Combining Prediction of Human Decisions with ISMCTS in Imperfect Information Games*

Extended Abstract

Moshe Bitan
Dept. of Computer Science, Bar-Ilan University
Ramat-Gan, Israel
moshe.bitan78@gmail.com

Sarit Kraus
Dept. of Computer Science, Bar-Ilan University
Ramat-Gan, Israel
sarit@cs.biu.ac.il

ABSTRACT

We present agents that perform well against humans in imperfect information games with partially observable actions. We introduce the Semi-Determinized-MCTS (SDMCTS), a variant of the Information Set MCTS algorithm (ISMCTS). SDMCTS generates a predictive model of the unobservable portion of the opponent's actions from historical behavioral data. Next, SDMCTS performs simulations on an instance of the game where the unobservable portion of the opponent's actions are determined. Thereby, it facilitates the use of the predictive model in order to decrease uncertainty. We present an implementation of the SDMCTS applied to the Cheat Game. Results from experiments with 120 subjects playing a head-to-head Cheat Game against our SDMCTS agents suggest that SDMCTS performs well against humans, and its performance improves as the predictive model's accuracy increases.

KEYWORDS

Agents competing against humans

ACM Reference Format:

Moshe Bitan and Sarit Kraus. 2018. Combining Prediction of Human Decisions with ISMCTS in Imperfect Information Games. In *Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), Stockholm, Sweden, July 10–15, 2018,* IFAAMAS, 3 pages.

1 INTRODUCTION

Monte Carlo Tree Search (MCTS) has had significant success in perfect information games such as Scrabble, Hex and Go [1, 10, 11, 23, 24] MCTS was adapted by researchers [4] to solve imperfect information games. One of the first popular extensions of MCTS to imperfect information games is a variation where a perfect information search is performed on a determinized instance of the game [3, 13, 23, 25]. However, several problems have been reported when applying determinization [20]. To address these problems, Information Set MCTS (ISMCTS) was developed [5, 15, 17]. ISMCTS performs simulations directly on information sets (IS). ISMCTS assumes that every game state in the IS has the same probability of being the game state, which is not always the case. Thus, it prevents the use of a predictive model. In addition, empirical studies suggest that people rarely converge to the sub-game perfect equilibrium [7, 8]. Therefore, a more integrative approach that incorporates prediction of a human opponent's decision-making may yield better game performance when interacting with humans [6, 9, 14, 18, 19, 21, 22].

We introduce the Semi-Determinized-MCTS (SDMCTS), a variant of the ISMCTS algorithm combined with a determinization technique. SDMCTS is designed to reduce uncertainty by utilizing a predictive model of the unobservable portion of the opponent's actions. SDMCTS performs simulations directly on IS. However, it performs simulations on an instance of the game where the unobservable portion of the opponent's actions is determined. This facilitates the use of the predictive model for decreasing uncertainty with regards to hidden actions. We evaluate the SDMCTS in *Cheat Game.* A behavioural-based predictive model of human player actions in the Cheat Game was used. The predictive model was trained on data collected from 60 players playing human-vs-human games reaching 0.821 Area Under the Curve (AUC). In the experiments, the SDMCTS agents played head-to-head cheat games against 120 subjects. The results suggest that SDMCTS performs well against humans, reaching a win ratio of 88.97%, and its performances improve as the predictive model's accuracy increases.

