

Top-k Selection based on Adaptive Sampling of Noisy Preferences

Róbert Busa-Fekete^{1,2} Balázs Szörényi^{2,3} Paul Weng⁴
Weiwei Cheng¹ Eyke Hüllermeier¹

¹Computational Intelligence Group, Philipps University Marburg, Marburg, Germany

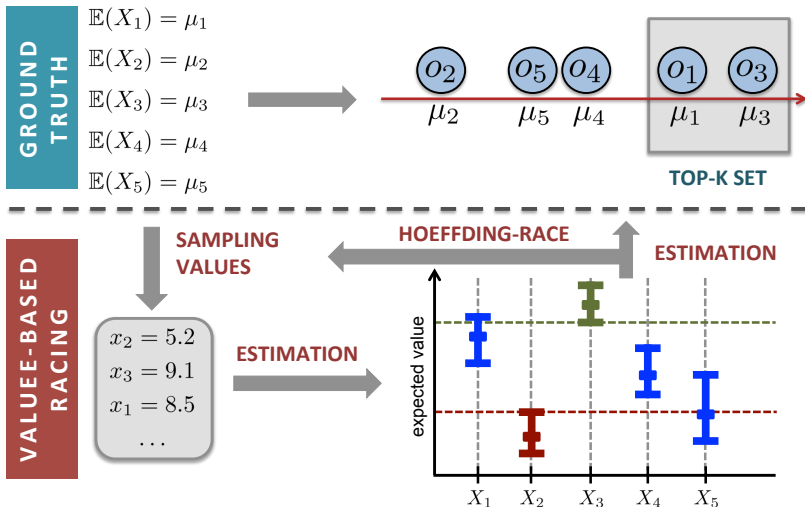
²Research Group on Artificial Intelligence, Hungarian Academy of Sciences and University of Szeged, Szeged, Hungary

³INRIA Lille - Nord Europe, Sequel project, Villeneuve d'Ascq, France

⁴Laboratory of Computer Science of Paris 6, University Pierre and Marie Curie, Paris, France

June 19, 2013, ICML

Value-based Top-k Selection (TKS) (Maron&Moore, 1994)

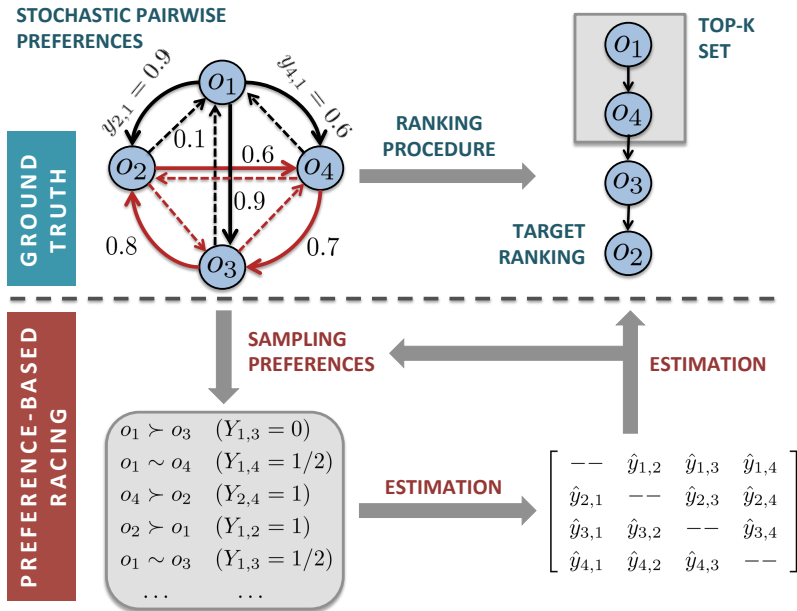


Preference-based Top-k Selection (TKS)

