A Critical Review of Multi-agent Evaluation

Xue Yan

2020.10.25

Outlines

- Problem background：problem statement，importance
- Classical methods：ELO，Glicko，TrueSkill
- Improved methods：mELO，Nash averaging，α-Rank
- Sampling complexity analysis
- Challenges

Problem Background

Problem statement：

- Input：Given a group of agents，and game outcome
- Output： rank／score／distribution of the group of agents
- efficient、robust、validity、general

Importance ： Evaluating agents；Promote the improvement of the algorithm

Question：What is optimal？How to find optimal？

Game Type Discussion

Game out come	Single		Team	
1 v 1	A > B Go,Chess	ELO、Glicko、mELO、Nash Averaging、RD、α−rank	[A, B, …] > [X, Y, …] Glory of Kings	TrueSkill、ELO+weight
multiplayer	StarCraft Poker	TrueSkill A > B > C > D α−rank Strategy profile	[A,B] > [C,D] > [E,F]	TrueSkill

AVT：Agent Vs Task such as Atari, Nash Averaging

cooperative game → competitive game

Outlines

- Problem background：problem statement，importance
- Classical methods：ELO，Glicko，TrueSkill
- Improved methods：mELO，Nash averaging，α-Rank
- Sampling complexity analysis
- Challenges

ELO Rating System

- Assumption: transitive、fixed variance
- Elo assigns a rating r_i to each player $i \in [n]$ based on their wins and losses
- Prediction
  - probability of i beating j

[2] A. E. Elo, The Rating of Chess players, Past and Present. Ishi Press International, 1978.

$$\hat{p}_{ij} := \frac{10^{r_i/400}}{10^{r_i/400} + 10^{r_j/400}} = \sigma(\alpha r_i - \alpha r_j), \quad \text{where} \quad \sigma(x) = \frac{1}{1+e^{-x}} \quad \text{and} \quad \alpha = \frac{\log(10)}{400}.$$

$$\ell_{\text{Elo}}(p_{ij}, \hat{p}_{ij}) = -p_{ij} \log \hat{p}_{ij} - (1 - p_{ij}) \log(1 - \hat{p}_{ij}), \quad \text{where} \quad \hat{p}_{ij} = \sigma(r_i - r_j)$$

$$r_i^{t+1} \leftarrow r_i^t - \eta \cdot \nabla_{r_i} \ell_{\text{Elo}}(S_{ij}^t, \hat{p}_{ij}^t) = r_i^t + \eta \cdot (S_{ij}^t - \hat{p}_{ij}^t).$$

Glicko

- The reliability of a player's rating.
  - ELO： only a rating
  - Glicko：Rating Deviation (RD) + rating
- The explanation for RD
  - A high RD indicates that a player may not be competing frequently , a low RD indicates that a player competes frequently.
  - Confidence interval
    - Rating 1800, RD is 50 , with 95% confidence in [1750,1850]

Glicko

- Algorithm

1. Ageing

$$\mathrm{RD} = \min(\sqrt{\mathrm{RD}_{old}^2 + c^2},\ 350)$$

2. update

$$d^2 = \left(q^2 \sum_{j=1}^{m} (g(\mathrm{RD}_j))^2 \mathrm{E}(s|r, r_j, \mathrm{RD}_j)(1 - \mathrm{E}(s|r, r_j, \mathrm{RD}_j)) \right)^{-1}$$

Low RD, Large influence !

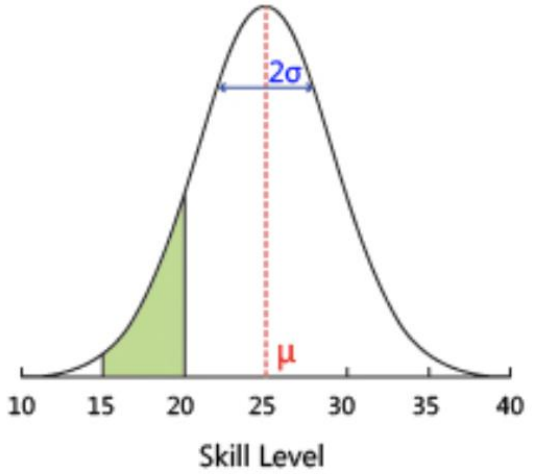
$$r' = r + \frac{\frac{\ln 10}{400}}{1/\mathrm{RD}^2 + 1/d^2} \sum_{j=1}^{m} g(\mathrm{RD}_j)(s_j - \mathrm{E}(s|r, r_j, \mathrm{RD}_j))$$

$$\mathrm{RD}' = \sqrt{\left(\frac{1}{\mathrm{RD}^2} + \frac{1}{d^2} \right)^{-1}}$$

TrueSkill

- TrueSkill ranking system skill is characterized by **two** numbers.
  - The average skill of the gamer (μ in the picture).
  - The degree of uncertainty in the gamer's skill (σ in the picture).

