

Effective Short-Term Opponent Exploitation in Simplified Poker

Bret Hoehn
Finnegan Southey
Robert C. Holte

University of Alberta, Dept. of Computing Science

Valeriy Bulitko

Centre for Science, Athabasca University

Abstract

Uncertainty in poker stems from two key sources, the shuffled deck and an adversary whose strategy is unknown. One approach is to find a pessimistic game theoretic solution (i.e. a Nash equilibrium), but human players have idiosyncratic weaknesses that can be exploited if a model of their strategy can be learned by observing their play. However, games against humans last for at most a few hundred hands so learning must be fast to be effective. We explore two approaches to opponent modelling in the context of Kuhn poker, a small game for which game theoretic solutions are known. Parameter estimation and expert algorithms are both studied. Experiments demonstrate that, even in this small game, convergence to maximally exploitive solutions in a small number of hands is impractical, but that good (i.e. better than Nash or breakeven) performance can be achieved in a short period of time. Finally, we show that amongst a set of strategies with equal game theoretic value, in particular the set of Nash equilibrium strategies, some are preferable because they speed learning of the opponent's strategy by exploring it more effectively.

Introduction

Poker is a game of imperfect information against an adversary with an unknown, stochastic strategy. It represents a tough challenge to artificial intelligence research. Game theoretic approaches seek to approximate the Nash equilibrium (i.e. minimax) strategies of the game (Koller & Pfeffer 1997; Billings *et al.* 2003), but this represents a pessimistic worldview where we assume optimality in our opponent. Human players have weaknesses that can be exploited to obtain winnings higher than the game-theoretic value of the game. Learning by observing their play allows us to exploit their idiosyncratic weaknesses. This can be done either directly, by learning a model of their strategy, or indirectly, by identifying an effective counter-strategy. Several factors render this difficult in practice. First, real-world poker games like Texas Hold'em have huge

game trees and the strategies involve many parameters (e.g. two-player, limit Texas Hold'em requires $O(10^{18})$ parameters (Billings *et al.* 2003)). The game also has high variance, stemming from the deck and stochastic opponents, and folding gives rise to partial observations. Strategically complex, the aim is not simply to win but to maximize winnings by enticing a weakly-positioned opponent to bet. Finally, we cannot expect a large amount of data when playing human opponents. You may play only 50 or 100 hands against a given opponent and want to quickly learn how to exploit them.

This research explores how rapidly we can gain an advantage by observing opponent play given that only a small number of hands will be played in total. Two learning approaches are studied: *maximum a posteriori parameter estimation (parameter learning)*, and an "experts" method derived from Exp3 (Auer *et al.* 1995) (*strategy learning*). Both will be described in detail.

While existing poker opponent modelling research focuses on real-world games (Korb & Nicholson 1999; Billings *et al.*), we systematically study a simpler version, reducing the game's intrinsic difficulty to show that, even in what might be considered a best case, the problem is still hard. We start by assuming that the opponent's strategy is fixed. Tracking a non-stationary strategy is a hard problem and learning to exploit a fixed strategy is clearly the first step. Next, we consider the game of Kuhn poker (Kuhn 1950), a tiny game for which complete game theoretic analysis is available. Finally, we evaluate learning in a two-phase manner; the first phase exploring and learning, while the second phase switches to pure exploitation based on what was learned. We use this simplified framework to show that learning to maximally exploit an opponent in a small number of hands is not feasible. However, we also demonstrate that some advantage can be rapidly attained, making short-term learning a winning proposition. Finally, we observe that, amongst the set of Nash strategies for the learner (which are "safe" strategies), the exploration inherent in some strategies facilitates faster learning compared with other members of the set.

Copyright © 2005, American Association for Artificial Intelligence (www.aaai.org). All rights reserved.

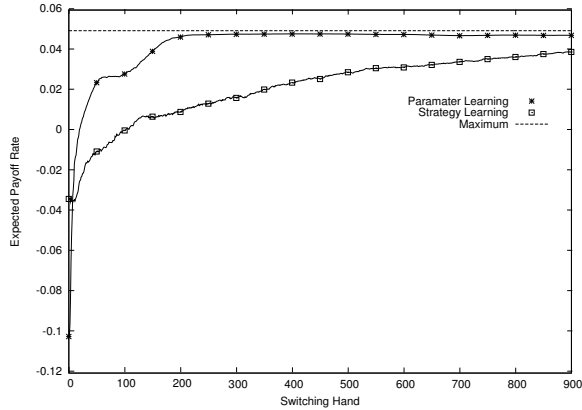


Figure 3: Convergence Study: Expected payoff rate vs. switching hand for parameter and strategy learning

