Adaptive Negotiating Agents in Dynamic Games: Outperforming Human Behavior in Diverse Societies

(Extended Abstract)

Eunkyung Kim, Luyan Chi, Yu Ning, Yu-Han Chang, Rajiv Maheswaran
Information Sciences Institute, University of Southern California
4676 Admiralty Way, Marina del Rey, CA 90292
{eunkyung, lchi, yuning}@usc.edu, {ychang, maheswar}@isi.edu

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Intelligent agents, Multiagent systems*

General Terms

Algorithms, Economics, Experimentation

Keywords

Multi-Agent Systems, Game Theory, Ultimatum Game, Mathematical Models of Human Behavior, Learning, Adaptation

1. INTRODUCTION

Creating software agents that can negotiate effectively is an important problem that has been studied by agent researchers in contexts such as the trading agent competition and the virtual agents community. In the former, the goal is typically to find optimal policies in settings with uncertain and incomplete information, and where policies are typically evaluated in societies of entirely artificial agents [6]. In the latter, a goal is to create agents that can interact with humans — in many cases, to train them in negotiation with individuals from particular cultures or different value settings [5].

In this paper, we examine a problem that combines the complexities of these goals. We want to create negotiating agents that can perform effectively in multiple environments, specifically in a multitude of societies where values and styles of negotiation might be significantly different. Since agent performance is highly dependent on the interaction environment, the design of such an agent is not a straightforward optimization problem.

As context for this investigation, we use the Social Ultimatum Game [2], a multi-agent multi-round extension of the Ultimatum Game, a classical game-theoretic problem which has been studied for decades due to the behavioral variance it elicits. It has been shown through many investigations that humans exhibit a wide range of behaviors that deviate from a "rational" payoff-maximizing strategy based on factors such as cultural background, occupation and emotional factors among others in the classical Ultimatum Game.

Appears in: *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)*, Conitzer, Winikoff, Padgham, and van der Hoek (eds.), 4-8 June 2012, Valencia, Spain.

2. SOCIAL ULTIMATUM GAME

The Ultimatum Game, is a two-player game where a player, P_1 proposes a split of an endowment $e \in \mathbb{N}$ to another player P_2 where P_2 would receive $q \in \{0, \delta, 2\delta, \ldots, e-\delta, e\}$ for some value $\delta \in \mathbb{N}$. If P_2 accepts the offer, they receive q and P_1 receives $e-q$. If P_2 rejects, neither player receives anything. The subgame-perfect Nash or Stackelberg equilibrium states that P_1 offer $q = \delta$, and P_2 accept. This is because a "rational" P_2 should accept any offer of $q > 0$, and P_1 knows this. Yet, humans make offers that exceed δ, even making "fair" offers of $e/2$, and reject offers less than the minimum.

To represent the characteristics that people operate in societies of multiple agents and repeated interactions, we introduce the Social Ultimatum Game. There are N players, denoted $\{P_1, P_2, \ldots, P_N\}$, playing K rounds, where $N \geq 3$. The requirement of having at least three players in necessary to give each player a choice of whom to interact with.

In each round k, every player P_m chooses a single potential partner P_n and makes an offer $q^k_{m,n}$. Each player P_n then considers the offers they have received and makes a decision $d^k_{m,n} \in \{0, 1\}$ with respect to each offer $q^k_{m,n}$ to either accept (1) or reject (0) it. If the offer is accepted by P_m, P_m receives $e - q^k_{m,n}$ and P_n receives $q^k_{m,n}$, where e is the endowment to be shared. If an offer is rejected by P_n, then both players receive 0 for that particular offer in round k. Thus, P_m's reward in round k is the sum of the offers they accept from other players (if any are made to them) and their portion of the proposal they make to another player, if accepted, $r^k_m = (e - q^k_{m,n})d^k_{m,n} + \sum_{j=1\ldots N, j\neq m} q^k_{j,m} d^k_{j,m}$. The total rewards for P_m over the game is the sum of per-round winnings, $r_m \sum_{k=1}^{K} r^k_m$.

3. AUTONOMOUS AGENTS

We summarize the types of agents that we implemented.

- **Tit-for-Tat** : This is a fully reciprocal agent that chooses responders who previously made them offers, and offers an amount that reciprocates that previous offer,
- **Regret Minimization** : This agent minimizes worst-case regret by hedging [1] among a set of available actions. It hedges by increasing the weights associated with high payoff actions during gameplay, and probabilistically chooses actions based on these weights, which are initialized using human data,
- **Expected Reward QRE** : This agent learns the ex-

pected rewards of various actions based on human play data and acts using a quantal response equilibrium [3] strategy based on these rewards.

- **SIGAL QRE** : This agent also uses a quantal response equilibrium strategy but the utility is based on the sigmoid acceptance learning [4] approach which incorporates a model of social utility into the rewards.
- **Adaptive Fairness** : This agent is characterized by a fairness threshold which is dynamically updated based on an adaptability parameter and an exploration parameter [2]. It accurately replicates human dynamic reciprocity behavior and is used as a stand-in for various human-like behaviors that are learned from data.
- **Marginal Value Optimization** : This agent chooses an action based on the marginal value of being seen as the preferred partner of each agent in the society. The value is a product of the expected value of the offer received from a particular agent and the marginal increase of the likelihood of receiving an offer.