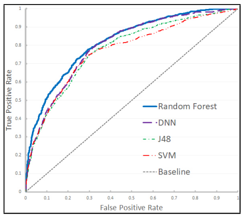
2 SEMI-DETERMINIZED MCTS

SDMCTS utilizes a predictive model of the unobservable portion of the opponent's actions. SDMCTS searches for an optimal strategy as a response to the opponent's strategies in an instance of the game where hidden actions are revealed to all players. Next, SDMCTS uses the predictive model to estimate the expected reward for each response action. SDMCTS performs MCTS simulations directly on the IS of an instance of the game where **only** the unobservable portion of the opponent's actions are determined. Formally, let u_k^i be the information state (IS) for player i at round k of the game. We denote by a_{k-1}^o the action performed by an opponent in the previous round $k-1$ which led to the IS u_k^i. Note that a_{k-1}^o is not fully observable by player i. Therefore, we define an alternative semi-determinized IS $\tilde{u}_k^i(a_{k-1}^o)$ where the opponent's previous action is determined to be a_{k-1}^o. SDMCTS performs simulations on the semi-determinized IS, resulting in estimates of $Q(\tilde{u}_k^i(a_{k-1}^o), a)$ for performing response action a at the IS u_k^i where the opponent's previous action is determined to be a_{k-1}^o.

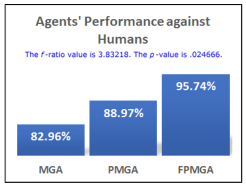
Next, the predictive model is used for calculating the expected payoff for each of the current player's response actions. The predictive model provides a probability distribution over the opponent's possible previous actions. Formally, for a given IS u^i, the predictive

*This work was supported in part by the LAW-TRAIN project that has received funding from the European Union Horizon 2020 research and innovation program under grant agreement 653587, a grant from MAAFAT and by a grant from the Ministry of Science & Technology, Israel & the Japan Scienceand Technology Agency (jst), Japan

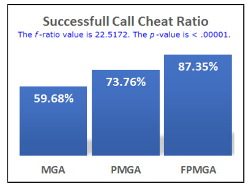
Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), M. Dastani, G. Sukthankar, E. André, S. Koenig (eds.), July 10–15, 2018, Stockholm, Sweden.



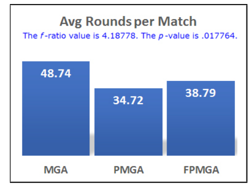
(a) ROC Curves



(b) Agents' matches win ratio



(c) Call Cheat Success Rate



(d) Matches' Avg. Rounds

Figure 1: Predictive Model & Experimental results.

model estimates $\mathcal{P}(a^o \mid u^i)$, the probability that the opponent has performed action a^o in the previous round. The expected payoff $\mathbb{E}_u[u^i, a]$ for performing action a at IS u^i is calculated by:

$$\mathbb{E}_u[u^i, a] = \sum_{a^o} \left(\mathcal{P}(a^o \mid u^i) \cdot Q(\tilde{u}^i(a^o), a) \right)$$

where $Q(\tilde{u}^i(a^o), a)$ is the estimated payoff for performing action a at the semi-determinized IS $\tilde{u}^i(a^o)$. Lastly, the algorithm chooses the response action that maximizes the expected payoff.

3 THE CHEAT GAME

The *Cheat Game* is an Imperfect Information Game with partially unobservable actions. In our version, players start with 8 cards. The objective is to be the first player to get rid of his cards. Players can place between one and four face-down cards on the table and make a claim as to what those cards' rank is. However, a player is permitted to lie about the cards' rank. The first claim is chosen as the top card of the deck; subsequent calls must be exactly one rank higher/lower. A player can avoid making a claim by *Taking a Card* from the deck. The opponent can challenge claims by performing the Call-Cheat action. If the challenge was correct, the player who is lying must take the stack that is on the table. However, if the challenger was wrong, he must take the stack. The first player to empty their hand is the winner.

3.1 The Cheat Game Agent

The *Predictive MCTS Cheat Game Agent* (PMCA) is an instantiation of the proposed method. PMCA combines a variation of the MCTS with a predictive model of human decisions. The chosen MCTS adaptation is Smooth-UCT [15]. The predictive model was developed based on a human behavioral data-set (for more details see [2]). Prior to developing the agent, a preliminary experiment was conducted with 60 participants.

