Strategic Advice Provision in Repeated Human-Agent Interactions

Amos Azaria,[1] Zinovi Rabinovich, [1] Sarit Kraus[1]
Claudia V. Goldman,[2] Ya'akov Gal[3]
[1] Department of Computer Science, Bar-Ilan University, Ramat Gan 52900, Israel
[2] General Motors Advanced Technical Center, Herzliya 46725, Israel
[3] Department of Information Systems Engineering, Ben-Gurion University of the Negev, Israel
{azariaa1,sarit}@cs.biu.ac.il, zr@zinovi.net, claudia.goldman@gm.com, kobig@bgu.ac.il

Abstract

This paper addresses the problem of automated advice provision in settings that involve repeated interactions between people and computer agents. This problem arises in many real world applications such as route selection systems and office assistants. To succeed in such settings agents must reason about how their actions in the present influence people's future actions. The paper describes several possible models of human behavior that were inspired by behavioral economic theories of people's play in repeated interactions. These models were incorporated into several agent designs to repeatedly generate offers to people playing the game. These agents were evaluated in extensive empirical investigations including hundreds of subjects that interacted with computers in different choice selections processes. The results revealed that an agent that combined a hyperbolic discounting model of human behavior with a social utility function was able to outperform alternative agent designs. We show that this approach was able to generalize to new people as well as choice selection processes that were not used for training. Our results demonstrate that combining computational approaches with behavioral economics models of people in repeated interactions facilitates the design of advice provision strategies for a large class of real-world settings.

Introduction

This paper addresses problems central to the design of advice provision strategies for computer agents that interact with people in repeated settings. We model these interactions as a family of repeated games of incomplete information called *choice selection processes* comprising a human and computer player. Both of the participants in a choice selection process are self-interested. The computer possesses private information regarding the states of the world which influences both participants' outcome; this information is not known to the person. At each round, the computer suggests one of several choices to the person, and the person then selects his or her choice, which may or may not correspond to the computer's suggestion. This choice affects the outcome for both the person and the computer agent.

We designed several models of human behavior in choice selection processes that incorporated quantile response, exponential smoothing, and hyperbolic discounting theories from behavioral economics. The parameters of these models were estimated using maximum likelihood, and their predictive power was measured on a sampled set of human play using ten-fold cross-validation. The best model, a combination of hyperbolic discounting and quantile response, was incorporated into an agent.

Clearly, an agent which only takes its own outcome into account when providing advice, will not perform as well in the long run, since the user will learn not to listen to its advice. Therefore, we propose that, in order for the agent to optimize its overall outcome, it should base its advice on a weighted sum of both the agent and user outcomes and recommend the action with the optimum weighted sum each round. The agents were evaluated in simulations using thousands of rounds as well as in studies comprising actual people.

Past work on advice provision spans the computational and social sciences. The majority of works on persuasion games study one-shot interactions (Sher 2011). A notable exception is the work by Renault et al. (2011) who considered repeated interactions that follow a Markov chain observed solely by the Sender. These works make the strong assumption that people follow equilibrium strategies. However, agents using equilibrium approaches to interact with people are not successful in repeated settings (Peled, Gal, and Kraus 2011).

Empirical Methodology

Our empirical methodology comprises a family of selection processes that are analogous to a route-selection task between a driver (a human Receiver) and a navigation system (an agent Sender). At each round of the interaction, the driver needs to choose one of 4 possible routes to get to work. The system can advise the driver to take one of the routes before the driver makes a choice. The road conditions (i.e., travel time and fuel consumption) constitute the state of the world, and vary due to traffic and maintenance. They are unknown to the driver when it makes its decision. The driver's goal is to minimize the travel time over all rounds, and the system's goal is to reduce fuel consumption over all rounds. After the driver chooses a route, both participants incur a cost which depends on the road conditions for the chosen route. At this point the interaction continues to the next round with probability 0.96. The conditions for the

Table 1: Simulation results comparing agent strategies

agent strategy	fuel (liters)	time (minutes)
Random	6.120	64.40
Silent	6.297	63.04
MDP	5.792	65.92
SAP	**5.520**	64.54

Table 2: Performance Results against People

method	selfishness	fuel	time	acceptance
Silent	–	6.20	64	–
Receiver	0	6.44	56.6	63.6%
Sender	1	5.88	64.32	31.0%
SAP-ϵ	0.29	5.76	56.6	70.8%
SAP-hyper	0.58	**5.08**	64.8	52.6%

roads at each round are sampled from a joint distribution that is known to the agent, but not to the driver.

