Carnegie Mellon University

Faster Game Solving via Hyperparameter Schedules

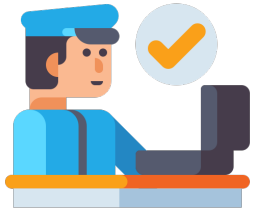
Naifeng Zhang[1], Stephen McAleer[2], Tuomas Sandholm[1345]

[1]Carnegie Mellon University, [2]Anthropic, [3]Strategy Robot, Inc., [4]Strategic Machine, Inc., [5]Optimized Markets, Inc.

AAAI 2026

Many real-world settings are games

Specifically, **imperfect-information games (IIGs)**



...to deceive and to understand deception

Our focus: solving two-player zero-sum IIGs

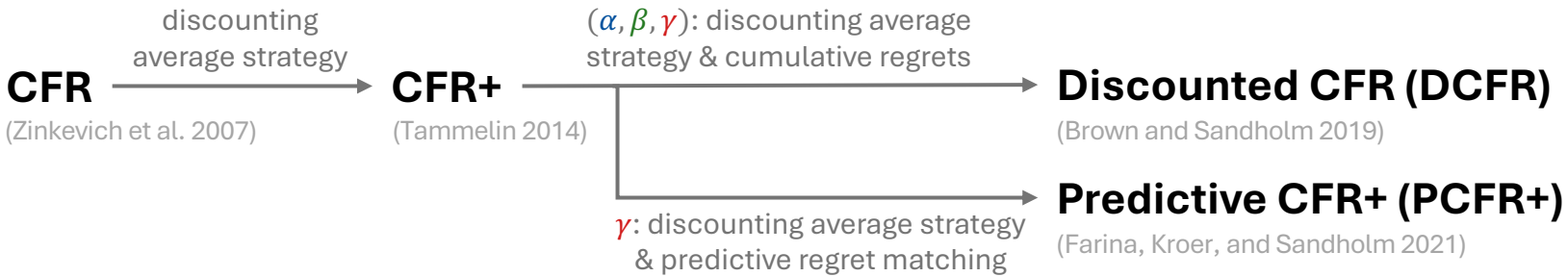
by converging to a Nash equilibrium

Counterfactual regret minimization (CFR) algorithms

Iteratively reducing regret to guide the average strategy toward a Nash equilibrium

Evolution of CFR variants

Discounting the contribution of early iterations



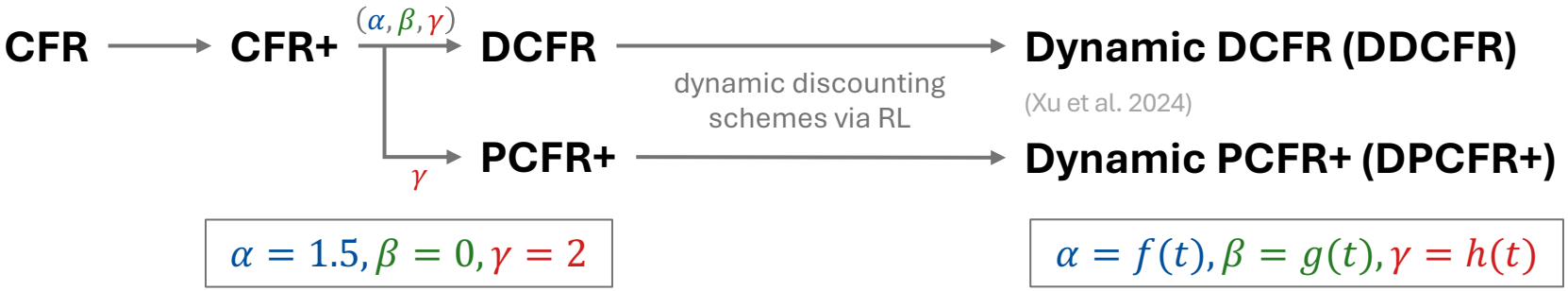
On iteration $t + 1$, DCFR multiplies

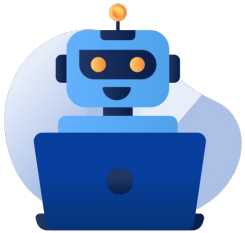
- positive cumulative regrets by $\frac{t^{\alpha}}{t^{\alpha}+1}$
- negative cumulative regrets by $\frac{t^{\beta}}{t^{\beta}+1}$
- contributions to the average strategy by $\left(\frac{t}{t+1}\right)^{\gamma}$

$$\alpha = 1.5, \beta = 0, \gamma = 2$$

Evolution of CFR variants, cont.

Discounting the contribution of early iterations





Training: 24 hours on 200 CPU cores
Inference: game-specific, multiprocessing required

Evolution of CFR variants, cont.

