The agent already performed quite good for its own utility and the overall welfare, but with the minor tweaks to the parameters it performs even better, its utility averaging the highest of all agents.