Regret Minimization in Multiplayer Extensive Games*

Richard Gibson and **Duane Szafron**
Department of Computing Science, University of Alberta
Edmonton, Alberta, T6G 2E8, Canada
{rggibson | duane}@cs.ualberta.ca

Abstract

The counterfactual regret minimization (CFR) algorithm is state-of-the-art for computing strategies in large games and other sequential decision-making problems. Little is known, however, about CFR in games with more than 2 players. This extended abstract outlines research towards a better understanding of CFR in multiplayer games and new procedures for computing even stronger multiplayer strategies. We summarize work already completed that investigates techniques for creating "expert" strategies for playing smaller sub-games, and work that proves CFR avoids classes of undesirable strategies. In addition, we provide an outline of our future research direction. Our goals are to apply regret minimization to the problem of playing multiple games simultaneously, and augment CFR to achieve effective on-line opponent modelling of multiple opponents. The objective of this research is to build a world-class computer poker player for multiplayer Limit Texas Hold'em.

1 Introduction

An *extensive form game* [Osborne and Rubenstein, 1994] is a rooted directed tree, where nodes represent decision states, edges represent actions, and terminal nodes hold end-game utility values. For each player i, the decision states are partitioned into information sets $\mathcal{I}_i$ such that game states within an information set are indistinguishable to player i. For example, poker can be modelled as an extensive game where $\mathcal{I}_i$ disguises the private cards held by the opponent(s). Extensive games are very versatile due to their ability to represent multiple agents, imperfect information, and stochastic events.

A *strategy* σ_i for player i in an extensive game is a mapping from the set of information sets for player i to a probability distribution over actions. A *strategy profile* in an n-player game is a vector of strategies $\sigma = (\sigma_1, ..., \sigma_n)$, one for each player. We denote $u_i(\sigma)$ to be the expected utility for player i given that all players play according to the strategy profile σ.

*This research is supported by NSERC and Alberta Ingenuity, now part of Alberta Innovates - Technology Futures.

Counterfactual regret minimization (CFR) [Zinkevich *et al.*, 2008] is an iterative procedure for computing a strategy profile in an extensive game. The algorithm constructs a sequence of profiles $(\sigma^1, \sigma^2, ...)$ that minimize the counterfactual regret, $R_i^T(I, a)$, at every information set I and action a at I. In a nutshell, $R_i^T(I, a)$ tells us how much player i would rather always play action a than follow σ_i^t at I at times $t = 1..T$. The output of CFR after T iterations is the average of the sequence of profiles, $\bar{\sigma}^T$, and is an approximate Nash equilibrium profile in 2-player zero-sum games. Though CFR strategies have also been found to compete very well in multiplayer (more than 2 players) settings [Abou Risk and Szafron, 2010], very little is known as to why they do well and how performance can be improved.

2 Strategy Stitching

For many real-world problems, the extensive game representation is too large to feasibly apply CFR. To address this limitation, strategies are often computed in abstract versions of the game that group similar states together into single abstract states. For example, in poker, a common approach is to group many different card dealings into single abstract states according to some similarity metric. For very large games, these abstractions need to be quite coarse, leaving many different states indistinguishable. However, for smaller sub-trees of the full game, strategies can be computed in much finer abstractions. Such "expert" strategies can then be pieced together, typically connecting to a "base strategy" computed in the full coarsely-abstracted game.

We have investigated stitched strategies in extensive games, focusing on the trade-offs between the sizes of the abstractions versus the assumptions made by the experts and the cohesion among the computed strategies when stitched together. We defined two strategy stitching techniques: (i) *static* experts that are computed in very fine abstractions with varying degrees of assumptions and little cohesion, and (ii) *dynamic* experts that are contained in abstractions with lower granularity, but make fewer assumptions and have perfect cohesion. We generalized previous strategy stitching efforts [Billings *et al.*, 2003; Waugh *et al.*, 2009; Abou Risk and Szafron, 2010] under a more general static expert framework. In poker, we found that experts can create much stronger overall agents than the base strategy alone. Furthermore, under a fixed memory limitation, a specific class of static ex-

perts were preferred because of the increase in granularity of abstraction allowed by the static approach. As a final validation of our results, we built two 3-player Limit Texas Hold'em agents with static experts and entered them into the 2010 Annual Computer Poker Competition.[1] Our agents won the 3-player events by a significant margin.

3 Domination

A *pure strategy* s_i for player i assigns a probability of 1 to a single action at each information set I; denote this action by $s_i(I)$. A pure strategy s_i is *strictly dominated* if there exists another strategy σ'_i such that $u_i(s_i, \sigma_{-i}) < u_i(\sigma'_i, \sigma_{-i})$ for all opponent strategies σ_{-i}. In addition, a pure strategy s_i is recursively defined to be *iteratively strictly dominated* if either s_i is strictly dominated, or if there exists another strategy σ'_i such that $u_i(s_i, \sigma_{-i}) < u_i(\sigma'_i, \sigma_{-i})$ for all non-iteratively dominated opponent strategies σ_{-i}.

