Notes on Regret Minimization in Games with Incomplete Information

Rom Parnichkun

April 22, 2022

1 Review Regret Matching

Algorithm

1. Compute the policy from the regret.
2. Play the action.
3. Compute regrets and add it to the cumulative regrets

2 Counterfactual Regret Minimization

Theoretically, regret matching can be utilized for games that require historical context by simply creating a state for every historical pattern. However, the size of the state space may be prohibitively large. As the name implies, counterfactual regret minimization is a modularization of regret matching by individually minimizing the counterfactual regrets.

Keywords

- **Information set**: Are set of game states that the controlling player cannot distinguish and so must choose actions for such states with the same distribution.
- **Strategy**: A strategy of player i σ_i is a function that assigns a distribution over $A(I_i)$ to each $I_i \in \mathcal{I}_i$. $\Sigma_i = \{\sigma(I_i) : I_i \in \mathcal{I}_i\}$ refers to player i's set of strategies.
- **Strategy Profile**: A strategy profile σ is the set of every player's strategy $\sigma_1, \sigma_2, \ldots$. We denote σ_{-i} as all the strategies in σ except σ_i.

A finite extensive game with imperfect information has the following components

- A finite set N of **players**.
- A finite set H of sequences, which represent the possible histories of actions, such that the empty sequence is in H, and every prefix of a sequence in H is also in H. $Z \subseteq H$ are the terminal histories (those which are not a prefix of any other sequences). $A(h) = a : (h, a) \in H$ are the actions avaialble after a nonterminal history $h \in H$.
- A function P that assigns to each nonterminal history a member of $N \cup \{c\}$. P is the **player function**. $P(h)$ is the player who takes an action after the history h. If $P(h) = c$ then chance determines the action taken after history h.
- A function f_c that (associatees with every history h for which $P(h) = c$) a probability measure $f_c(\cdot \mid h)$ on $A(h)$, where each probability measure is independent of every other such measure.
- For each player $i \in N$ an **information partition** $\mathcal{I}_i$ is a set of **information sets** I_i. In which $A(h) = A(h')$ whenever both $h, h' \in I_i$. Therefore $A(h) = A(I_i)$ and $P(h) = P(I_i)$ for all $h \in I_i$.
- For each player $i \in N$ a utility function u_i maps each terminal state Z to $\mathbb{R}$. If $N = \{1, 2\}$ and $u_1 = -u_2$, it is a **zero-sum extensive game**. We additionally define $\Delta_{u,i} = \max_z u_i(z) - \min_z u_i(z)$ as the range of utilities.

Let $\pi^\sigma(h)$ be the probability of history h occuring if players choose actions according to σ. We can decompose $\pi^\sigma = \prod_{i \in N \cup \{c\}} \pi_i^\sigma(h)$. Hence $\pi_i^\sigma(h)$ is the probability if player i picks all the actions in h. We denote $\pi_{-i}^\sigma(h)$ be the product of all player's contribution (including chance c) except player i. $\pi^\sigma(I) = \sum_{h \in I} \pi^\sigma(h)$ is the probability of reaching a particular information set I given σ.

The overall value to player i of a strategy profile is formulated as $u_i(\sigma) = \sum_{h \in Z} u_i(h) \pi^\sigma(h)$.

The **average overall regret** of player i at time T is:

$$R_i^T = \frac{1}{T} \max_{\sigma_i^* \in \Sigma_i} \sum_{t=1}^{T} (u_i(\sigma_i^*, \sigma_{-i}^t) - u_i(\sigma^t)). \tag{1}$$

The **immediate counterfactual regret** is:

$$R_{i,imm}^T(I) = \frac{1}{T} \max_{a \in A(I)} \sum_{t=1}^{T} \pi_{-i}^{\sigma^t}(I)(u_i(\sigma^t|_{I \to a}, I) - u_i(\sigma^t, I)). \tag{2}$$

Here, $\sigma^t|_{I \to a}$ is the strategy profile identical to σ except that player i always chooses action a when in information set I, $u(\sigma, I)$ is the expected utility given that information set I is reached formulated as follows.

$$u_i(\sigma, I) = \frac{\sum_{h \in I, h' \in Z} \pi_{-i}^\sigma(h) \pi^\sigma(h, h') u_i(h')}{\pi_{-i}^\sigma(I)}. \tag{3}$$

To minimize the counterfactual regret, we define

$$R_{i,imm}^T(I, a) = \frac{1}{T} \sum_{t=1}^{T} \pi_{-i}^{\sigma^t}(I)(u_i(\sigma^t|_{I \to a}, I) - u_i(\sigma^t, I)), \tag{4}$$

and $R_i^{T,+}(I, a) = \max(R_i^T(I, a), 0)$. Then using the Blackwell's algorithm for approachability, the strategy for time $T + 1$ is:

$$\sigma_i^{T+1}(I)(a) = \begin{cases} \frac{R_i^{T,+}(I,a)}{\sum_{a \in A(I)} R_i^{T,+}(I,a)} & \text{if} \sum_{a \in A(I)} R_i^{T,+}(I, a) > 0 \\ \frac{1}{|A(I)|} & \text{otherwise} \end{cases} \tag{5}$$