Counterfactual Regret Minimization

COMP3801 Report

Introduction

- Imperfect information games: Players lack full knowledge of the game state or past actions
- Subgames are interdependent
  - Probabilities with which actions are taken in one area affect others
- Counterfactual regret minimization
  - An iterative algorithm that converges to a Nash Equilibrium

Preliminaries

- Normal-form games
  - Simultaneous moves
  - Represented by payoff matrices
- Extensive-form games
  - Sequential moves
  - Modeled as a Game Tree (Nodes = States, Edges = Actions)
- Strategy Types
  - Pure strategy: Deterministic plan
  - Mixed strategy: Probability distribution over pure strategies
  - Behavioral strategy: Independent probability distribution at each information set
- Nash Equilibrium
  - A profile where no player can improve their payoff by unilaterally changing their strategy.

Regret

- Regret is a measure of what could have been
  - $r^t(a) = u^t(a) - u^t(a^t)$
- Regret Minimization algorithms
  - If average cumulative regret approaches zero as $T \to \infty$, and players play according to the strategy
  - In two player zero sum games, the strategy converges to a set of Nash equilibria

Regret Matching Algorithm

1. For T iterations
    a. For every action calculate the immediate regret for the chosen action
    b. Add immediate regret to the cumulative total. $R^t(a_i) = R^{t-1}(a_i) + r^t(a_i)$
    c. Normalize the strategy, the probability of choosing an action is proportional to its positive cumulative regret $\sigma^{t+1}(a_i) = \frac{R^{t,+}(a_i)}{\sum_{j=1}^{K} R^{t,+}(a_j)}$

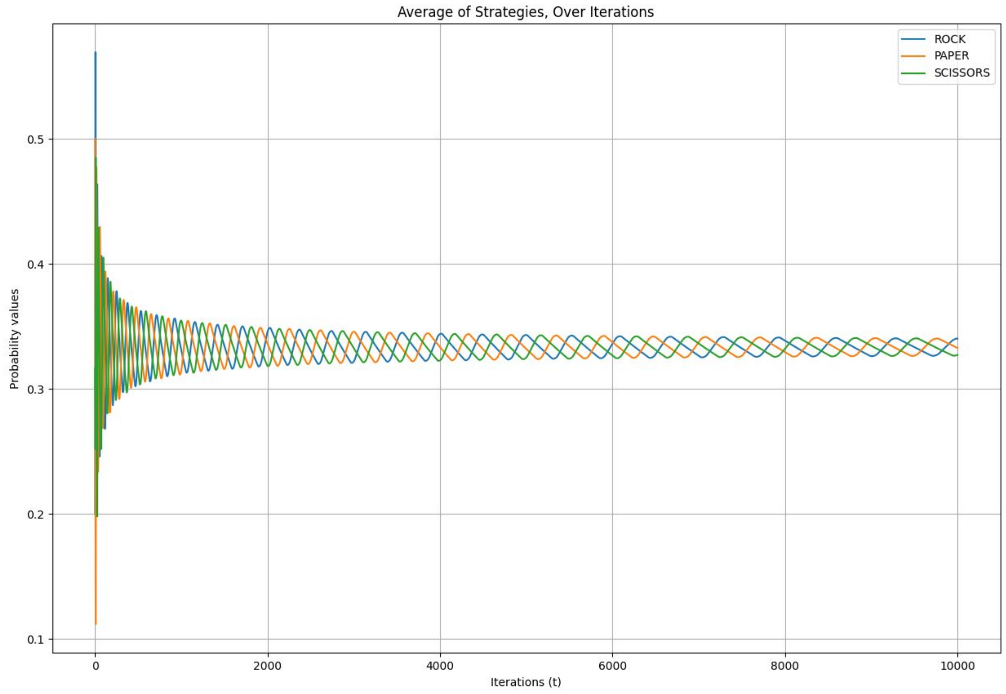
2. Take the average strategy over the T iterations $\bar{\sigma}^T(a_i) = \frac{\sum_{t=1}^{T} \sigma^t(a_i)}{T}$

Regret Matching Analysis

- Regret grows sublinearly, as T increases $R^T \leq \Delta u \sqrt{KT}$
- Average regret approaches zero, at $O(1/\sqrt{T})$
- If regret is bounded by ϵ, it guarantees a strategy profile where no player can gain more than ϵ by deviating
- To reach an error threshold ϵ iterations are required are $T \geq \frac{(\Delta u)^2 K}{\epsilon^2}$
- $O(1/\epsilon^2)$

Implementation Rock Paper Scissors

- Random seed values to avoid an immediate solution
- Converges to (⅓,⅓ ,⅓) the optimal solution



Counterfactual Regret Minimization(CFR)

- Extending from normal form to extensive form
  - How to assign regret to early actions, when outcomes are not known yet.
- Counterfactual regret: Expected utility of current strategy minus expected utility of current strategy if alternate action were selected.