- ▶ Pairwise preferences over the set of options
- ▶ Four possible outcomes of a single pairwise comparison between o_i and o_j :
 - ▶ $o_i \prec o_j \Rightarrow Y_{i,j} = 0$ ((strict) preference for o_j)
 - ▶ $o_i \succ o_j \Rightarrow Y_{i,j} = 1$ ((strict) preference for o_i)
 - ▶ $o_i \sim o_j \Rightarrow Y_{i,j} = 1/2$ (indifference)
 - ▶ $o_i \perp o_j \Rightarrow Y_{i,j} = 1/2$ (incomparability)
- ▶ $y_{i,j} = \mathbb{E}[Y_{i,j}]$
 - ▶ if $y_{i,j} > 1/2$ then o_i is preferred to o_j
 - ▶ It can be estimated on the basis of a finite sample

$$y_{i,j} \approx \bar{y}_{i,j} = \frac{1}{n} \sum_{\ell=1}^n y_{i,j}^{\ell}$$

Preference-based Top-k Selection (TKS)



Ranking procedures

$$y_{i,j} = \mathbb{E}[Y_{i,j}]$$

1. **Copeland's ranking**: (CO): $o_i \prec^{\text{CO}} o_j$ if and only if $d_i < d_j$,
where

$$d_i = \#\{k \in [K] \mid 1/2 < y_{i,k}\} ,$$

- ▶ An option o_i is preferred to o_j whenever o_i “beats” more options than o_j does.

2. **Sum of expectations** (SE) ranking: $o_i \prec^{\text{SE}} o_j$ if and only if

$$y_i = \frac{1}{K-1} \sum_{k \neq i} y_{i,k} < \frac{1}{K-1} \sum_{k \neq j} y_{j,k} = y_j .$$

3. The idea of the **Random walk** (RW) ranking is to handle the matrix $\mathbf{Y} = [y_{i,j}]_{K \times K}$ as a transition matrix \mathbf{S} of a Markov chain and order the options based on its stationary distribution.

Thanks!

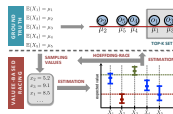
Top-k Selection based on Adaptive Sampling of Noisy Preferences

Róbert Busa-Fekete^{1,2} Balazs Székely^{2,3} Paul Weng⁴ Weiwei Cheng¹ Eyle Hüllermeier²

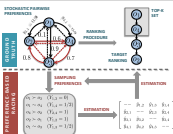
¹Computational Intelligence Group, Philipps University Marburg, Germany
²Research Group on Artificial Intelligence, Hungarian Academy of Sciences and University of Szeged, Szeged, Hungary
³MPSI Lab, Sorbonne University, Paris, France
⁴University of Computer Science, University of Paris, France



Value-based Top-k Selection (TKS) [1]



Preference-based Top-k Selection (TKS)



Resolving Preferential Cycles

1. Copeland's ranking: $\text{CO} = \{x_i\}_{i=1}^K$ if and only if $d_i < d_j$ where $d_i = \pi(|\{k \in [K] : x_i \succ x_k\}|/2 + \kappa)$.
2. Sum of expectations (SE) ranking: $\text{SE} = \{x_i\}_{i=1}^K$ if and only if $\sum_{j=1}^K \mathbb{E}[x_i \succ x_j] > \sum_{j=1}^K \mathbb{E}[x_j \succ x_i]$.
3. The idea of the Random walk (RW) ranking is to handle the matrix \mathbf{V} as a transition matrix \mathbf{S} of a Markov chain and order the options based on its stationary distribution.

Theorem (Expected sample complexity for Copeland's ranking)

Let $\mathcal{O} = \{o_1, \dots, o_K\}$ be a set of options such that $d_{o_i} = |o_i| - 1/2 \neq 0$ for all $i, j \in [K]$. The expected number of pairwise comparison times by PTH-CO is bounded by

$$\sum_{i=1}^K \sum_{j=1}^K \left(\frac{1}{d_{o_i} d_{o_j}} \log \frac{2K^2 n_{\text{max}}}{d_{o_i} d_{o_j}} \right).$$

Moreover, the probability that no optimal solution is found by PTH-CO is at most δ if $n_{\text{max}} \geq \frac{1}{\delta} \log \frac{2K^2 n_{\text{max}}}{\delta}$ for all $i, j \in [K]$.

References

- [1] M. Busa-Fekete, A. Hüllermeier, and P. Weng. Accelerating model selection search for classification and function approximation. *NIPS*, pp. 30-40 (2009).
- [2] Adam, J.A. and Shalizi, C.R. General bounds on statistical query learning and PAC learning with noise via hypothesis learning. *Inf. Comput.* 183(2):85-110 (2006).
- [3] Feinberg, E.E. and Meyer, C.D. Smoothness of the stationary distribution vector for an ergodic Markov chain. *Linear Algebra and its Applications* 183(1):51-57 (1993).

