Counterfactual Regret Minimization

CFR with Khun Poker

Introduction

Regret minimization is a theoretical technique utilized in game theory where the performance of an arbitrary algorithm $\mathcal{H}$, is compared against another theoretical algorithm $\mathcal{G}$. Regret is then defined as:

$$\mathcal{R} = \mathcal{H} - \mathcal{G}$$

The algorithm $\mathcal{H}$ is defined in this context as the comparison class.

The concept of regret is broad, and there exists a multitude of valid regret mathing algorithms, although in this paper, I will be discussing one algorithm in particular–namely: counterfactual regret.

Backround

To explain how counterfactual regret minimization works, first we must define a set of variables to describe an extensive form game played by two decision making players and one chance player.

- We first define the set $\mathcal{N} = \{1, 2, c\}$ which represents the players in our extensive form game. Here, players 1 and 2 are players that can change and update their strategies. Player c represents chance, a player who plays a behaviour strategy, a static strategy that is known by players 1 and 2. Each player has perfect recall, meaning at every stage of the game, each player remembers the actions they took previously to get to that point.
- We now define histories. Let the set H represent all possible sequence of actions (including the empty sequence) the players of a game can make to reach any node in the game tree. $h \in H$ is defined as a history, which consists of a sequence of actions to reach a node in the game tree. $Z \subset H$ represents the set of all possible terminal histories. Terminal histories are a sequence of actions that end on a leaf node in the game tree.

We say that $h \sqsubseteq h'$ if from history h we can reach h'. Then we can say that for all $h \in H, z \in Z \quad h \sqsubseteq z$ [@lanctot2009monte].

- An information set is defined as the sequence of actions a player has knowledge of at any node in the game tree. $\mathcal{I}_i$ represents the set of information sets for player i, for each node in the game tree. Then, $I_i(h) \in \mathcal{I}_i$ represents the information set given to player i at history h. $A(I_i)$ represents the set of actions player i can make at information set I [@lanctot2009monte].
- A strategy $\sigma_i(I)$, is defined as the mixed strategy of player i at information set I. Where $\sigma_i(I, a)$ is the probability that player i plays action a at information set I.
- $\pi_i^\sigma(h)$ represents the probability that player i reaches history h while playing the strategy σ, independent of the actions taken by other players. Then we can define $\pi^\sigma(h) = \prod_{i \in \mathcal{N}} \pi_i^\sigma(h)$, which represents the probability, given the strategy of each player, we reach a history h. $\pi_{-i}^\sigma(h)$ represents the probability of all players excluding i reach history h. Finally, we define $\pi^\sigma(h, z) = \pi^\sigma(z)/\pi^\sigma(h)$, which represents the probability that from history h, we reach terminal history z.

Immediate Counterfactual Regret

Immediate counterfactual regret is the technique that was used to converge mixed strategies of extensive form, 2 player games, to an ϵ Nash equilibrium. To explain how it works, first we need to define the counterfactual value:

$$v_i^{\sigma^t}(I) = \sum_{h \in I, z : h \sqsubseteq z} \pi_{-i}^{\sigma^t}(h) \pi^{\sigma^t}(h, z) u_i(z)$$

At each time step t, each player (excluding chance) is given a new strategy σ^t. These strategies are used to compute the value of each information set by the chance we reach the set, multiplied by the probabilty we reach each terminal node from the current history, multiplied by the value we get from that terminal node. Here, $u_i(z)$ represents the utility we get from the terminal node z (Lanctot, Waugh, and Bowling 2009).

Next, we define the counterfactual value for each action a in an information set I such that:

$$v_i^{\sigma^t}(I, a) = \sum_{h \in I, z : h \sqsubseteq z} \pi_{-i}^{\sigma^t}(ha) \pi^{\sigma^t}(ha, z) u_i(z)$$

Finally, Immediate counter factual regret is defined as:

$$R_{i,imm}^T(I, a) = \frac{1}{T} \sum_{t=1}^{T} (v_i^{\sigma^t}(I, a) - v_i^{\sigma^t}(I))$$