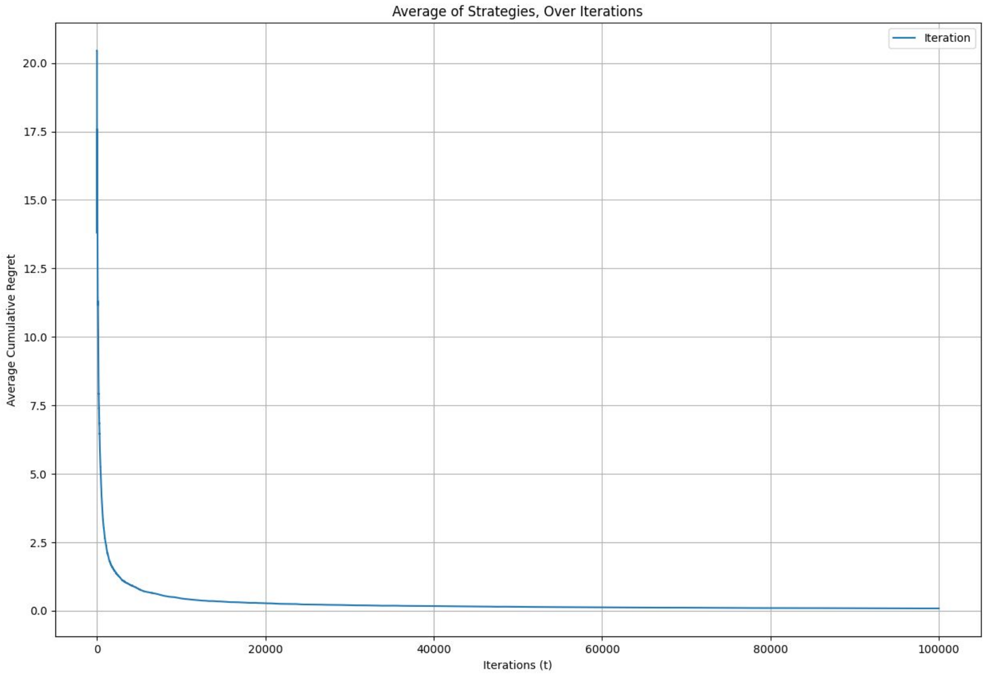
1. Walk the game tree, using a depth first traversal
2. Send probabilities of reaching current node in the traversal
3. Return utility values, weighted by probability of opponent reaching the node.
4. Update counterfactual regrets, as recursion unwinds
5. Update average strategy, compute weight using the players reach probability.

CFR Analysis

- Bound on regret: $R_i^T \leq |\mathcal{I}_i| \Delta u_i \sqrt{|A_{max}| T}$
- Overall time complexity $O\left(\frac{M \cdot |\mathcal{I}|^2}{\epsilon^2}\right)$

Leduc Holdem Implementation

- To measure correctness a best response function is used
- Traverses the game tree measuring the maximal exploitability



Conclusion

- Regret minimization algorithms
- Regret matching: Solutions for normal form games
- Counterfactual regret minimization: Solutions for extensive form games
- Potential next steps

References

[1] M. Zinkevich, M. Johanson, M. Bowling and C. Piccione. Regret minimization in games with incomplete information. *Advances in neural information processing systems*, 20:1729–1736, 2007.

[2] S. Hart and A. Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68(5):1127–1150, 2000.

[3] M. Maschler, E. Solan and S. Zamir. *Game Theory*. Cambridge: Cambridge University Press, 2013.

[4] T. W. Neller and M. Lanctot. An introduction to counterfactual regret minimization. Technical Report, Gettysburg College, 2013.

[5] N. Brown and T. Sandholm. Safe and nested subgame solving for imperfect-information games. *Advances in neural information processing systems*, 30:896–906, 2017.