Convergence Rate Study

Figure 3 shows the expected payoff rate plot of the two learning methods against a single opponent. The straight-line near the top shows the maximum exploitation rate for this opponent (i.e. the value of the best response to P2's strategy). It takes 200 hands for parameter learning to almost converge to the maximum and strategy learning does not converge within 900 hands. Results for other opponents are generally worse, requiring several hundred hands for near-convergence. This shows that, even in this tiny game, one cannot expect to achieve maximal exploitation in a small number of hands. The possibility of maximal exploitation in larger games can reasonably be ruled out on this basis and we must adopt more modest goals for opponent modellers.

Game Length Study

This study is provided to show that our total winnings results are robust to games of varying length. While most of our results are presented for games of 200 hands, it is only natural to question whether different numbers of hands would have different optimal switching points. Figure 4 shows overlaid total winnings plots for 50, 100, 200, and 400 hands using parameter learning. The lines are separated because the possible total winnings is different for differing numbers of hands. The important observation to make is that the highest value regions of these curves are fairly broad, indicating that switching times are flexible. Moreover, the regions of the various curves overlap substantially. Thus, switching at hand 50 is a reasonable choice for all of these game lengths, offering close to the best possible total winnings in all cases. This means that even if we are unsure, *a priori*, of the number of hands to be played, we can be confident in our choice of switching time. Moreover, this result is robust across our range of opponents. A switch at hand 50 works well in all cases.

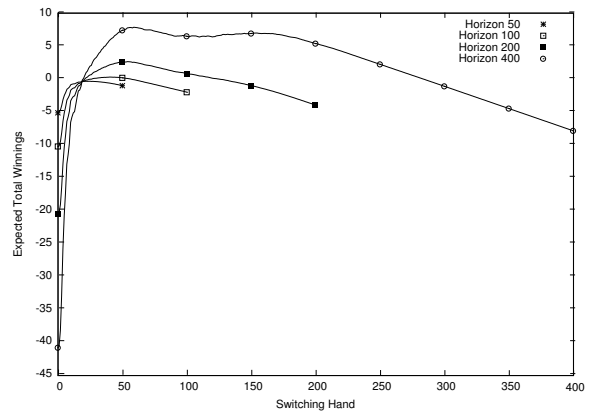


Figure 4: Game Length Study: Expected total winnings vs. switching hand for game lengths of 50, 100, 200, and 400 hands played by parameter learning

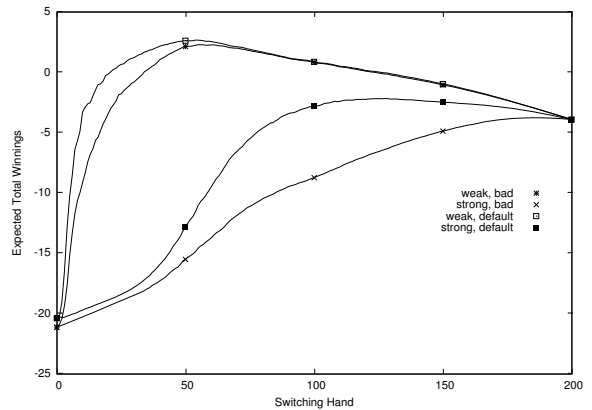


Figure 5: Prior Study: Four different priors for parameter learning against a single opponent.

Parameter Learning Prior Study

In any Bayesian parameter estimation approach, the choice of prior is clearly important. Here we present a comparison of various priors against a single opponent ($O_6 = (.17, .2)$). Expected total winnings are shown for four priors: a weak, default prior of $(.5, .5)$, a weak, bad prior of $(.7, .5)$, a strong, default prior of $(.5, .5)$, and a strong, bad prior of $(.7, .5)$. The weak priors assume 2 fictitious points have been observed and the strong priors assume 20 points. The “bad” prior is so called because it is quite distant from the real strategy of this opponent. Figure 5 shows that the weak priors clearly do better than the strong, allowing for fast adaptation to the correct opponent model. The strong priors perform much more poorly, especially the strong bad prior.

