An Extension of Counterfactual Regret Minimization for Multi-player Card Games

Cao Yu[1] Kaneko[2]

,

1. Introduction

Counterfactual Regret Minimization (CFR)[3] is one of state-of-the-art methods for solving large imperfect-information games. It shows great performance in solving 1-to-1 poker games[5]. And there is still little research about how to apply it to multi-player poker games. In this paper, we will apply CFR to an extension of poker which is played by 4 players(2-to-2), and compare its performance with random policy and human player policy.

2. Background

2.1 Extensive Games

An extensive-form game[2] is a kind of model of sequential decision-making multi-player imperfect information games. It is composed of:

- N: A finite set N of players.
- c: a chance player.
- $-i$: players except player i.
- H: a sequence of actions that were played, Z is a set of terminal histories, $Z \subset H$.
- $A(h)$: a set of available actions after history h if $h \notin Z$.
- $P(h)$: player function that assigns a player p to take an action after history h, $p \in N \bigcup \{c\}$.
- $\mathcal{I}_i$: information partition of player i, which is denoted by $\{h \in H : P(h) = i\}$. If h and h' are in the same member of an information partition, $A(h) = A(h')$ will be satisfied whenever h and h'. I_i is an information set for player i, $I_i \in \mathcal{I}_i$.
- $f_c(a|I)$: probability that a occurs when given information set I.
- u: a utility function that assigns a reward to each player when a terminal state z is reached.

2.2 An Extension of Poker(AEOP)

The game uses a set of poker except Jokers, which is 2-to-2 game, each player has 13 cards per round. The playing rule is just like Dou Di Zhu[4], but we replace winning rule by setting that the first player who plays out cards in his hand gets 20 scores, the second player gets 15 scores, the third player gets 10 scores, and the fourth player gets nothing. And players who sit face to face belong to same team. Team score is the sum of scores of all players in the team. When a round finishes, the winner is determined by team scores, the team with the highest score wins.

3. Previous Research

3.1 CFR

Counterfactual Regret Minimization[3] is one of efficient methods for solving imperfect-information games. It computes regret value which is the difference between the action chosen and other actions in action profile to show how regretful if it chose another action. CFR can reach Nash Equilibrium by tracking past plays' counterfactual regrets, making its strategies just as following:

$$\sigma_i^{T+1}(I,a) = \begin{cases} \frac{R_i^{T,+}(I,a)}{\sum_{a\in A(I)} R_i^{T,+}(I,a)} & \text{if } \sum_{a\in A(I)} R_i^{T,+}(I,a) > 0 \\ \frac{1}{|A(I)|} & \text{otherwise.} \end{cases} \tag{1}$$

In this equation, $R_i^{T,+}(I,a)$ denotes non-negative counterfactual regret value of player i util turn T for choosing action a after getting information set I. Player i's strategy in turn $T+1$ is proportional to positive cumulative regrets of past plays if the cumulative regret is positive. If it is not, a uniform random strategy is used.

3.2 MCCFR

It is unfeasible to solve large incomplete information games by CFR because of its large cost in traversing large information set. An extension of CFR,Monte Carlo Counterfactual Regret Minimization[2], is proposed by Marc Lanctot and Kevin Waugh etc to reduce time cost of traversing the game tree on each iteration. On each iteration only some of terminal histories will be considered to update counterfactual value. There are many sampling methods, such as outcome-sampling, external sampling and average strategy sampling.

[1] souyu@g.ecc.u-tokyo.ac.jp

[2] Tokyo University, Japan

3.3 Pluribus

Pluribus[1] is an efficient algorithm shown stronger than top human professional players in multi-player card games. It makes abstraction by removing some actions from consideration and bucketing similar decision states into one state. And it trains AI offline by self-play with Monte Carlo Counterfactual Regret Minimization, in which AI starts by playing randomly, and improves itself by beating previous versions it took. When playing with human player, real-time search is taken to adjust strategies, in which AI looks some moves ahead at a leaf node in a limited depth unless it reaches terminal states to estimate expected utility value at the node. It supposes that opponents takes k different strategies according to their bias. A more balanced strategy can be found by this method because choosing an unbalanced strategy will be more likely to lose when opponent choose a strategy that just dominates it. Pluribuss success shows that it is possible to produce superhuman strategies for large multi-player imperfect-information games with well-designed algorithm, even though there is still not theorerical gaurantees that multi-player games can be well solved.

4. Method

In this paper, we will use vector sets of strategies, information partitions, actions and utilities of members in the same team to represent each teams information, so that we can take it as a two-player zero-sum game. Hence, an extension of Counterfactual Regret Minimization (CFR) can be adopted to compute Nash equilibrium. And abstraction method and depth-limited search in Pluribus can be used to reduce time cost.

References

[1] Noam Brown and Tuomas Sandholm. “Superhuman AI for multiplayer poker”. In: *Science* (2019), eaay2400.

[2] Marc Lanctot et al. “Monte Carlo sampling for regret minimization in extensive games”. In: *Advances in neural information processing systems*. 2009, pp. 1078–1086.

[3] Todd W. Neller and Marc Lanctot. “An Introduction to Counterfactual Regret Minimization”. In: *Proceedings of Model AI Assignments, The Fourth Symposium on Educational Advances in Artificial Intelligence (EAAI-2013)*. 2013.

[4] Wikipedia. *Dou dizhu*. `https://en.wikipedia.org/wiki/Dou_dizhu`.

[5] Martin Zinkevich et al. “Regret minimization in games with incomplete information”. In: *Advances in neural information processing systems*. 2008, pp. 1729–1736.