Regret Matching

Regret matching is the algorithm that is used to compute σ_i^t. That is, at every time step T:

$$\sigma_i^T(I,a) = \frac{R_{i,imm}^{T,+}(I,a)}{\sum_{a \in A(I)} R_{i,imm}^{T,+}(I,a)},$$
$$R_{i,imm}^{T,+}(I,a) = max(0, R_{i,imm}^{T}(I,a))$$

The literature details that by using our regret matching algorithm for both players, eventually, through repeated traversal of our game tree traversals,

$$\lim_{T \to \infty} \sum_{I \in \mathcal{J}} R_{i,imm}^{T} = 0$$

When the average regret is 0, our strategy is said to be a Nash equilibrium. To explain why this is true, we first need to explain why regret matching is able to converge average regret to 0. This simple fact is a result of Blackwell's Approachability theorem, a derivation of Jon Von Neumann's minimax theorem.

Approachability

Approachability was a concept created by David Blackwell to demonstrate how an algorithm could be constructed to converge a value to a set S. To fully understand where David Blackwell derived approachability, we must first understand Jon Von Neumann's Minimax theorem (Neumann 1928).

Minimax Theorem

In a two player, zero-sum, normal form game, players 1 and 2 can select any strategy from the sets $P = \Delta^r, Q = \Delta^s$ respectively. Where Δ^n is a simplex in n space. The minimax theorem states that there exists strategies $p \in P$ and $q \in Q$ for players 1 and 2 such that :

$$\max_{p \in P} \min_{q \in Q} b(p,q) = \min_{p \in P} \max_{q \in Q} b(p,q)$$

Where

$$b(p,q) = p^T A q, \quad A \in \mathbb{R}^{r \times s}$$

and A is the payoff matrix of a zero-sum game (Neumann 1928).

To see how this relates to approachability, David Blackwell defines a new game (Blackwell 1956). In this game, two players play a normal form game n times. On each iteration, player 1 and player 2 choose strategies from sets P and Q respectively. After choosing a strategy, each player receives a valued vector $x_i \in \mathbb{R}^N$. The strategies, picked by the players, depend on the values incurred at each stage of the game, where:

$$f_n \in P : f_n(x_1, x_2, x_3, ..., x_n)$$
$$g_n \in P : f_n(x_1, x_2, x_3, ..., x_n)$$

Then we can define sets f and g, where:

$$f = \{f_1, f_2, ..., f_n\}$$
$$g = \{g_1, g_2, ..., g_n\}$$

The value x incurred on iteration i is chosen from matrix,

$$M = \|m(i,j)\|, \quad 0 \leq i \leq r, 0 \leq j \leq s$$

Where each element in the matrix M is a probability distribution. Given a set S in N space, S is said to be approachable with f^* in M, if for every $\epsilon > 0$, there exists an N_0, such that for every g:

$$Prob\{\delta_n \geq \epsilon \text{ for some } N_0\} < \epsilon$$

Where δ_n is the shortest distance from the point $\frac{1}{n}\sum_{i=1}^{n} x_i$ to s.

Then, it is said that S is excludable with g^* in M, if there exists a $d > 0$ such that for every $\epsilon > 0$, there is an N_0 such that for every f:

$$Prob\{\delta_n \geq d \text{ for all } n \geq N_0\} > 1 - \epsilon$$

Blackwell asserts that Jon Von Neumann's Minimax theorem could be described in terms of approachability/exclusivity when $N = 1$. Then, with every M there exists number v and vectors $p \in P$ and $q \in Q$ such that the set $S = \{x \geq t\}$ is approachable for $t \leq v$ with $f : f_n \equiv p$ and excludable for $t > v$ with $g : g_n \equiv q$ (Blackwell 1956). To demonstrate, why this is true, let us assume that p and q are vectors that satisfy:

$$\max_{p \in P} \min_{q \in Q} b(p,q) = \min_{p \in P} \max_{q \in Q} b(p,q)$$

then, let us assume that $f : f_n \equiv p$ is approachable to S and $g : g_n \equiv q$ is excludable. We will show that under these assumptions, from our definitions of approachability, we can derive the Minimax theorem. Given our set M which holds probability distributions on each index (i,j). Let $\hat{M} = \bar{m}$ be the average values of each distribution on index (i,j), then:

$$f : f_n \equiv p, \implies$$
$$\frac{1}{n}\sum_{i=1}^{n} f_i^T \hat{M} g_i = \frac{1}{n}\sum_{i=1}^{n} p^T \hat{M} q$$