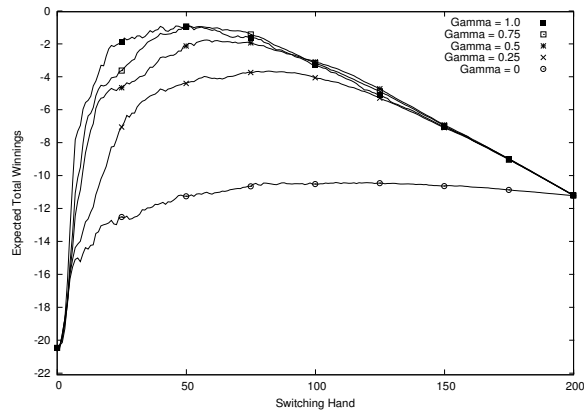


Figure 6: Nash Study: Expected total winnings vs. switching hand for parameter learning with various Nash strategies used during the learning phase.

Nash Exploration Study

Figure 6 shows the expected total winnings for parameter learning when various Nash strategies are played by the learner during the learning phase. The strategies with larger γ values are clearly stronger, more effectively exploring the opponent's strategy during the learning phase. This advantage is typical of Nash strategies with $\gamma > 0.7$ across all opponents we tried.

Learning Method Comparison

Figure 7 directly compares strategy and parameter learning (both balanced and Nash exploration ($\gamma = 1$)), all against a single opponent. Balanced parameter learning outperforms strategy learning substantially for this opponent. Over all opponents, either the balanced or the Nash parameter learner is the best, and strategy learning is worst in all but one case.

Conclusions

This work shows that learning to maximally exploit an opponent, even a stationary one in a game as small as Kuhn poker, is not generally feasible in a small number of hands. However, the learning methods explored are capable of showing positive results in as few as 50 hands, so that learning to exploit is typically better than adopting a pessimistic Nash equilibrium strategy. Furthermore, this 50 hand switching point is robust to game length and opponent. Future work includes non-stationary opponents, a wider exploration of learning strategies, and larger games. Both approaches can scale up, provided the number of parameters or experts is kept small (abstraction can reduce parameters and small sets of experts can be carefully selected). Also, the exploration differences amongst equal valued strategies (e.g. Nash) deserves more attention. It may be pos-

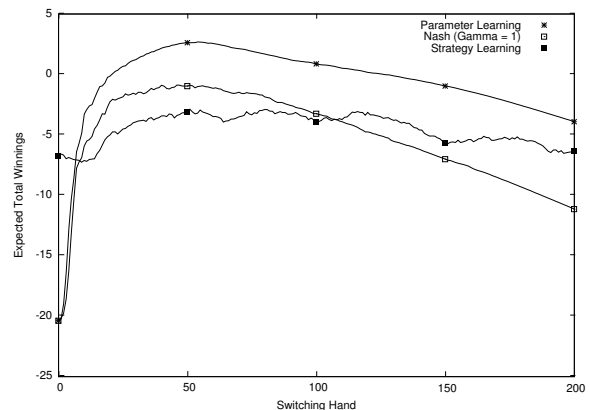


Figure 7: Learning Method Comparison: Expected total winnings vs. switching hand for both parameter learning and strategy learning against a single opponent.

sible to more formally characterize the exploratory effectiveness of a strategy. We believe these results should encourage more opponent modelling research because, even though maximal exploitation is unlikely, fast opponent modelling may still yield significant benefits.

Acknowledgements

Thanks to the Natural Sciences and Engineering Research Council of Canada and the Alberta Ingenuity Centre for Machine Learning for project funding, and the University of Alberta poker group for their insights.

References

- Auer, P.; Cesa-Bianchi, N.; Freund, Y.; and Schapire, R. E. 1995. Gambling in a rigged casino: the adversarial multi-armed bandit problem. In *Proc. of the 36th Annual Symp. on Foundations of Comp. Sci.*, 322–331.
- Billings, D.; Davidson, A.; Schauenberg, T.; Burch, N.; Bowling, M.; Holte, R.; Schaeffer, J.; and Szafron, D. Game Tree Search with Adaptation in Stochastic Imperfect Information Games. In *Computers and Games'04*.
- Billings, D.; Burch, N.; Davidson, A.; Holte, R.; Schaeffer, J.; Schauenberg, T.; and Szafron, D. 2003. Approximating game-theoretic optimal strategies for full-scale poker. In *18th Intl. Joint Conf. on Artificial Intelligence (IJCAI'2003)*.
- Koller, D., and Pfeffer, A. 1997. Representations and solutions for game-theoretic problems. *Artificial Intelligence* 94(1):167–215.
- Korb, K., and Nicholson, A. 1999. Bayesian poker. In *Uncertainty in Artificial Intelligence*, 343–350.
- Kuhn, H. W. 1950. A simplified two-person poker. *Contributions to the Theory of Games* 1:97–103.