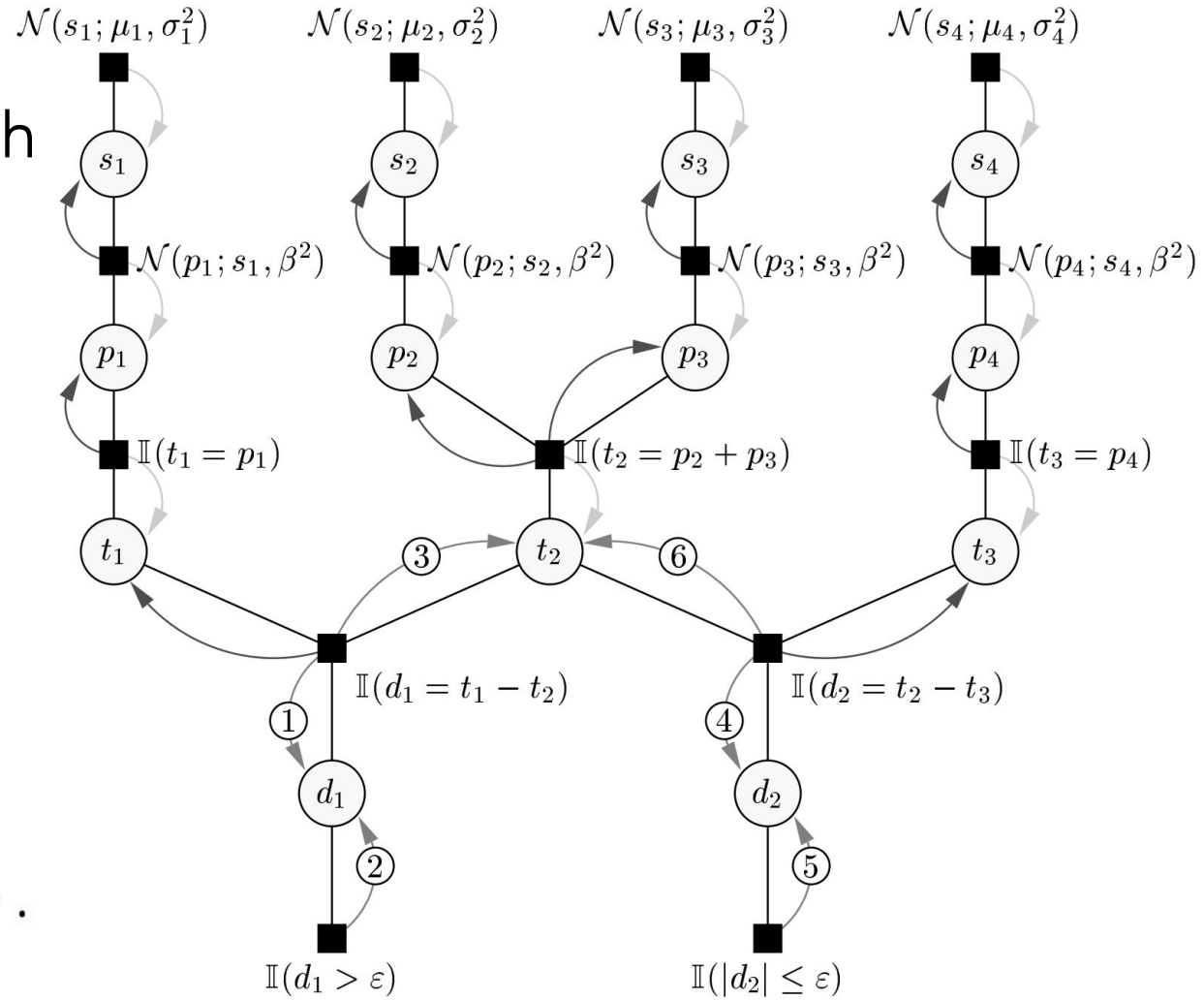
- More complex battle forms
  - Multi—team , Multi—player
  - Gaussian distribution(skill of player and team)



[3] R. Herbrich, T. Minka, and T. Graepel, “TrueSkill: a Bayesian skill rating system,” in NIPS, 2007.