Discounting the contribution of early iterations

CFR → CFR+ → DCFR → Dynamic DCFR (DDCFR)

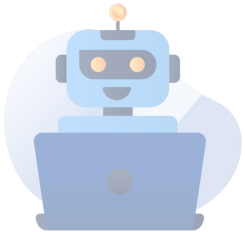
(α, β, γ)

dynamic discounting schemes via RL

(Xu et al. 2024)

PCFR+ → Dynamic PCFR+ (DPCFR+)

Is there a simple, training-free method to achieve strong performance without game-specific tuning?

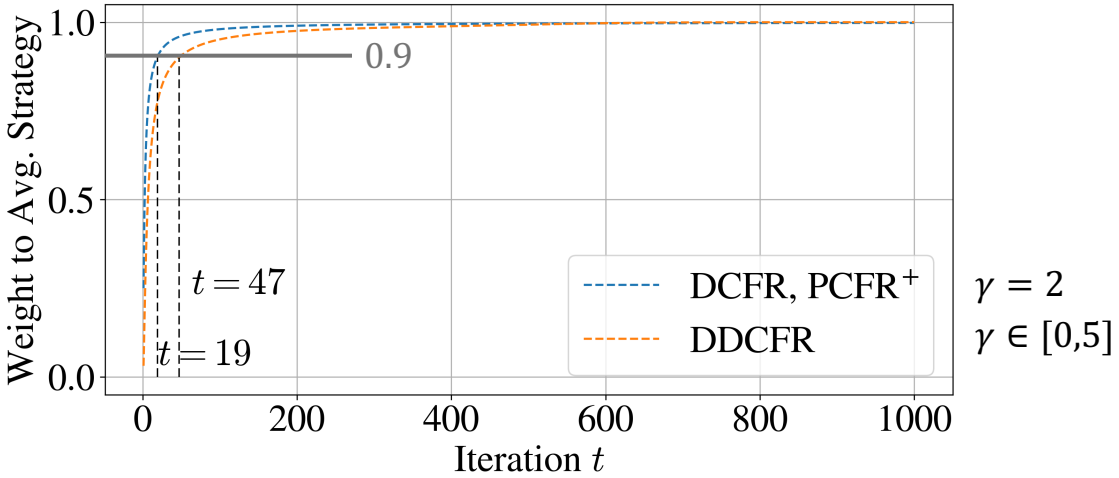


Training: 24 hours on 200 CPU cores
Inference: game-specific, multiprocessing required

Solution: Hyperparameter Schedules (HSs)

Discounting schemes that control how hyperparameter changes

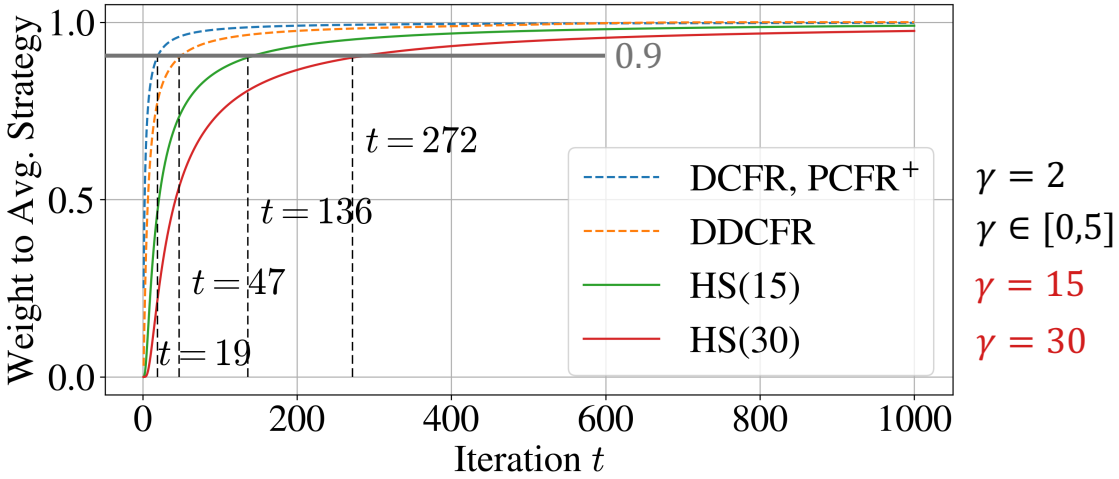
Prior CFR variants' discounting schemes are **not sufficiently aggressive**



Evolution of the weight $\big(t/(t+1)\big)^{\gamma}$ applied to the contribution to the average strategy

