Learning One-shot Strategic Interactions Among Humans and Robots

Chia-Yin Shih
Machine Learning Department
Carnegie Mellon University
Pittsburgh, PA
chiayins@andrew.cmu.edu

Geoffrey J. Gordon
Machine Learning Department
Carnegie Mellon University
Pittsburgh, PA
ggordon@cs.cmu.edu

Abstract

As intelligent systems have been increasingly introduced to the everyday lives of people, it becomes important to develop accurate models for the humans. A key component in deploying effective and reliable intelligent systems around humans is to be able to make predictions of the behavior of the humans. Game theory is a framework that models strategic interactions among multiple agents. While the literature has focused on developing and evaluating models in the case where humans are interacting with humans, we provide a comprehensive evaluation of different models when the humans believe that they are playing against robots. We found that when humans believe that they are playing against robots, there are no significant differences in the prediction results of the four models we considered. In addition, we found that these models have on average higher prediction performance when applied to the case where humans believe they are playing against humans based on data in existing literature, although we recognize that the data collection methodology is different and that the significance is not large. Based on these results, we provide directions for future research.

1 Introduction

In recent years, intelligent systems have been making their ways into the everyday lives of people. For example, the Google self-driving car project [9] has been developing autonomous vehicles that can navigate urban roads with a mix of human drivers and autonomous drivers. Google Assistant [8] has been tapping into the market as a convenient assistive device through speech recognition and human intention modelling. Assistive robotics [11, 19, 6, 10, 3] has also been of huge interest in academia for its potential in providing a higher quality for people with disabilities. Furthermore, intelligent tutoring systems [13, 14, 7] have also been an exciting area of research, and have the potential to make education more personalized and effective. The future of the integration of intelligent systems and humans in everyday lives is limitless. Hence, the importance of developing algorithms for these intelligent systems to enable them to interact effectively with humans is at an all time high.

Key to seamless and successful interactions between intelligent systems and humans is the ability for the intelligent systems to infer human intentions. For example, when observing the past trajectory of a human car, it should be able to infer whether the human may be planning to make a turn [15]; when a human is trying to operate a robot arm to reach a goal, the robot should be able to infer which goal the human is reaching towards [12]; when a human speaks, the voice and intent recognition system should be able to correctly infer the goal and intent of the human [8]. To correctly infer intention, we often need past data and a good human model. For example, in inverse reinforcement learning, we assume that the human is a best-response optimizer where his/her goal is to optimize a cost function when the world is modeled by a Markov Decision Process. In other places, we could a model that the human will take the path with the smallest cost where the cost could be a very simple cost function like the distance to the goal.

However, all of the human models mentioned in the previous paragraph assume that the human does not reason about what the intelligent system would do if she takes an action; instead these models only take into account the current state of the world. However, this is critical as during any interaction with two agents because often times interactions are strategic. For example, when we have an intelligent system and a human interacting with each other, we may be able to make a better prediction of what a human would do by modelling a larger number of iterative steps of reasoning, meaning, we model that the human forms a model of the intelligent systems' thought process based on different factors, including how the intelligent system would model the human. And this goes on recursively.

Strategic interactions form the basis of how humans interact with their surrounding environment every day. As more and more intelligent systems are deployed in our daily lives, having an accurate strategic model of the human becomes very important as this provides crucial information of how we should develop our algorithms for intelligent systems. Game theory provides a framework for modelling strategic interactions among multiple agents. In this project, we consider the case where we make predictions of one-shot strategic interactions among humans and robots. In particular, humans are made to believe that they're playing strategic games with a robot. We are interested in evaluating the predictive performance of existing models and extensions of existing models in the literature on modelling strategic behavior among humans. We would also like to gain insight on the parameters learned by these models based on the data we collected. We discover that these models have on average higher predictive performance on the existing dataset [18] where humans believe they are playing against humans than the case where humans are made to believe they are playing with robots, although we recognize the data collection methodology is different and that there is no significant difference among the two datasets. We applied several existing models and extensions of existing models to our dataset. Our experiments show that there are no significant differences in performance among the models that we considered. Based on these results and the existing literature, we gained insightful information in the model and design of human-robot interaction algorithms in the future.