Algorithm

- Based on Gaussian factor graph
  - Skills of all players
  - Performance of all players
  - Skills of teams(sum)
  - Differences between teams
- Bayesian methods
  - $$p(\mathbf{s}|\mathbf{r}, A) = \frac{P(\mathbf{r}|\mathbf{s}, A)\, p(\mathbf{s})}{P(\mathbf{r}|A)}.$$



Weakness

- Only Sum—product
  - There is no effective modeling of cooperative relationships, just simply adding up each player
- Transitive
  - Like ELO, it's still a Gaussian probability model.

[1] The transitive relation "1 draws with 2" is not modelled exactly by the relation $|t_1 - t_2| \leq \varepsilon$, which is non-transitive. If $|t_1 - t_2| \leq \varepsilon$ and $|t_2 - t_3| \leq \varepsilon$ then the model generates a draw among the three teams despite the possibility that $|t_1 - t_3| > \varepsilon$.

Outlines

- Problem background：problem statement，importance
- Classical methods：ELO，Glicko，TrueSkill
- Improved methods：mELO，Nash averaging，α-Rank
- Sampling complexity analysis
- Challenges

Cyclic game

Re-evaluating Evaluation，NIPS 2018

- Elo bakes—in the assumption that relative skill is transitive

$$\mathbf{C} = \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{T} = \begin{pmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \\ -2 & -1 & 0 \end{pmatrix}$$

- Cyclic Game (Intransitive)
  - rock，scissors，paper will all receive the same Elo ratings
  - $\hat{p}_{ij} = \frac{1}{2}$ for all i, j actually paper beats rock with p = 1

Multidimensional Elo (mElo2k)

- Antisymmetric matrices

$$\mathbf{A} + \mathbf{A}^\intercal = 0.$$

$$P = \begin{bmatrix} 0 & p_{12} & \dots & p_{1n} \\ p_{21} & 0 & \dots & p_{2n} \\ \dots & \dots & \dots & \dots \\ p_{n1} & p_{n2} & \dots & 0 \end{bmatrix} \qquad A = logit(P) = \begin{bmatrix} 0 & a_{12} & \dots & a_{1n} \\ a_{21} & 0 & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & 0 \end{bmatrix}$$

$$a_{ij} = ln\frac{p_{ij}}{1 - p_{ij}} \qquad p_{ij} + p_{ji} = 1 \qquad a_{ij} = -a_{ji}$$

- Schur decomposition

$$\mathbf{A}_{n\times n} = \mathbf{Q}_{n\times n} \cdot \mathbf{\Lambda}_{n\times n} \cdot \mathbf{Q}^\intercal_{n\times n},$$

Multidimensional Elo (mElo2k)

- Combinatorial Hodge theory

$$\mathbf{A} = \{\textit{transitive component}\} + \{\textit{cyclic component}\} = \mathrm{grad}(\mathbf{r}) + \mathrm{rot}(\mathbf{A}) \quad \textit{where} \quad \mathbf{r} = \mathrm{div}(\mathbf{A}).$$

$$\mathbf{C} = \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{T} = \begin{pmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \\ -2 & -1 & 0 \end{pmatrix}$$

- Elo ratings just capture transitive component, but ignore the cyclic component rot(A).

Multidimensional Elo (mElo2k)

- Combining the Schur and Hodge decompositions allows to construct low-rank approximations that extend Elo

$$\mathbf{A}_{n\times n} = \operatorname{grad}(\mathbf{r}) + \tilde{\mathbf{A}} \approx \operatorname{grad}(\mathbf{r}) + \mathbf{C}^\intercal \begin{pmatrix} 0 & 1 & & \\ -1 & 0 & & \\ & & \ddots & \end{pmatrix} \mathbf{C} =: \operatorname{grad}(\mathbf{r}) + \mathbf{C}^\intercal_{n\times 2k} \mathbf{\Omega}_{2k\times 2k} \mathbf{C}_{2k\times n}$$