One should never play a dominated strategy. Also, if we assume our opponents are rational and will not play dominated strategies, one should also avoid iteratively dominated strategies. We have extended the notion of dominance to actions at information sets: We say that a is a *strictly dominated action at information set* I if there exists another action a' at I such that $v_i(\sigma_{(I \to a)}, I) < v_i(\sigma_{(I \to a')}, I)$ for all strategy profiles σ. Here, $v_i(\sigma, I)$ is the *counterfactual value* of σ at I as defined by Lanctot *et al.* [2009, Eq. (4)], and $\sigma_{(I \to a)}$ is the profile σ except at I, action a is always taken. Iteratively strictly dominated actions are defined analogously.

We have two main results regarding dominance in CFR. The first proves that if the opponents continue to reach an information set I with positive probability, then eventually the probability of playing a strictly dominated action at I becomes zero.[2] This implies that $\bar{\sigma}^T$ plays iteratively strictly dominated actions with vanishing probability. Our second result shows that if s_i is a strictly dominated strategy, then eventually the regret $R_i^T(I, s(I))$ for action $s_i(I)$ must be negative at some information set I. Future work will look into measuring how quickly the dominated elements are removed from $\bar{\sigma}^T$ and strengthening these results.

4 Simultaneous Game Playing

Strategy stitching allows us to employ finer abstractions to sub-games. For very large games, however, it is not feasible to build expert strategies for every sub-game. Currently, we play in a coarsely abstracted base game for much of the tree, where granularity is restricted by our resource limitations.

Our goal is to develop a regret minimization procedure that produces a strong strategy for playing multiple (abstract) games simultaneously. The motivation for this comes from overlapping tilings in reinforcement learning [Sutton and Barto, 1998, Figure 8.5]. By coarsely abstracting the space in multiple ways, we have fewer total information sets than if we considered the single "product" abstraction. Consequently, less memory and less time are required to run a CFR-type algorithm. The goal is to find strategies that perform well in each of the individual abstractions and improve play in the full game.

[1] http://www.computerpokercompetition.org

[2] A minor but straightforward modification to CFR is required.

5 On-line Opponent Modelling

Finally, in repeated games, monitoring opponent behaviors and exploiting them is a challenging but important task. In multiplayer settings, following a single static strategy can be problematic; even playing a Nash equilibrium does not provide a worst case guarantee. While there are some positive results in 2-player games [Davidson, 2002; Lockett and Miikkulainen, 2008], little has been achieved with modelling multiple opponents on-line.

Our goal is to apply an augmentation of CFR on-line that will perturb our strategy according to the opponents' play. Since hidden information is not always revealed after each repetition, we plan to model opponent behavior based only on their public actions and restricting the opponents' strategy space accordingly. By leaving our strategy space unrestricted, the goal is to update our strategy to better exploit these restricted opponents. Opponent modelling is essential to our primary objective: Building a world-class poker player for multiplayer Limit Texas Hold'em.

References

[Abou Risk and Szafron, 2010] N. Abou Risk and D. Szafron. Using counterfactual regret minimization to create competitive multiplayer poker agents. In *AAMAS*, pages 159–166, 2010.

[Billings *et al.*, 2003] D. Billings, N. Burch, A. Davidson, R. Holte, J. Schaeffer, T. Schauenberg, and D. Szafron. Approximating game-theoretic optimal strategies for full-scale poker. In *IJCAI*, pages 661–668, 2003.

[Davidson, 2002] A. Davidson. Opponent modeling in poker: Learning and acting in a hostile and uncertain environment. Master's thesis, University of Alberta, 2002.

[Lanctot *et al.*, 2009] M. Lanctot, K. Waugh, M. Zinkevich, and M. Bowling. Monte carlo sampling for regret minimization in extensive games. In *NIPS-22*, pages 1078–1086, 2009.

[Lockett and Miikkulainen, 2008] A.J. Lockett and R. Miikkulainen. Evolving opponent models for Texas Hold'em. In *CIG*, 2008.

[Osborne and Rubenstein, 1994] M. Osborne and A. Rubenstein. *A Course in Game Theory*. The MIT Press, Cambridge, Massachusetts, 1994.

[Sutton and Barto, 1998] R.S. Sutton and A.G. Barto. *Reinforcement Learning: An Introduction*. The MIT Press, Cambridge, Massachusetts, 1998.

[Waugh *et al.*, 2009] K. Waugh, M. Bowling, and N. Bard. Strategy grafting in extensive games. In *NIPS-22*, pages 2026–2034, 2009.

[Zinkevich *et al.*, 2008] M. Zinkevich, M. Johanson, M. Bowling, and C. Piccione. Regret minimization in games with incomplete information. In *NIPS-20*, pages 905–912, 2008.