2 Related Work

Several methods have been proposed to model strategic decisions among humans. Game theory is a field first initiated in economics to model human behaviors in the markets. These models often assume that humans have perfect rationality and make predictions based on the idea that the humans would perform actions at equilibrium, that is both players play actions such that none of the players can achieve a better payoff if they deviate from their current action unilaterally. However, in controlled experimental settings, previous work has reported that equilibrium is not a good model for human behavior prediction [20]. Instead, humans often exhibit *bounded rationality* [17]. This suggests that instead of thinking very deeply, humans tend to have limited depth of reasoning. Hence, papers have proposed to use hierarchical models, like the level-k model (Lk)[5] and quantal level-k model (QLk) [18]. However, these models have only been applied to scenarios where humans believe they are playing against other humans.

Recently, there has been a significant rise of interest in interactions between humans and robots. However, very few tried to model humans as strategic players against robots. Many works have used best response model to model the humans, i.e., the humans will best respond to the robot and use one level of reasoning — if the robot performs a certain action, the human will perform based on models like inverse reinforcement learning [2], [1] etc. In addition, these models are used in situations where robots are interacting with the humans over a certain time horizon and the robots have the ability to influence the behavior of the human through the duration of the interaction [16]. In this work, we instead focus on how a human would perceive the robot before any interaction and focus on evaluating models for a one-shot strategic decision making process.

3 Method

Before we describe the methods in detail, we would like to define necessary terms in game theory.

Defintion 1 ***Mixed strategy*** $s_i(a_j)$***:*** *A mixed strategy of an agent i is defined by the probability of taking each action a_j, i.e., $s_i(a_j)$ is the probability that agent i will take action j.*

Defintion 2 ***Strategy profile*** s_{-i}***:*** *A strategy profile* s_{-i} *represents the mixed strategy of each of the agent in the game except agent i.*

Defintion 3 ***Utility function*** $u_i(a_j, s_{-1})$***:*** *The utility function* $u_i(a_j, s_{-i})$ *represents the utility that agent* i *will get if she played action* a_j *and the other agents play according to the strategy profile* s_{-i}*.*

Now we will describe the models that we are using for our experiments. In the literature, instead of modelling humans as equilibrium agents, there have been two central ideas aimed at modelling human strategic behaviors: **quantal best response** and **level-k** modelling.

Defintion 4 ***Quantal best response (QBR):*** *A quantal best response* $QBR_i(s_{-i}|\lambda)$ *of an agent i is the best response mixed strategy* $s_i(a_j)$ *to the strategy profile* s_{-i} *such that*

$$s_i(a_j) = \frac{e^{\lambda u_i(a_j, s_{-i})}}{\sum_{a'_j} e^{\lambda u_i(a'_j, s_{-i})}}$$

where λ *is a nonnegative real number.*

The definition of QBR suggests that the probability of the agent taking an action a_j is proportional to $e^{\lambda u_i(a_j, s_{-i})}$. This is essentially taking a softmax of λ times the utility function. λ is called the *precision parameter*. We can see that when $\lambda = 0$, the agent chooses an action uniformly at random and when $\lambda \to \infty$, the agent chooses the action with the maximum utility.

Level-k is a key concept in modelling that humans have *bounded rationality* and perform *iterated strategic reasoning*. The idea is that a population consists of agents with different levels. A level-0 agent is non-strategic and in the majority of the literature, in this case, it picks an action uniformly at random. For example, a level-1 agent best responds to level-0 agents; similarly, a level-2 agent best responds to level-1 agents. Generally, a level-k agent best responds to level-$(k-1)$ agents. There has also been proposed models in literature that a level-k agent best responds to a distribution of lower level agents, level-0, ..., level-$(k-1)$ agents [4].