- The mElo2k win-loss prediction is

$$\mathbf{mElo}_{2k}\mathbf{:}\ \hat{p}_{ij} = \sigma\left(r_i - r_j + \mathbf{c}_i^\intercal \cdot \mathbf{\Omega}_{2k\times 2k} \cdot \mathbf{c}_j\right) \text{ where } \mathbf{\Omega}_{2k\times 2k} = \sum_{i=1}^{k} (\mathbf{e}_{2i-1}\mathbf{e}_{2i}^\intercal - \mathbf{e}_{2i}\mathbf{e}_{2i-1}^\intercal).$$

Application

- In a non—transitively case，mElo2 (Table mElo2) correctly predicts likely winners in all cases (Table empirical), with more accurate probabilities：
- $\alpha_v > \alpha_p > Zen > \alpha_v$

Elo	α_v	α_p	Zen
α_v	-	0.41	0.58
α_p	0.59	-	0.67
Zen	0.42	0.33	-

empirical	α_v	α_p	Zen
α_v	-	0.7	0.4
α_p	0.3	-	1.0
Zen	0.6	0.0	-

mElo$_2$	α_v	α_p	Zen
α_v	-	0.72	0.46
α_p	0.28	-	0.98
Zen	0.55	0.02	-

$\alpha_p > \alpha_v > Zen$

Nash averaging

- Given antisymmetric logit matrix A, define a two-player meta-game with *payoffs* $\mu_1(\mathbf{p}, \mathbf{q}) = \mathbf{p}^\intercal \mathbf{A} \mathbf{q}$ and $\mu_2(\mathbf{p}, \mathbf{q}) = \mathbf{p}^\intercal \mathbf{B} \mathbf{q}$

$\mathbf{A}$	A	B	C
A	0.0	4.6	-4.6
B	-4.6	0.0	4.6
C	4.6	-4.6	0.0

- Two player pick "teams " of agents ,p,q correspond to the mixed strategy distribution

$\mathbf{B} = \mathbf{A}^\intercal$. The game is symmetric because $\mathbf{B} = \mathbf{A}^\intercal$ and zero-sum because $\mathbf{B} = -\mathbf{A}$.

Nash averaging

- Nash equilibria are teams that are unbeatable in expectation
  - In rock—paper—scissors, the only unbeatable—on—average team is the uniform distribution.

- A problem with Nash equilibria (NE) is that they are not unique for zero—sum game.

Maxent Nash Evaluation

- Fortunately, for zero-sum games there is a natural choice of Nash:

Proposition 4 (maxent NE). *For antisymmetric* $\mathbf{A}$ *there is a* unique symmetric *Nash equilibrium* $(\mathbf{p}^*, \mathbf{p}^*)$ *solving* $\max_{\mathbf{p}\in\Delta_n} \min_{\mathbf{q}\in\Delta_n} \mathbf{p}^\intercal \mathbf{A}\mathbf{q}$ *with greater entropy than any other Nash equilibrium.*

- The maxent Nash evaluation method

Definition 2. *The* ***maxent Nash evaluation method*** *for AvA is*

$$\mathcal{E}_m : \{\textit{evaluation data}\} = \{\textit{antisymmetric matrices}\} \xrightarrow{\textit{maxent NE}} \left[\{\textit{players}\} \xrightarrow{\textit{Nash average}} \mathbb{R}\right],$$

where $\mathbf{p}^*_\mathbf{A}$ *is the* ***maxent Nash equilibrium*** *and* $\mathbf{n}_\mathbf{A} := \mathbf{A} \cdot \mathbf{p}^*_\mathbf{A}$ *is the* ***Nash average.***

Interpretable

Interpretable: ***(i)*** *The maxent NE on* $\mathbf{A}$ *is the uniform distribution,* $\mathbf{p}^* = \frac{1}{n}\mathbf{1}$*, iff the meta-game is* cyclic*, i.e.* $\mathrm{div}(\mathbf{A}) = \mathbf{0}$*.* ***(ii)*** *If the meta-game is* transitive*, i.e.* $\mathbf{A} = \mathrm{grad}(\mathbf{r})$*, then the maxent NE is the uniform distribution on the player(s) with highest rating(s) – there could be a tie.*

$\mathbf{A}$	A	B	C
A	0.0	4.6	-4.6
B	-4.6	0.0	4.6
C	4.6	-4.6	0.0

$$\mathbf{p}^*_{\mathbf{A}} = (\tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3})$$

$$\mathbf{n}_{\mathbf{A}} = \mathbf{0}_{3\times 1}$$

Interpretable

$$\mathbf{C} = \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{T} = \begin{pmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \\ -2 & -1 & 0 \end{pmatrix}$$