3.2 The Cheat Game - Information State

An information state u^i is the visible portion of the game state s to player i. In the Cheat Game, the player is granted access to the following attributes of the game state. Clearly, the player can view his own cards and his own actions. In addition, the player can view the number of cards in the opponent's hand, the facedown table cards, and the remaining cards in the shuffled deck. Once a *Call-Cheat* action is performed, the last claim that was made is examined. An abstraction aggregates similar information states, resulting in an alternative information space with a considerably smaller size [12, 16]. Formally, an abstraction $F = \{f_A^i : \mathcal{U}^i \to \tilde{\mathcal{U}}^i | i \in \mathcal{N}\}$ is a set of functions that maps the information state space $\mathcal{U}^i$ onto an information state space $\tilde{\mathcal{U}}^i$, where $|\mathcal{U}^i| \gg |\tilde{\mathcal{U}}^i|$. It is important to choose a suitable abstraction that both reduces state space and partially preserves its strategic structure.

Attributes from the full IS u^i were selected based on their importance.The search is performed on the alternative IS where claim actions are determinized. Therefore, the alternative IS contains public information about the nature of the claim, i.e. whether the previous claim is a *true* claim or a *false* claim.

4 EXPERIMENTS AND RESULTS

120 participants from the US, aged 20-50 (46% females and 54% males), were recruited using Amazon Mechanical Turk. They were asked to play a two-player Cheat Game for three to five matches. Participants were randomly divided into three groups. Each group played against a different instantiation of the proposed method. The first group played against a *Smooth-UCT Cheat Game Agent* (*MGA*) without a predictive model. The second group played against the *Predictive Smooth-UCT Cheat Game Agent* (***PMGA***) that incorporates the predictive model. Participants in the third group played against the ***FPMGA*** agent, which had an unfair advantage. The *FPMGA* agent was allowed to peek at his opponent's real claim. FPMGA was restricted to a prediction rate of 85%. The 85% prediction rate was chosen as a plausible prediction rate that can be achieved when predicting human decisions.

One-way ANOVA was conducted to compare the three conditions ($p < .05$). The *p-value* and *f-ratio* can be observed in figures 1c. Both the PMGA and FPMGA performed significantly better than the non-predictive agent (MGA). Furthermore, FPMGA performed significantly better than PMGA. Figure 1b presents the percentage of won matches by the agent. The results suggest that by combining the predictive model with the ISMCTS, the agents were able to reduce the uncertainty that was derived from hidden actions. Thereby, the predictive agents were able to choose better response actions to deceptive claims. Figure 1d can further demonstrate the improvement of game performance. The average number of rounds it took the predictive agents to conclude a match is significantly lower than the MGA. Another measure for play-dominance is the average difference between the cards held by the agent and the humans. The PMGA's and FPMGA's average card difference was significantly lower than the MGA's (see figure 1d).

REFERENCES

[1] B Arneson, R B Hayward, and P Henderson. 2010. Monte Carlo tree search in Hex. *IEEE Transactions on Computational Intelligence and AI in Games* 2, 4 (2010), 251–258.

[2] Moshe Bitan and Sarit Kraus. 2017. Combining Prediction of Human Decisions with ISMCTS in Imperfect Information Games. *CoRR* abs/1709.09451 (2017). arXiv:1709.09451 http://arxiv.org/abs/1709.09451

[3] Y Bjornsson and H Finnsson. 2009. Cadiaplayer: A simulation-based general game player. *IEEE Transactions on Computational Intelligence and AI in Games* 1, 1 (2009), 4–15.

[4] C. B Browne, E. Powley, D. Whitehouse, S. M. Lucas, P. I. Cowling, P. Rohlfshagen, S. Tavener, D. Perez, S. Samothrakis, and S. Colton. 2012. A survey of monte carlo tree search methods. *IEEE Transactions on Computational Intelligence and AI in games* 4, 1 (2012), 1–43.

[5] Peter I Cowling, Edward J Powley, and Daniel Whitehouse. 2012. Information set monte carlo tree search. *IEEE Transactions on Computational Intelligence and AI in Games* 4, 2 (2012), 120–143.