3.1 Lk model

In this model, a level-k agent best responds to the level-$(k-1)$ agents with probability $1-\epsilon_k$ and chooses a non-optimal action with probability ϵ_k. In addition, the higher level agents do not know that the lower level agents make mistakes.

First we define the following variables:

- A_i: player i's action set
- $BR_i(s_{-i})$: the set of best response actions of agent i to the strategy profile s_{-1}
- $IBR_{i,k}$: the set of iterative best response of a level-k agent i. This means $IBR_{i,0} = A_i$ and $IBR_{i,k} = BR_i(IBR_{-i,k-1})$.
- $\pi_{i,k}$: the distribution of actions that a level-k agent i plays

Then the model predicts the agents play according to the following:

$$\pi_{i,0}(a_i) = \frac{1}{|A_i|}$$

$$\pi_{i,k}(a_i) = \begin{cases} \frac{1-\epsilon_k}{|IBR_{i,k}|} & \text{if } a_i \in IBR_{i,k} \\ \frac{\epsilon_k}{|A_i| - |IBR_{i,k}|} & \text{otherwise} \end{cases}$$

As previous literature suggested, we experiment with the case where $k = 2$, i.e., there are three levels of agents: level-0, level-1, and level-2 agents.

ϵ_1, ϵ_2 denotes the probability that level-1 and level-2 agents do not best respond respectively. $\alpha_0, \alpha_1, 1-\alpha_0-\alpha_1$ denotes the distribution of level-0, level-1, and level-2 agents. Hence, we have four parameters in this model: $\epsilon_1, \epsilon_2, \alpha_0, \alpha_1$.

3.2 Quantal Level-k (QLk) model

We again choose $k = 2$ for this model. In this model, instead of modeling the agents as best responding with some probability of error, the model assumes that a level-k agent will quantally best respond to the level-$(k-1)$ agents. Mathematically, the model predicts the agents play according to the following

$$\begin{aligned}
\pi_{i,0}(a_i) &= \frac{1}{|A_i|} \\
\pi_{i,1}(a_i) &= QBR_i(\pi_{-i,0}|\lambda_1) \\
\pi_{i,2}(a_i) &= QBR_i(\gamma|\lambda_2)
\end{aligned}$$

γ is the perceived precision parameter of the level-1 agents to the level-0 agents by the level-2 agents.

$\alpha_0, \alpha_1, 1-\alpha_0-\alpha_1$ denotes the distribution of level-0, level-1, and level-2 agents. λ_1, λ_2 denotes the precision of the level-1 and level-2 agents respectively. γ is the level-2 agents' belief about the precision of the level-1 agents. Hence, we have four parameters in this model: $\lambda_1, \lambda_2, \gamma, \alpha_0, \alpha_1$.

3.3 QLkNoZero model

This model is similar to the QLk model except that we adopt the idea that there are no 0-level agents in the population.

$\alpha_1, 1-\alpha_0$ denotes the distribution of level-0, level-1, and level-2 agents. λ_1, λ_2 denotes the precision of the level-1 and level-2 agents respectively. γ is the level-2 agents' belief about the precision of the level-1 agents. Hence, we have four parameters in this model: $\lambda_1, \lambda_2, \gamma, \alpha_1$.

3.4 QLkWeighted model

This model is an extension of the QLk model. There are three levels of agents in this model, level-0, level-1, and level-2 agents. We adopt a richer representation for the parameters in the quantal best response (QBR) parameterization. Instead of having the action distribution be the softmax of $\lambda u_i(a_j, s_{-i})$, we have the action distribution be the softmax of $w_j^T u_i(a, s_{-i})$ where $w_j \in \mathbb{R}^{|A_i|}$ and $u_i(a, s_{-i}) = \begin{bmatrix} w_i(a_1, s_{-i}) & \cdots & w_i(a_{|A_i|}, s_{-i}) \end{bmatrix}^T$.

$$s_i(a_j) = \frac{e^{w_j^T u_i(a,s_{-i})}}{\sum_{a'_j} e^{w_{j'}^T u_i(a,s_{-i})}}$$

Mathematically, the model predicts the agents play according to the following

$$\begin{aligned}
\pi_{i,0}(a_i) &= \frac{1}{|A_i|} \\
\pi_{i,1}(a_i) &= QBR_i(\pi_{-i,0}|w_j) \\
\pi_{i,2}(a_i) &= QBR_i(w_j|w_j)
\end{aligned}$$

$\alpha_0, \alpha_1, 1-\alpha_0-\alpha_1$ denotes the distribution of level-0, level-1, and level-2 agents. We learn the parameters w_j for $j \in \{1, \ldots, |A_i|\}$. Hence we learn $|A_i|^2 + 2$ parameters in this model.

