Estimating α-Rank from A Few Entries with Low Rank Matrix Completion

Yali Du*[1], Xue Yan*[2], Xu Chen[3], Jun Wang[1], Haifeng Zhang[2]

[1]University College London, [2]Institute of Automation, CAS, [3]Renmin University of China

June 21, 2021

Introduction

Background

- Multi-agent evaluation aims at the assessment of an agent's strategy on the basis of interaction with others.
- Renowned evaluation algorithms: Elo rating system [Elo78], α-rank [Omi+19].

α-**rank** for two-player game with a single population:

- build Markov transitive matrix according to payoff $\boldsymbol{M}$;

$$C_{i,j} = \begin{cases} \eta \frac{1-\exp\left(-\alpha\left(M_{ji}-M_{ij}\right)\right)}{1-\exp\left(-\alpha p\left(M_{ji}-M_{ij}\right)\right)} & \text{if } M_{ji} \neq M_{ij} \\ \frac{\eta}{p} & \text{otherwise} \end{cases}$$

- compute the invariant distribution $\boldsymbol{\pi}$;
- return the ranking of strategies according to $\boldsymbol{\pi}$.

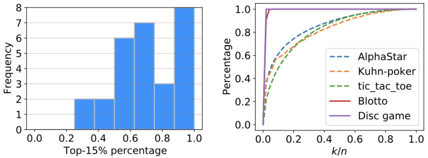
Low-rank Property

Motivation:

- Repeated strategies may exist in the empirical games.
- Agents who have similar skills might perform similarly.
- Rowland et al. estimate all pairs, which is computationally heavy.

Evidence from real-world games from [Cza+20]:

Game	# policies	rank	k
3-move parity game 2	160	14	9
Blotto	1001	50	16
hex(board_size=3)	766	764	232
Disc game	1000	2	2
Normal Bernoulli game	1000	1000	499
Elo game	1000	38	2
Random game of skill	1000	1000	515
Transitive game	1000	2	2
Triangular game	1000	1000	137
AlphaStar	888	888	238
tic_tac_toe	880	880	285
Kuhn-poker	64	64	24

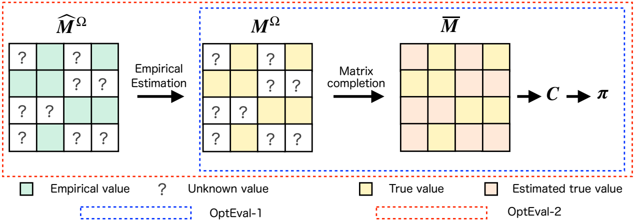


(a) Histograms of 15% singular values percentage (b) Percentage of top singular values

We need to remove the necessity of exhaustively comparing all strategy pairs.

Our Algorithms

We are motivated to consider low-rank property of payoff matrices. The overall architecture of OptEval



We propose two algorithms with noise-free and noisy payoffs.

Algorithm 1 optEval-1: α-rank with noise-free payoff.

Require: A chosen rank $\hat{r}$, sampling operator $\Omega \in [n] \times [n]$.
Ensure: The invariant distribution of n players $\bar{\pi}$.

Randomly sample m pairs from the entire sample space $[n] \times [n]$ by the sampling operator Ω.
Get pairwise comparison results M^{Ω}
Calculate the reconstructing payoff matrix $\bar{M}$ according to **OptSpace** with rank $\hat{r}$ in Algorithm 4.
Construct the Markov chain $\bar{C}$ through Eq. (1)
Solve the invariant distribution $\bar{\pi}$ of $\bar{C}$
Return $\bar{\pi}$

Algorithm 2 Learning α-rank with noisy payoff.

Require: n players' strategies, a chosen rank $\hat{r}$, a sampling operator $\Omega \in [n] \times [n]$
Ensure: The invariant distribution of n players $\bar{\pi}$.