The maxent Nash equilibria and Nash averages of $\mathbf{C} + \epsilon\mathbf{T}$ are

$$\mathbf{p}^*_{\mathbf{C}+\epsilon\mathbf{T}} = \begin{cases} \left(\frac{1+\epsilon}{3}, \frac{1-2\epsilon}{3}, \frac{1+\epsilon}{3}\right) & \text{if } 0 \le \epsilon \le \frac{1}{2} \\ (1,0,0) & \text{if } \frac{1}{2} < \epsilon \end{cases} \quad \text{and} \quad \mathbf{n}_{\mathbf{C}+\epsilon\mathbf{T}} = \begin{cases} (0,0,0) & 0 \le \epsilon \le \frac{1}{2} \\ (0,-1-\epsilon,1-2\epsilon) & \frac{1}{2} < \epsilon \end{cases}$$

- $\epsilon > \frac{1}{2}$, $C + \epsilon T$ is transitive, The first one has the largest Nash probability and Nash averages

Evaluation of the Environment

- Agent vs agent (AvA), where agents compete directly as in Go and Starcraft.
  - ELO, Glicko, TruesikII, Nash Averaging

- Agent vs task (AvT), where algorithms are evaluated on suites of datasets or environments as in Atari
  - How should environments be evaluated?
  - How should agents be evaluated?
- Nash averaging can compute which tasks and agents do and do not matter by a meta-game .
  - Using tasks to evaluate ability of agents
  - Using agents to evaluate difficulty of tasks

Evaluation of the Environment

Definition 1. *An **evaluation method** maps data to a real-valued function on players (that is, agents or agents and tasks):*

$$\mathcal{E} : \{\textit{evaluation data}\} = \{\textit{antisymmetric matrices}\} \to \Big[\{\textit{players}\} \to \mathbb{R}\Big].$$

- AvA Logit(P)
- AVT

$$\mathbf{A}_{(m+n)\times(m+n)} = \operatorname{grad}(\mathbf{r}) + \underbrace{\begin{pmatrix} \mathbf{0}_{m\times m} & \tilde{\mathbf{S}}_{m\times n} \\ -\tilde{\mathbf{S}}^{\intercal}_{n\times m} & \mathbf{0}_{n\times n} \end{pmatrix}}_{\tilde{\mathbf{A}}} = \begin{pmatrix} \operatorname{grad}(\mathbf{s}) & \mathbf{S}_{m\times n} \\ -\mathbf{S}^{\intercal}_{n\times m} & \operatorname{grad}(\mathbf{d}) \end{pmatrix}.$$

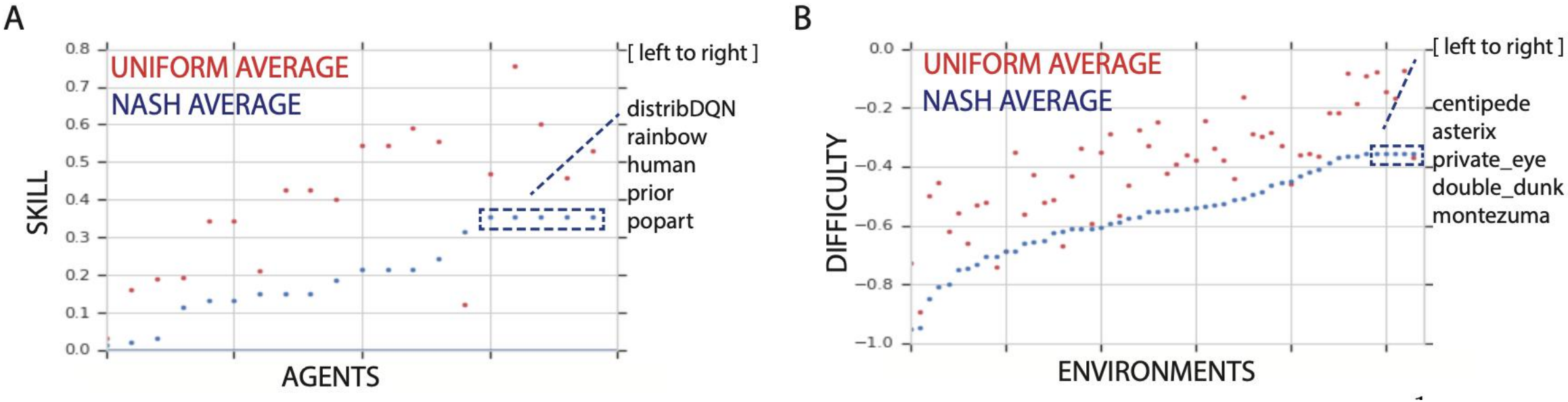
compares agents by their average skill on tasks

compares tasks by their average difficulty for agents

An (m × n) matrix S: rows are agents, columns are tasks, entries are scores (e.g. accuracy or total reward)