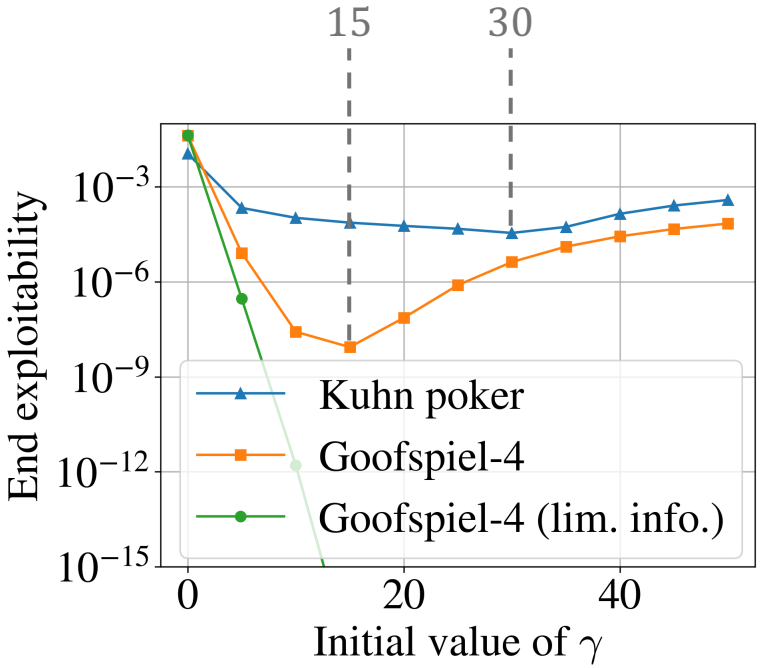
Hyperparameter Schedules (HSs), cont.

Aggressively downweighting unrefined strategies from early updates



Evolution of the weight $\big(t/(t+1)\big)^{\gamma}$ applied to the contribution to the average strategy

Identifying effective HSs



$$\text{HS}_\alpha : \alpha = 1 + \frac{3}{n}\,t, \qquad \text{HS}_\beta : \beta = -1 - \frac{2}{n}\,t,$$

$$\text{HS}_{\gamma 30} : \gamma_{30} = 30 - \frac{5}{n}\,t, \qquad \text{HS}_{\gamma 15} : \gamma_{15} = 15 - \frac{5}{n}\,t.$$

n is the total number of iterations

Implementing HS-powered algorithms

- **HS-powered DCFR (HS-DCFR):** $(HS_\alpha, HS_\beta, HS_\gamma)$

$$\frac{t^\alpha}{t^\alpha + 1} \to \frac{t^{HS_\alpha}}{t^{HS_\alpha} + 1}$$

$$\frac{t^\beta}{t^\beta + 1} \to \frac{t^{HS_\beta}}{t^{HS_\beta} + 1}$$

- **HS-powered PCFR+ (HS-PCFR+):** HS_γ

$$\left(\frac{t}{t+1}\right)^\gamma \to \left(\frac{t}{t+1}\right)^{HS_\gamma}$$

< 15 lines of code changes!

Theoretical guarantees

Provable convergence to a Nash equilibrium in two-player zero-sum games

HS-DCFR:

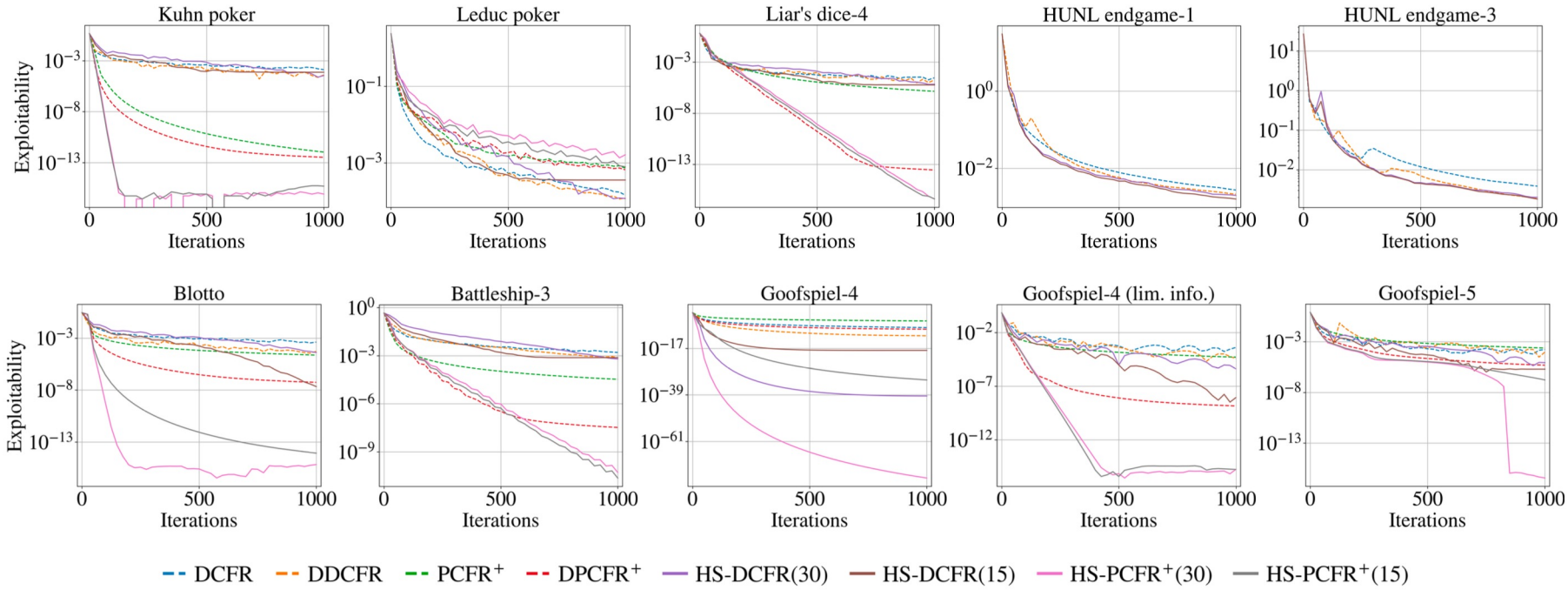
Theorem 3.1. *Suppose T iterations of HS-DCFR, with simultaneous updates, are played in a two-player zero-sum game, and U is the upper bound of γ across all iterations. If $\alpha \in [1, 5]$, $\beta \in [-5, 0]$, and $\gamma \in [0, U]$, the weighted average strategy profile is a $(U+1)\Delta|\mathcal{I}|\left(\frac{8}{3}\sqrt{|\mathcal{A}|}+\frac{2}{\sqrt{T}}\right)/\sqrt{T}$-Nash equilibrium.*

HS-PCFR+:

Theorem 3.2. *Suppose T iterations of HS-PCFR$^+$, with simultaneous updates, are played in a two-player zero-sum game, and U is the upper bound of γ across all iterations. If $\gamma \in [0, U]$, the weighted average strategy profile is a $(U+1)|\mathcal{I}|O(1)/\sqrt{T}$-Nash equilibrium.*

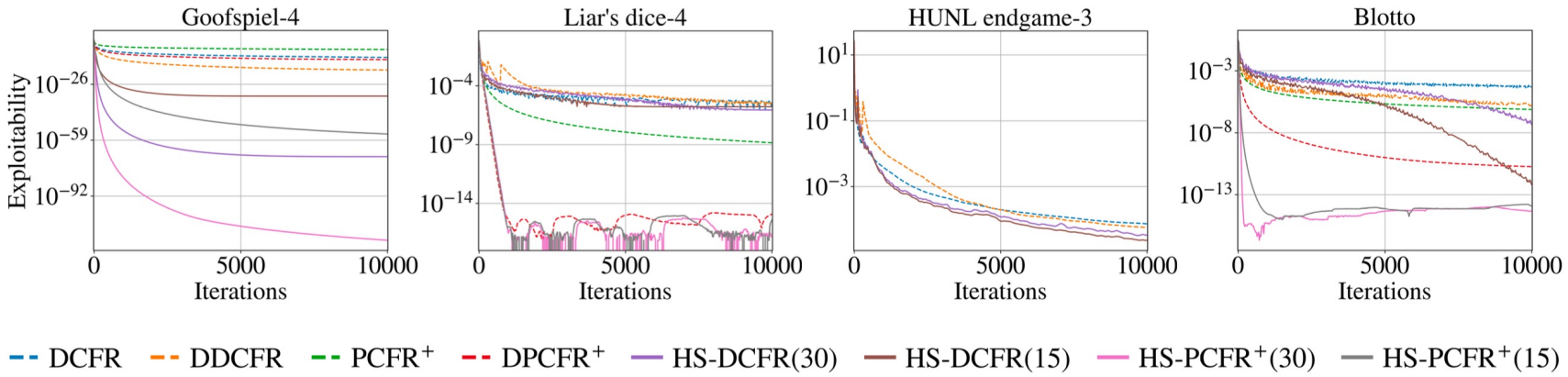
State of the art on ten diverse games

HS-PCFR+(30) outperforms prior SoTA by 12 orders of magnitude on average



SoTA performance with extended iterations

HS-powered algorithms consistently outperform prior SoTA



Thank you!

Extended version at arxiv.org/pdf/2404.09097

Naifeng Zhang

naifengz@cmu.edu

Stephen McAleer

mcaleer.stephen@gmail.com

Tuomas Sandholm

sandholm@cs.cmu.edu