[6] A Davidson, D Billings, J Schaeffer, and D Szafron. 2000. Improved opponent modeling in poker. In *International Conference on Artificial Intelligence, ICAI'00*. 1467–1473.

[7] Ido Erev and Alvin E Roth. 1998. Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American economic review* (1998), 848–881.

[8] Ya'akov Gal, Sarit Kraus, Michele Gelfand, Hilal Khashan, and Elizabeth Salmon. 2011. An adaptive agent for negotiating with people in different cultures. *ACM Transactions on Intelligent Systems and Technology (TIST)* 3, 1 (2011), 8.

[9] Ya'akov Gal, Avi Pfeffer, Francesca Marzo, and Barbara J Grosz. 2004. Learning social preferences in games. In *Proceedings of the National Conference on Artificial Intelligence*. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 226–231.

[10] S Gelly, L Kocsis, M Schoenauer, M Sebag, D Silver, C Szepesvári, and O Teytaud. 2012. The grand challenge of computer Go: Monte Carlo tree search and extensions. *Commun. ACM* 55, 3 (2012), 106–113.

[11] S Gelly, Y Wang, O Teytaud, M Uct Patterns, and P Tao. 2006. Modification of UCT with patterns in Monte-Carlo Go. (2006).

[12] A. Gilpin. 2009. *Algorithms for abstracting and solving imperfect information games*. Carnegie Mellon University.

[13] M L Ginsberg. 2001. GIB: Imperfect information in a computationally challenging game. *Journal of Artificial Intelligence Research* 14 (2001), 303–358.

[14] H He, J Boyd-Graber, K Kwok, and H Daumé III. 2016. Opponent Modeling in Deep Reinforcement Learning. In *Proceedings of The 33rd International Conference on Machine Learning*. 1804–1813.

[15] J Heinrich and D Silver. 2015. Smooth UCT Search in Computer Poker.. In *IJCAI*. 554–560.

[16] M B Johanson. 2016. *Robust Strategies and Counter-Strategies: From Superhuman to Optimal Play*. Ph.D. Dissertation. University of Alberta.

[17] V. Lisỳ, M. Lanctot, and M. Bowling. 2015. Online monte carlo counterfactual regret minimization for search in imperfect information games. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 27–36.

[18] S Markovitch and R Reger. 2005. Learning and exploiting relative weaknesses of opponent agents. *Autonomous Agents and Multi-Agent Systems* 10, 2 (2005), 103–130.

[19] N. Peled, M. Bitan, J. Keshet, and S. Kraus. 2013. Predicting Human Strategic Decisions Using Facial Expressions.. In *IJCAI*.

[20] M Ponsen, S De Jong, and M Lanctot. 2011. Computing approximate nash equilibria and robust best-responses using sampling. *Journal of Artificial Intelligence Research* 42 (2011), 575–605.

[21] Ariel Rosenfeld and Sarit Kraus. 2016. Providing arguments in discussions on the basis of the prediction of human argumentative behavior. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 6, 4 (2016), 30.

[22] A Rosenfeld, I Zuckerman, A Azaria, and S Kraus. 2012. Combining psychological models with machine learning to better predict people's decisions. *Synthese* 189, 1 (2012), 81–93.

[23] B Sheppard. 2002. World-championship-caliber Scrabble. *Artificial Intelligence* 134, 1-2 (2002), 241–275.

[24] D Silver, A Huang, C J Maddison, A Guez, L Sifre, G Van Den Driessche, J Schrittwieser, I Antonoglou, V Panneershelvam, M Lanctot, et al. 2016. Mastering the game of Go with deep neural networks and tree search. *Nature* 529, 7587 (2016), 484–489.

[25] N R Sturtevant. 2008. An analysis of UCT in multi-player games. In *International Conference on Computers and Games*. Springer, 37–49.