Randomly sample m pairs from the entire sample space $[n] \times [n]$ by Ω.
Call Algorithm 3 on Ω to get noisy pairwise comparison results $\hat{M}^{\Omega}$
Perform **OptSpace** on $\hat{M}^{\Omega}$ with rank $\hat{r}$ and calculate the reconstructing payoff matrix $\hat{\bar{M}}$
Construct the Markov chain $\hat{\bar{C}}$ through Eq. (1)
Solve the invariant distribution $\hat{\bar{\pi}}$ of $\hat{\bar{C}}$
Return $\bar{\pi}$

Theoretical Results (Informal)

Theorem (Noise-free evaluations)

Let $\Omega \subseteq [n] \times [n]$ be a randomly selected set of pairs to be evaluated, then there exists a constant C such that if Ω satisfies

$$|\Omega| \geq Cnr\kappa^2 \max\left\{\mu_0 \log n, \mu^2 r\kappa^4\right\},$$

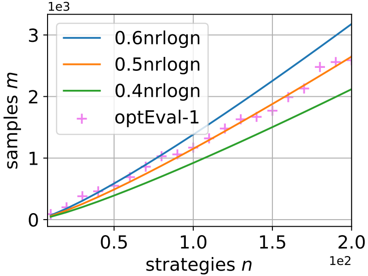
then we can obtain the exact invariant distribution with high probability.

Theorem (Noisy evaluations)

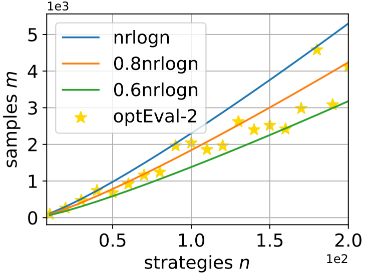
Let Ω be a sample operator, for each pair $(i,j) \in \Omega$, let $\hat{\mathrm{M}}_{ij}$ be an empirical payoff constructed by taking K i.i. d. interactions of player i and j. There exist constants C and C' such that if the number of randomly selected pairs m satisfies $|\Omega| \geq C\kappa^2 n \max(\mu_0 r \log(n), \mu_0^2 r^2\kappa^4, \mu_1^2 r^2\kappa^4)$ and K satisfies $K \geq \frac{2592 M_{\max}^2 \log(2mn^3) L(\alpha, M_{\max})^2 \left(\sum_{i=1}^{n-1} \binom{n}{i} i^n\right)^2 C'^2 \kappa^4 rn^2}{\epsilon^2 g(\alpha,\eta,p,M_{\max})^2}$*, Then* $\max_{i\in[n]} |\bar{\boldsymbol{\pi}}(i) - \boldsymbol{\pi}(i)| \leq \epsilon$ *is satisfied with probability at least* $1 - \frac{2}{n^3}$.

Experiments I

Empirical sample complexity: Results on twenty $n \times n$ **Gaussian games** with $n = 10, 20, \ldots, 200$, $r = 5$. (a) the empirical sampling complexity of OptEval-1 when $\epsilon \leq 10^{-4}/n$ at a chosen rank $r = 5$. (b) the empirical sampling complexity when OptEval-2 outperforms RG-UCB.



(a) Noise-free case by OptEval-1



(b) Noisy case by OptEval-2

Experiments II

Real-world games: OptEval approximate the games that are both low-rank and high-rank, i.e. approximating full-rank AlphaStar by $r = 32$.

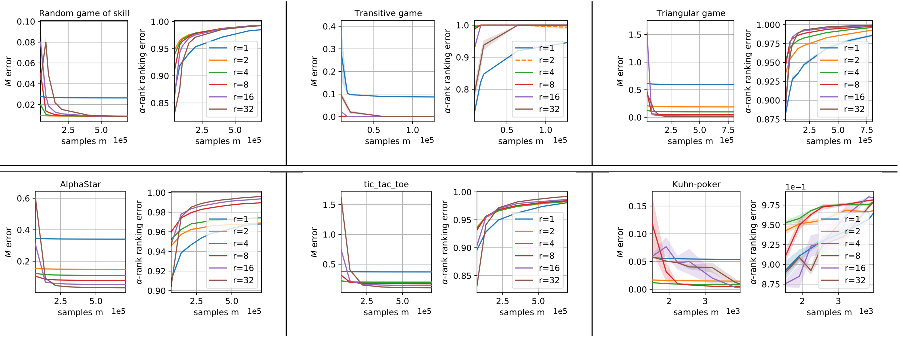


Table: Results on selected real world games with noise free evaluations. The number of entries are reduced by 60%-80%.

The End

GitHub Page