Let the set $S = \{x \geq t\} : t = v$, then there exists an N_0 for an arbitrailiy small ϵ, such that for every g:

$$Prob\{\delta_n \geq \epsilon \text{ for some } N_0\} < \epsilon \implies$$
$$\min_{s \in S} |s - \frac{1}{N_0} \sum_{i=1}^{N_0} p^T \hat{M} g_i| = 0 \implies$$
$$\frac{1}{N_0} \sum_{i=1}^{N_0} p^T \hat{M} g_i \geq v$$

Similarly:

$$g : g_n \equiv q, \implies$$
$$\frac{1}{n} \sum_{i=1}^{n} f_i^T \hat{M} g_i = \frac{1}{n} \sum_{i=1}^{n} f_i^T \hat{M} q$$

Then, because q is excludable to the set $S = \{x \geq t\} : t > v$, then when we set t sufficiently close to v, we can say that there exists N_0 and $d > 0$ for all $\epsilon > 0$ such that for all f:

$$Prob\{\delta_n \geq d \text{ for all } n \geq N_0\} > 1 - \epsilon \implies$$
$$\min_{s \in S} |s - \frac{1}{N_0} \sum_{i=1}^{N_0} f_i^T \hat{M} q| > 0 \implies$$
$$\frac{1}{N_0} \sum_{i=1}^{N_0} f_i^T \hat{M} q \leq v$$

Thus,

$$\frac{1}{N_0} \sum_{i=1}^{N_0} f_i^T \hat{M} q \leq v \leq \frac{1}{N_0} \sum_{i=1}^{N_0} p^T \hat{M} g_i \implies$$
$$\max_{f_i \in P} f_i^T \hat{M} q = v, \quad \min_{g_i \in Q} p^T \hat{M} g_i = v \implies$$
$$\max_{f_i \in P} \min_{g_i \in Q} f_i^T \hat{M} g_i = \min_{g_i \in Q} \max_{f_i \in P} f_i^T \hat{M} g_i$$

Next, blackwell claims that given an arbitrary vector $\mathbf{x} \in \mathbb{R}^N$ not in set S, there exists a vector $\mathbf{y} \in S$ such that $\mathbf{y}$ is the closest point in S to $\mathbf{x}$. If for every $\mathbf{x} \notin S$ there exists a strategy p such that the convex hull of s points of $R(p) = convexHull(p\hat{M}) \subseteq \mathcal{H}$. Where $\mathcal{H} = \{\mathbf{z} | (\mathbf{y} - \mathbf{z}) \cdot (\mathbf{y} - \mathbf{x}) \leq 0\}$. Then S is approachable with the sequence of strategies:

$$f : f_n = \begin{cases} p(\bar{\mathbf{x}}_n) \text{ if } \bar{\mathbf{x}}_n = \sum_{i=1}^{n} \mathbf{x}_i \notin S \\ \text{any strategy to remain in S otherwise} \end{cases}$$

To prove this, let us assume that for every point $\mathbf{x} \notin S$, that there exists a strategy $p(\mathbf{x})$, where $R(p(\mathbf{x})) \subseteq \mathcal{H}$. That is, we want to prove that the average value of our sequence of strategies, $f : f_n$, approaches our set S (Farina 2023).

Let $\mathbf{y}_i$ be the closest point from S to $\bar{\mathbf{x}}_i$. Then, let $\mathbf{u}_n = \mathbf{y}_n - \bar{\mathbf{x}}_n$. Next, define the halfspace: $\mathcal{H}_n = \{\mathbf{z} | (\mathbf{y}_n - \mathbf{z}) \cdot \mathbf{u}_n \leq 0\}$.

$$\begin{aligned}
&\bar{\mathbf{x}}_{n+1} = \frac{n}{n+1}\bar{\mathbf{x}}_n + \frac{1}{n+1}\mathbf{x}_{n+1}, \\
&\|\mathbf{y}_n - \bar{\mathbf{x}}_{n+1}\|_2^2 = \|\mathbf{y}_n - (\frac{n}{n+1}\bar{\mathbf{x}}_n + \frac{1}{n+1}\mathbf{x}_{n+1})\|_2^2 = \\
&\|\mathbf{y}_n - \frac{n}{n+1}\bar{\mathbf{x}}_n - \frac{1}{n+1}\mathbf{x}_{n+1})\|_2^2 = \\
&\|\frac{n}{n+1}(\mathbf{y}_n - \bar{\mathbf{x}}_n) + \frac{1}{n+1}(\mathbf{y}_n - \mathbf{x}_{n+1})\|_2^2
\end{aligned}$$