4 Results

4.1 Experiment methodology

We recruited 45 participants on Amazon Mechanical Turk for the experiment. To ensure the quality of the data gathered, only participants from the US with an approval rate of at least 95% can participate.

We provide thorough instructions on how the strategic matrix game works and each participant has to receive a 100% mark on the quiz involving the rules of the games and how the game works. Each participant is given 10 different 3×3 matrix games. To avoid the learning of human subjects during the experiment, we did not reveal the robot's action in each round of the game. In total, we collected 450 data points for the purpose of our experiment.

4.2 Experiment results

In our analysis, we look at the models from a machine learning perspective and we report the prediction accuracy of each model on the test set.

We also applied these methods to the data collected from [18]. This paper collected data for 40 participants with each participant playing 10 games. The games used in our experiment are identical to the ones used in this paper. However, our subject pool is different so our results may not be perfectly comparable.

The metric reported is the average negative log likelihood on the test data per game, i.e., the average negative log likelihood is the sum of the negative log likelihood of the test data divided by (number of players in the test data $\times$ 10). We use the Nelder–Mead algorithm to optimize the objective function in these experiments. In addition, to make the optimization procedure robust to local optima, we used 500 random restarts and picked the parameters resulted in the smallest objective value. We use 5-fold cross-validation and report the average of the mean negative log likelihood from these 5 folds. We kept each subject's data together in a single fold. We also introduce a baseline model where the agent chooses one of the three actions uniformly at random.

The prediction performance of each model on the two datasets is summarized below:

	Lk	QLk	QLkNoZero	QLkWeighted	Baseline
Human-Robot (HR)	0.877	0.871	0.886	0.888	1.099
Human-Human (HH)	0.775	0.853	0.847	0.800	1.099

Table 1: Cross-validated likelihood per observation for different models. 0 suggests perfect prediction and $\ln 3$ suggests uniform random prediction

The Human-Robot dataset is the data we collected. The Human-Human dataset is the data collected by [18].

We also computed the p-values for to compare the prediction accuracy of each pair of the two models for the two datasets using the signed rank test on the difference in likelihood measurements. However, we acknowledge that the likelihoods from cross-validation are only approximately i.i.d. The p-values are reported in the format of HR/HH, where the value on the left is the p-value based on the Human-Robot dataset and the value on the right is the p-value based on the Human-Human dataset. The results are as follows:

	Lk	QLk	QLkNoZero	QLkWeighted	Baseline
Lk	x	0.705/0.282	0.748/0.333	0.739/0.545	**<0.001/<0.001**
QLK	x	x	0.826/**0.01**	0.352/0.746	**<0.001/<0.001**
QLKNoZero	x	x	x	0.739/0.798	**<0.001/<0.001**
QLKWeighted	x	x	x	x	**<0.001/<0.001**

Table 2: P-values for each pair of models

We also computed the p-values for the two datasets given the same model using the Mann-Whitney test on the difference in likelihood measurements. A p-value closer to 0 suggests that there is higher statistical significance that the HH data results in a higher likelihood when a given model is applied to it compared with the HR data.

We see that for the data obtained by making humans believe they are playing against robots, all methods show a significant difference compared to the baseline model. However, there are not significant differences among these four methods. The same conclusion is mostly true for the dataset obtained from the literature for human-human strategic games from [18], except that there is a

	Lk	QLk	QLkNoZero	QLkWeighted
p-value	0.052	0.428	0.331	0.060

Table 3: P-values for the two datasets given the same model

significance ($p < 0.05$) for the QLk v.s. QLkNoZero model. However, when combining many existing datasets from multiple papers and on specific individual datasets, [21] showed that the QLk model performs better than the Lk model.