Sampling Strategies

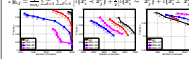
- Algorithm PTH ($\mathbf{V}_{1:n}, \dots, \mathbf{V}_{K:n}, \kappa, n_{\text{max}}, \delta$)
- 1. Initially sample each \mathbf{V}_i .
- 2. In each iteration, calculate $\hat{d}_i = \frac{1}{n} \sum_{j=1}^n \hat{v}_{ij}$ and its confidence interval $[\hat{d}_i - c_{i,n}, \hat{d}_i + c_{i,n}]$ with $c_{i,n} = \sqrt{\frac{\log 2/\delta}{n}}$.
- 3. and decide which \mathbf{V}_i will be sampled next based on the specific ranking procedure of interest.
- Copeland's ranking: $1/2 \leq \hat{d}_i - c_{i,n} < \hat{d}_j + c_{j,n}$
- Sum of expectations (SE): $\hat{d}_i = \frac{1}{K-1} \sum_{j=1, j \neq i}^K (\hat{d}_i - c_{i,n}) < \hat{d}_j + c_{j,n}$ where $\hat{d}_i = \frac{1}{K-1} \sum_{j=1, j \neq i}^K \hat{v}_{ij}$ and $\hat{d}_j = \frac{1}{K-1} \sum_{i=1, i \neq j}^K \hat{v}_{ji}$
- Random walk (RW) ranking: transform \mathbf{V} into stochastic matrix \mathbf{S} where $S_{ij} = \frac{1}{K-1} \sum_{k=1, k \neq i}^K \hat{v}_{ik}$ where $S_{ij} = \frac{1}{K-1} \sum_{k=1, k \neq i}^K \hat{v}_{ik}$ (for $i \neq j$)
- Let $S_{ii} = 1 - \sum_{j=1, j \neq i}^K S_{ij}$. Then, according to [3], we have $\|x - y\|_{\text{max}} \leq \max_i \sum_{j=1}^K |x_j - y_j| \leq \max_i \sum_{j=1}^K |x_j - y_j|$

Experiments: Bundesliga

- soccer matches of the last ten seasons from the German Bundesliga
- uniform sampling as baseline
- $\delta = 0.1, \kappa = 3, n_{\text{max}} \in \{100, 500, 1000, 5000, 10000\}$

A Special Case

- Each option x_i is associated with a random variable X_i .
- The random variables X_i take values in a set Ω that is *partially ordered* by a preference relation \succ .
- $\mathbb{W}_i = \mathbb{P}(X_i < X_j) + \frac{1}{2}(\mathbb{P}(X_i = X_j) + \mathbb{P}(X_j = X_i))$
- $\mathbb{W}_i = \frac{1}{2} \sum_{j=1}^K \sum_{\omega \in \Omega} [\mathbb{I}(\omega_i < \omega_j) + \frac{1}{2}(\mathbb{I}(\omega_i = \omega_j) + \mathbb{I}(\omega_j = \omega_i))]$



Theorem (Expected sample complexity for SE ranking)

Let $\mathcal{O} = \{o_1, \dots, o_K\}$ be a set of options. Assume $d_{o_i} < d_{o_j}$ for all $i < j \leq K$. Let $\hat{d}_i = \frac{1}{K-1} \sum_{j=1, j \neq i}^K \hat{v}_{ij}$ for $i \in [K-1]$ and $\hat{d}_K = \frac{1}{K-1} \sum_{j=1}^{K-1} \hat{v}_{jK}$ for $j = K - \kappa + 1, \dots, K$. Then, whenever $n_{\text{max}} \geq \log_{1-\delta} \frac{1}{\delta} \log \frac{2K^2 n_{\text{max}}}{\delta}$, PTH-SE terminates after $\sum_{i=1}^K \sum_{j=1}^K \log \frac{1}{\delta} \log \frac{2K^2 n_{\text{max}}}{\delta} \log \frac{1}{\delta} \log \frac{2K^2 n_{\text{max}}}{\delta}$ pairwise comparisons and outputs the optimal solution with probability at least $(1 - \delta)$.