Evaluating Agents and Environments

- Atari : The 20 agents evaluated on 54 environments are represented by matrix $S_{20 \times 54}$



	Nash probability	Nash average	Uniform average
DQN_(w/o_MC)	0.000	0.030	0.189
DQN_(with_MC)	0.000	0.115	0.191
DQN-PixelCNN_(w/o_MC)	0.000	0.022	0.161
DQN-PixelCNN_(with_MC)	0.000	0.148	0.212
DQN	0.000	0.132	0.343
A3C	0.000	0.149	0.426
DDQN	0.000	0.244	0.556
PriorDDQN	0.000	0.213	0.543
DuelDDQN	0.034	0.354	0.600
DistribDQN	0.000	0.185	0.400
NoisyDQN	0.297	0.354	0.755
Rainbow	0.000	0.314	0.122
RANDOM	0.000	0.012	0.032
HUMAN	0.328	0.354	0.470
DQN_	0.000	0.132	0.343
DDQN_	0.000	0.149	0.426
DUEL	0.000	0.213	0.543
PRIOR	0.135	0.354	0.529
PRIOR_DUEL	0.000	0.214	0.590
PopArt	0.206	0.354	0.457

Figure 6: Evaluation of agents. Note, there are redundancies since agents are taken from multiple papers; these are ignored by Nash averaging.

Conclusion

- *Maxent* entropy Nash equilibrium can obtain the agents with the strongest ability(the most difficult question), whose probability is greater than 0, and has the maximum Nash average.

- This method can discover the existence of circular games when there are multiple maximum Nash averages.

- Unlike ELO, this approach only finds the most valuable set of agents but cannot rank all players

α −Rank

Given match outcomes for a K-player game, α-Rank computes rankings as follows:

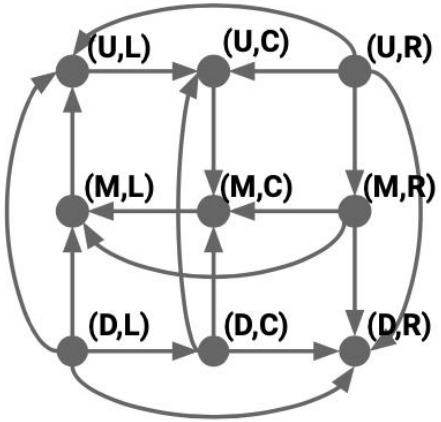
1. Construct meta-payoff tables $\mathbf{M}^k$ for each player $k \in \{1, \ldots, K\}$ (e.g., by using the win/loss ratios for the different strategy/agent match-ups as payoffs)
2. Compute the transition matrix $\mathbf{C}$, as detailed in Section 2
3. Compute the stationary distribution, $\boldsymbol{\pi}$, of $\mathbf{C}$
4. Compute the agent rankings by ordering the masses of $\boldsymbol{\pi}$

α −Rank

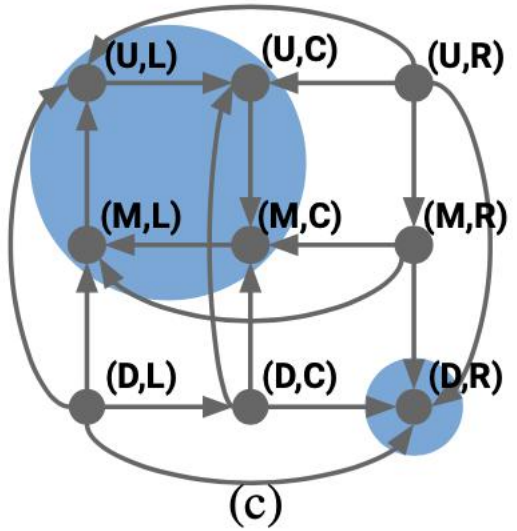
		II		
		L	C	R
I	U	2, 1	1, 2	0, 0
	M	1, 2	2, 1	1, 0
	D	0, 0	0, 1	2, 2

(a)

- Response Graph
  - (U,L)−>(U,C) player2's payoff 1−>2
  - Only change one player
- Markov−Conley chains(MCCs)
  - The sink strongly connected components(SSCC) of the response graph.