For the human-robot strategic interaction setting, it's interesting that removing the modeling of level-0 players has little influence on the likelihood. This could suggest that there are few level-0 players in the population. [18] reached a similar conclusion on their dataset.

Introducing more parameters in the quantal best response formulation also did not result in significant differences. This suggests that simpler models are sufficient to describe the strategic behavior of the humans. We also see that there's no significant difference in performance on the two datasets for all the models ($p > 0.05$).

5 Discussion & Future Work

In this project, we applied several existing methods and extensions of existing methods to the situation where humans believe they are playing against artificial agents like robots. We investigated how well each model predicts human strategic behavior and how this compares to existing literature in human-human strategic behavior. From a machine learning perspective, there is no significant difference among the models. However, they all perform better than the baseline model. Compared to an existing dataset drawn from human-human strategic interaction behavior, these models on average better describe the behavior when humans believe they are playing against other humans. However, there is no significance among the two datasets. Although this is an interesting distinction and may suggest that there are opportunities for novel models distinctive from the existing models and variations of existing models for the case where humans believe they are interacting with robots, we recognize that participants in the Amazon Mechanical Turk are drawn from a population that may be different from participants in a university, as in the case of [18]. We think to do the most fair comparison, we could collect data of human-human strategic actions through the same means and compare the results with the case of human-robot strategic actions. We are also interested in collecting much more large-scale data for the case of human-robot strategic interaction like [21] to see if we can find significant differences in performances of different models. This can reduce the variance among the data and hopefully allow us to get a more accurate picture of how humans behave when playing against robots. We are also interested in designing different strategic games and quantify through models how participants choose actions differently when playing against another human versus a robot. We are very excited to explore these directions.

References

[1] Pieter Abbeel, Dmitri Dolgov, Andrew Y. Ng, and Sebastian Thrun. Apprenticeship learning for motion planning with application to parking lot navigation. In *Proceedings of the International Conference on Intellegent RObots and Systems (IROS)*, 2008.

[2] Pieter Abbeel and Andrew Y. Ng. Apprenticeship learning via inverse reinforcement learning. In *Proceedings of the 21st International Conference on Machine Learning (ICML)*, 2004.

[3] F. Arrichiello, P. Di Lillo, D. Di Vito, G. Antonelli, and S. Chiaverini. Assistive robot operated via p300-based brain computer interface. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6032–6037, May 2017.

[4] Colin F. Camerer, Teck-Hua Ho, and Juin-Kuan Chong. A cognitive hierarchy model of games*. *The Quarterly Journal of Economics*, 119(3):861–898, 2004.

[5] Miguel Costa-Gomes, Vincent Crawford, and Bruno Broseta. Cognition and behavior in normal-form games: An experimental study. 69:1193 – 1235, 09 2001.

[6] E. Demeester, A. Huntemann, E. Vander Poorten, and J. De Schutter. Ml, map and greedy pomdp shared control: comparison of wheelchair navigation assistance for switch interfaces. In *International Symposium on Robotics*, 2012.

[7] Shayan Doroudi, Kenneth Holstein, Vincent Aleven, and Emma Brunskill. Towards understanding how to leverage sense-making, induction/refinement and fluency to improve robust learning. In *EDM*, 2015.

[8] Google, Inc. Google assistant. *https://assistant.google.com/*, 2018.

[9] Google, Inc. Google self-driving car. *https://www.google.com/selfdrivingcar/*, 2018.

[10] D. Gopinath, S. Jain, and B. D. Argall. Human-in-the-loop optimization of shared autonomy in assistive robotics. *IEEE Robotics and Automation Letters*, 2(1):247–254, Jan 2017.