Model Selection and Agent Design

We collected 2250 rounds from 90 subjects to train and evaluate short-term memory (ST), hyperbolic discounting (Hyper), SoftMax (SM), and Exponential Smoothing (ES) models. For each of these models, we estimated the maximum-likelihood (ML) value of the parameters, and computed the fit-to-data of the test set using the ML values. All results were confirmed to be statistically significant. The Hyper model, which modeled the Receiver using hyperbolic discounting theory exhibited a higher fit-for-data than all other models of human Receivers.

We hypothesized that the use of a social utility approach, where the agent advises the road which minimizes a weighted linear combination of both the fuel consumption and the travel time, will lead to best performance for the agent, measured in terms of fuel consumption. Intuitively, this weight provides a trade-off between the short term gain received for recommending an action with low cost for the agent and the long term gain from increasing the user's trust in the agent by advising an action with low cost for the user.

To evaluate this hypothesis, we used different agent designs for generating offers to people which incorporate the decision-making strategies that were described in the previous section. Specifically, we used an agent that incorporated the social utility approach to make offers and found the optimal weights using search methods. We termed this agent Social agent for Advice Provision (*SAP*). We also used an agent approximating the optimal policy using an MDP model to make offers, termed *MDP*. Finally, we employed two baseline agents, *Random* which offered roads with uniform probability and *Silent* which didn't provide any advice.

We evaluated these agent designs in simulation as well as in experiments involving new people. The simulation studies consisted of sampling 10,000 road instances according to the distribution over the fuel consumption and travel time. As shown in Table 1, the cost accumulated by the SAP agent using the hyperbolic discounting mode was significantly lower than the cost to all other agents.

Evaluation with People

Given the demonstrated efficacy of the SAP agent in the simulation described above, we set out to evaluate the ability of the SAP agent to generalize to new types of settings and new people. We hypothesized that a SAP agent using the hyperbolic discounting model to describe Receiver behavior would be able to improve its performance when compared to the SAP agent using the $\epsilon-$greedy model. We randomly divided subjects into one of several treatment groups. Subjects in the *Silent* group received no advice. Subjects in the *SAP-hyper* group were advised by the SAP agent using a hyperbolic model to describe Receiver behavior. Subjects in the *SAP-ϵ* group were advised by the SAP agent that used an $\epsilon-$greedy strategy to describe Receiver behavior. Subjects in the *Receiver* group were consistently advised to choose the road that was most beneficial to them, (i.e., associated with the lowest travel time). Lastly, subjects in the *Sender* group were consistently advised to choose the road which was best for the Sender (i.e., associated with the lowest fuel consumption).

Table 2 shows results for evaluating the models with new people on the same instances used to train the models. The performance for agents and for people is measured in terms of overall fuel consumption and commuting time, respectively. The "selfishness" column measures the degree to which the agent was self-interested. As shown by table 2, the SAP-hyper agent significantly outperformed all other agent-designs.

Conclusions and Future Work

In this paper we consider a two player game, in which an agent repeatedly supplies advice to a human user followed by an action taken by the user which influences both the agent's and the user's costs. We present the Social agent for Advice Provision (SAP) which models human behavior combining principles known from behavioral science with machine learning techniques. We show that SAP outperforms other alternatives. This work provides the first computational study of strategic advice provision in repeated settings, and demonstrates the efficacy of using behavioral economic models to inform agent design for these tasks.

In future work we intend to study scenarios where the user has other agents providing advice which have different cost functions than our agent. We will also allow future users to turn on or off the advice.

References

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

Renault, J.; Solan, E.; and Vieille, N. 2011. Dynamic sender-receiver games. Unpublished manuscript.

Sher, I. 2011. Credibility and determinism in a game of persuasion. *Games and Economic Behavior* 71(2):409 – 419.