Given two arbitrary vectors $\mathbf{u}$ and $\mathbf{v}$, $\|\mathbf{u} + \mathbf{v}\|_2^2 = \|\mathbf{u}\|_2^2 + \|\mathbf{v}\|_2^2 + 2(\mathbf{u} \cdot \mathbf{v})$

$$\begin{aligned}
&\|\frac{n}{n+1}(\mathbf{y}_n - \bar{\mathbf{x}}_n) + \frac{1}{n+1}(\mathbf{y}_n - \mathbf{x}_{n+1})\|_2^2 = \\
&\frac{n^2}{(n+1)^2}\|\mathbf{u}_n\|_2^2 + \frac{1}{(n+1)^2}\|\mathbf{y}_n - \mathbf{x}_{n+1}\|_2^2 + \frac{2n}{n+1}(\mathbf{u}_n \cdot (\mathbf{y_n} - \mathbf{x}_{n+1}))
\end{aligned}$$

By our assumption, we know that:

$$\begin{aligned}
&\mathbf{x}_{n+1} \in \mathcal{H}_n \\
\implies &(\mathbf{y}_n - \mathbf{x}_{n+1}) \cdot \mathbf{u}_n \leq 0 \\
\implies &\|\mathbf{y}_n - \bar{\mathbf{x}}_{n+1}\|_2^2 \leq \frac{n^2}{(n+1)^2}\|\mathbf{u}_n\|_2^2 + \frac{1}{(n+1)^2}\|\mathbf{y}_n - \mathbf{x}_{n+1}\|_2^2
\end{aligned}$$

The value $\mathbf{x}_{n+1}$ is bounded by the possible utilities we can recieve from our game. Thus, given the game, we can bound $\|\mathbf{y}_n - \mathbf{x}_{n+1}\|_2^2 \leq \Omega^2$.

$$\begin{aligned}
&\|\mathbf{y}_n - \bar{\mathbf{x}}_{n+1}\|_2^2 \leq \frac{n^2}{(n+1)^2}\|\mathbf{u}_n\|_2^2 + \frac{\Omega^2}{(n+1)^2} \\
\implies &\|\mathbf{u}_{n+1}\|_2^2 \leq \frac{n^2}{(n+1)^2}\|\mathbf{u}_n\|_2^2 + \frac{\Omega^2}{(n+1)^2} \\
\Leftrightarrow \quad &(n+1)^2\|\mathbf{u}_{n+1}\|_2^2 - n^2\|\mathbf{u}_n\|_2^2 \leq \Omega^2
\end{aligned}$$

Telescoping terms from $i = 0...n-1$ results in (Farina 2023):

$$\begin{aligned}
&(n+1)^2\|\mathbf{u}_{n+1}\|_2^2 \leq n\Omega^2 \\
\Leftrightarrow \quad &\|\mathbf{u}_{n+1}\|_2 \leq \frac{\sqrt{n}\Omega}{n+1}
\end{aligned}$$

Which implies that when we assume S is approachable, we converge to S at a rate of $O(1/\sqrt{n})$ by playing strategy $f : f_n$.

References

Blackwell, David. 1956. "An Analog of the Minimax Theorem for Vector Payoffs." *Pacific Journal of Mathematics* 6 (1): 1–8. https://doi.org/10.2140/pjm.1956.6.1.

Farina, Gabriele. 2023. "Lecture 5A: Blackwell Approachability."

Lanctot, Marc, Kevin Waugh, and Michael Bowling. 2009. "Monte Carlo Sampling for Regret Minimization in Extensive Games." In *Advances in Neural Information Processing Systems*.

Neumann, John von. 1928. "Zur Theorie Der Gesellschaftsspiele." *Mathematische Annalen* 100: 295–320. https://doi.org/10.1007/BF01448847.