(b)



(c)

Transition Matrix C

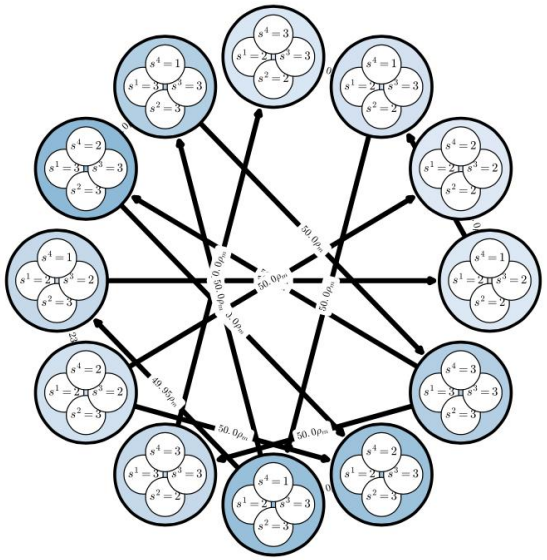
- Irreducible Markov chain ——>Unique invariant distribution π ——>Strategy profile rankings

$$\mathbf{C}_{s,\sigma} = \begin{cases} \eta \frac{1-\exp\left(-\alpha\left(\mathbf{M}^k(\sigma)-\mathbf{M}^k(s)\right)\right)}{1-\exp(-\alpha m(\mathbf{M}^k(\sigma)-\mathbf{M}^k(s)))} & \text{if } \mathbf{M}^k(\sigma) \neq \mathbf{M}^k(s) \\ \frac{\eta}{m} & \text{otherwise}, \end{cases} \quad \text{and} \quad \mathbf{C}_{s,s} = 1 - \sum_{\substack{k\in[K] \\ \sigma|\sigma^k \in S^k \setminus \{s^k\}}} \mathbf{C}_{s,\sigma},$$

- Large values of α corresponding to higher *selection pressure* in the evolutionary model considered.
- α is either set to a large but finite value, or a perturbed version of C under the infinite—α limit is used.

Ranking

- The stationary distribution indicating the average amount of time individuals in the underlying evolutionary model spend playing each strategy profile.



(a)

Agent	Rank	Score
(3,3,3,2)	1	0.08
(2,3,3,1)	2	0.07
(2,3,3,2)	3	0.07
(3,3,3,1)	4	0.06
(3,3,3,3)	5	0.06
(3,2,3,3)	6	0.05
(2,3,2,1)	7	0.04
(2,3,2,2)	8	0.04
(2,2,3,1)	9	0.04
(2,2,3,3)	10	0.03
(2,2,2,1)	11	0.03
(2,2,2,2)	12	0.03
...	...	...

(c)

Outlines

- Problem background：problem statement，importance
- Classical methods：ELO，Glicko，TrueSkill
- Improved methods：mELO，Nash averaging，α-Rank
- Sampling complexity analysis
- Challenges

Incomplete [4] "Multiagent evaluation under incomplete information." nips. 2019.

- Nash averaging and α-Rank assume noise-free (complete information, payoff matrix)

- The exact payoff table M is rarely known;
- An empirical payoff table $\hat{M}$ is typically constructed from observed agent interactions.

Sample complexity guarantees

Theorem 3.1 (Finite-α). *Suppose payoffs are bounded in the interval* $[-M_{\max}, M_{\max}]$, *and define* $L(\alpha, M_{\max}) = 2\alpha \exp(2\alpha M_{\max})$ *and* $g(\alpha, \eta, m, M_{\max}) = \eta \frac{\exp(2\alpha M_{\max})-1}{\exp(2\alpha m M_{\max})-1}$. *Let* $\varepsilon \in (0, 18 \times 2^{-|S|} \sum_{n=1}^{|S|-1} \binom{|S|}{n} n^{|S|})$, $\delta \in (0,1)$. *Let* $\hat{\mathbf{M}}$ *be an empirical payoff table constructed by taking* N_s *i.i.d. interactions of each strategy profile* $s \in S$. *Then the invariant distribution* $\hat{\boldsymbol{\pi}}$ *derived from the empirical payoff matrix* $\hat{\mathbf{M}}$ *satisfies* $\max_{s \in \prod_k S^k} |\pi(s) - \hat{\pi}(s)| \leq \varepsilon$ *with probability at least* $1 - \delta$, *if*

$$N_s > \frac{648 M_{\max}^2 \log(2|S|K/\delta) L(\alpha, M_{\max})^2 \left(\sum_{n=1}^{|S|-1} \binom{|S|}{n} n^{|S|} \right)^2}{\varepsilon^2 g(\alpha, \eta, m, M_{\max})^2} \qquad \forall s \in S\,.$$