[11] Michael Hillman, Karen Hagan, Sean Hagan, Jill Jepson, and Roger Orpwood. The weston wheelchair mounted assistive robot - the design story. *Robotica*, 20(2):125–132, March 2002.

[12] Shervin Javdani, Henny Admoni, Stefania Pellegrinelli, Siddhartha S. Srinivasa, and J. Andrew Bagnell. Shared autonomy via hindsight optimization for teleoperation and teaming. *CoRR*, abs/1706.00155, 2017.

[13] Kenneth R. Koedinger, Emma Brunskill, Ryan Shaun Joazeiro de Baker, Elizabeth A. McLaughlin, and John C. Stamper. New potentials for data-driven intelligent tutoring system development and optimization. *AI Magazine*, 34:27–41, 2013.

[14] Anna N. Rafferty, Emma Brunskill, Thomas L. Griffiths, and Patrick Shafto. Faster teaching by pomdp planning. In Gautam Biswas, Susan Bull, Judy Kay, and Antonija Mitrovic, editors, *Artificial Intelligence in Education*, pages 280–287, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[15] Tobias Rehder, Wolfgang Muenst, Lawrence Louis, and Dieter Schramm. Learning lane change intentions through lane contentedness estimation from demonstrated driving. In *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*, pages 893–898. IEEE, 2016.

[16] Dorsa Sadigh, S. Shankar Sastry, Sanjit A. Seshia, and Anca D. Dragan. Planning for autonomous cars that leverage effects on human actions. In *Proceedings of Robotics: Science and Systems*, RSS '16, 2016.

[17] Herbert A. Simon. Rational decision making in business organizations. *The American Economic Review*, 69(4):493–513, 1979.

[18] Dale Stahl and Paul W. Wilson. Experimental evidence on players' models of other players. 25:309–327, 02 1994.

[19] P. Trautman. Assistive planning in complex, dynamic environments: a probabilistic approach. In *IEEE International Conference on Systems, Man and Cybernetics (to be published; preprint at http://arxiv.org/abs/1506.06784)*, 2015.

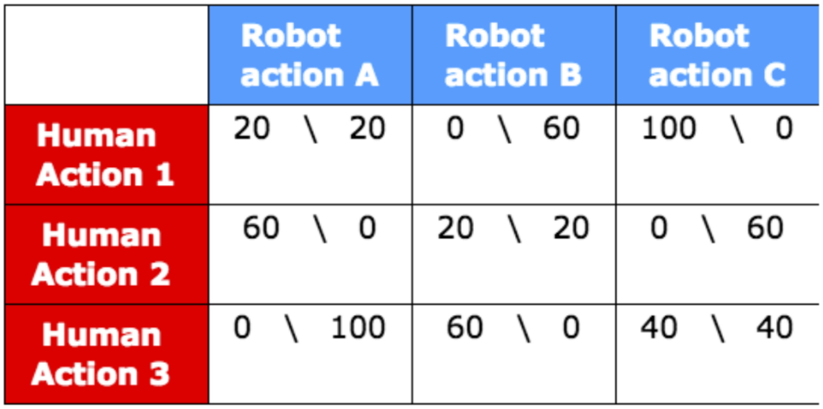
[20] James R. Wright and Kevin Leyton-brown. Beyond equilibrium: Predicting human behavior in normal-form games. In *Proceedings of the AAAI Conference on Artificial Intelligence.*, 2010.

[21] James R. Wright and Kevin Leyton-Brown. Behavioral game theoretic models: A bayesian framework for parameter analysis. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems - Volume 2*. International Foundation for Autonomous Agents and Multiagent Systems, 2012.

Appendix

Amazon Mechanical Turk interface

Below is the interface we show for each game to the participants. We collected the data by showing a game matrix to the participant and the participant can select one of the three available actions.



	Robot action A	Robot action B	Robot action C
Human Action 1	20 \ 20	0 \ 60	100 \ 0
Human Action 2	60 \ 0	20 \ 20	0 \ 60
Human Action 3	0 \ 100	60 \ 0	40 \ 40

Figure 1: Amazon Mechanical Turk interface for a sample game