4. EXPERIMENTS

In order to investigate adaptiveness of the agents and of the humans, we created 10 different societies. We first ran two sets of human experiments, one with undergraduates and staff at a U.S. university, and a second at an international conference with primarily computer science doctoral students and faculty. From this data we estimated parameters for the Adaptive Fairness (AF) agents, using different subsets of humans.

This includes the top 25% scorers at the conference, the top 25% scorers at the university, two clusters of the human population based on offer recipient entropy (people who spread their offers out the most and the least), and four humans drawn randomly from the populations. In addition, SIGAL-QRE, ER-QRE and Regret Minimization agents were created with data from the first two experiments. We then created the following 10 societies for 5-player games where one *test player* plays against four players:

- **AF-Conf-Top25** : 4 Conference Top 25% AF-agents
- **AF-Univ-Top25** : 4 University Top 25% AF-agents
- **AF-Cluster1** : 4 low recipient entropy AF-agents
- **AF-Cluster2** : 4 high recipient entropy AF-agents
- **AF-Alpha-7** : 4 AF-agents for Human #2
- **AF-4Types** : AF-agents for 4 types of humans
- **SIGAL-QRE** : 4 SIGAL-QRE agents
- **ER-QRE** : 4 ER-QRE agents
- **Regret** : 4 Regret Minimization agents
- **TFT-2** : 4 Tit-for-Tat agents with baseline $2 offers

We ran a third set of human experiments using Amazon Mechanical Turk where a human player could play against the societies above in 20-round games with a $10 endowment per round. We created on HIT (Human Intelligence Task) for each instance of a game in a society with 20 assignments, i.e., we had 20 human game play traces for each society type.

We then tested the following agents in each of the societies, running 1000 iterations of games for each: Regret, SIGAL-QRE, ER-QRE, TFT-2 and Marginal Value Optimization (MVO).

Mean of Payoffs		Society									
		AF-Conf-Top25	AF-Univ-Top25	AF-Cluster 1	AF-Cluster 2	AF-Alpha-7	AF-4Types	SIGAL-QRE	ER-QRE	Regret	TFT-2
Test Player	Human	199.2	193.5	149.7	173.1	86.9	211.6	167.4	186.4	138.7	192.0
	MVO	203.6	229.3	215.4	215.1	111.4	221.3	184.1	174.7	180.2	209.9
	SIGAL-QRE	185.7	109.6	147.8	140.1	64.3	168.4	200.0	210.2	178.3	205.9
	ER-QRE	146.2	138.5	141.5	131.5	88.0	167.4	190.8	200.0	170.5	202.7
	Regret	172.3	98.4	143.9	140.9	62.7	162.1	224.1	231.3	200.0	199.0
	TFT-2	93.4	162.9	95.7	113.8	104.8	167.3	209.0	207.5	199.3	200.0

Figure 1: Mean of Payoffs for Test Players

5. RESULTS

Figure 1 shows the mean of payoffs for the test players in the 10 different agent societies. The main result of the paper is that the marginal value optimization (MVO) agent outperforms human players in 9 out of 10 societies. In 7 out of 10 societies the gaps in mean payoff were very high (MVO advantages were 16.6, 17.9, 24.5, 35.9, 41.6, 42.1, 65.7) The only society where it does not outperform humans is the ER-QRE society (-11.7) which is made up of agents which follow a static policy. We see that MVO's assumptions about generating payoffs from others by being the top target is validated, as MVO is able to generate more payoffs from offers made to it by others, when compared to human players.

Furthermore, MVO is also able to generate more payoffs from its own offers when compared to humans in 7 out of 10 human societies. This is because the generous offer reduces the probability of rejection in several of the societies. However, it pays a price for this in societies where the probability of rejection is low (or zero). In two of the three cases, it is able to overcome this loss from improvement in the number and quality of offers made to it by others.

Acknowledgements. This material is based upon work supported by the AFOSR Award No. FA9550-10-1-0569 and ARO MURI award No. W911NF-11-1-0332.

6. REFERENCES

[1] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire. Gambling in a rigged casino: the adversarial multi-armed bandit problem. In *Proceedings of the 36th Symposium on Foundations of Computer Science*, 1995.

[2] Y.-H. Chang, T. Levinboim, and R. Maheswaran. The social ultimatum game. In *Decision Making with Imperfect Decision Makers*, 2011.

[3] R. McKelvey and T. Palfrey. Quantal response equilibria for extensive form games. *Experimental Economics*, 1:9–41, 1998.

[4] N. Peled, Y. Gal, and S. Kraus. A study of computational and human strategies in revelation games. In *Ninth International Conference on Autonomous Agents and Multi-Agent Systems*, 2011.

[5] D. Traum, S. C. Marsella, J. Gratch, J. Lee, and A. Hartholt. Multi-party, multi-issue, multi-strategy negotiation for multi-modal virtual agents. In *Proceedings of the 8th international conference on Intelligent Virtual Agents*, IVA '08, pages 117–130, Berlin, Heidelberg, 2008. Springer-Verlag.

[6] M. P. Wellman, A. Greenwald, and P. Stone. *Autonomous Bidding Agents: Strategies and Lessons from the Trading Agent Competition*. MIT Press, 2007.