Theorem 3.2 (Infinite-α). *Suppose all payoffs are bounded in* $[-M_{\max}, M_{\max}]$, *and that* $\forall k \in [K]$ *and* $\forall s^{-k} \in S^{-k}$, *we have* $|\mathbf{M}^k(\sigma, s^{-k}) - \mathbf{M}^k(\tau, s^{-k})| \geq \Delta$ *for all distinct* $\sigma, \tau \in S^k$, *for some* $\Delta > 0$. *Let* $\delta > 0$. *Suppose we construct an empirical payoff table* $(\hat{\mathbf{M}}^k(s) \mid k \in [K], s \in S)$ *through* N_s *i.i.d games for each strategy profile* $s \in S$. *Then the transition matrix* $\hat{\mathbf{C}}$ *computed from payoff table* $\hat{\mathbf{M}}$ *is exact (and hence all MCCs are exactly recovered) with probability at least* $1 - \delta$, *if*

$$N_s > 8\Delta^{-2} M_{\max}^2 \log(2|S|K/\delta) \qquad \forall s \in S\,.$$

Bounds for Elo

Theorem C.1. *Consider a symmetric, two-player win-loss game with finite strategy set S^1 and payoff matrix $\mathbf{M}$. Let $\mathbf{q}$ be the fitted payoffs obtained from the BatchElo model on the payoff matrix $\mathbf{M}$, and let $\hat{\mathbf{q}}$ be the fitted payoffs obtained from the BatchElo model on an empirical payoff table $\hat{\mathbf{M}}$, based on $N_{s,s'}$ interactions between each pair of strategies s, s'. If we take, for each pair of strategy profiles $s, s' \in S^1$, a number of interactions $N_{s,s'}$ satisfying*

$$N_{s,s'} > 0.5|S^1|^2\varepsilon^{-2}\log(|S^1|^2/\delta)\,. \tag{3}$$

Then it follows that with probability at least $1-\delta$,

$$\left|\sum_{s'}(\mathbf{q}_{s,s'} - \hat{\mathbf{q}}_{s,s'})\right| < \varepsilon \qquad \forall s \in S^1. \tag{4}$$

[0, 1]

the following form of Hoeffding's inequality: Let $X_1, \ldots, X_N$ be i.i.d. random variables supported on $[a, b]$. Let $\varepsilon > 0$ and $\delta > 0$. Then for $N > (b-a)^2\log(2/\delta)/(2\varepsilon^2)$, we have

$$\mathbb{P}\left(\left|\tfrac{1}{N}\sum_{n=1}^{N} X_n - \mathbb{E}[X_1]\right| > \varepsilon\right) < \delta.$$

Outlines

- Problem background：problem statement，importance
- Classical methods：ELO，Glicko，TrueSkill
- Improved methods：mELO，Nash averaging，α-Rank
- Sampling complexity analysis
- Challenges

Challenges

- Efficient
  - Simple calculation
  - Less sample
  - Incremental
- Robust
  - Small perturbation
  - redundant
- Validity
  - Correct rank
  - Adversarial attack
- General
  - General-sum game
  - Multi-player
  - Cooperative

Reference

[1] Re-evaluating Evaluation，NIPS 2018

[2] A. E. Elo, The Rating of Chess players, Past and Present. Ishi Press International, 1978.

[3] R. Herbrich, T. Minka, and T. Graepel, “TrueSkill: a Bayesian skill rating system,” in NIPS, 2007.

[4] Ro"Multiagent evaluation under incomplete information." nips. 2019.

[5] Jordan, Scott M., et al. "Evaluating the Performance of Reinforcement Learning Algorithms." arXiv preprint arXiv:2006.16958 (2020).

[6] Morrison, Breanna. "Comparing Elo, Glicko, IRT, and Bayesian IRT Statistical Models for Educational and Gaming Data." (2019).