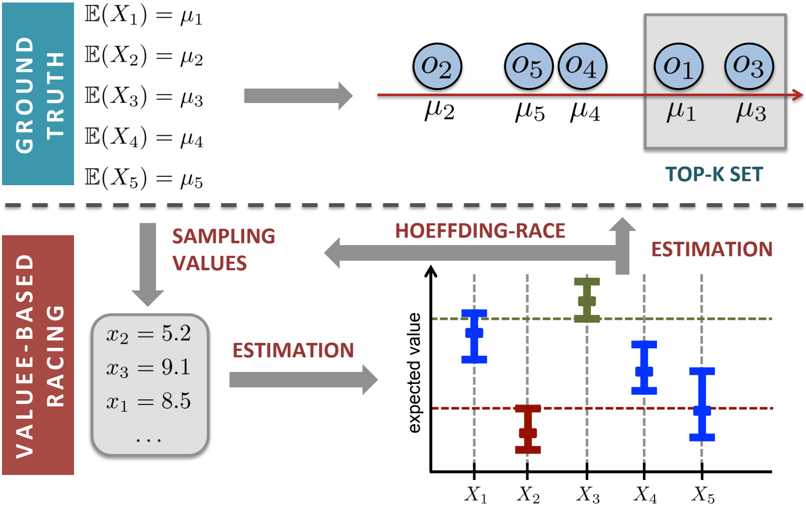
Top-k Selection based on Adaptive Sampling of Noisy Preferences

Róbert Busa-Fekete[1,2] Balázs Szörényi[2,3] Paul Weng[4]
Weiwei Cheng[1] Eyke Hüllermeier[1]

[1]Computational Intelligence Group, Philipps University Marburg, Marburg, Germany
[2]Research Group on Artificial Intelligence, Hungarian Academy of Sciences and University of Szeged, Szeged, Hungary
[3]INRIA Lille - Nord Europe, SequeL project, Villeneuve d'Ascq, France
[4]Laboratory 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)

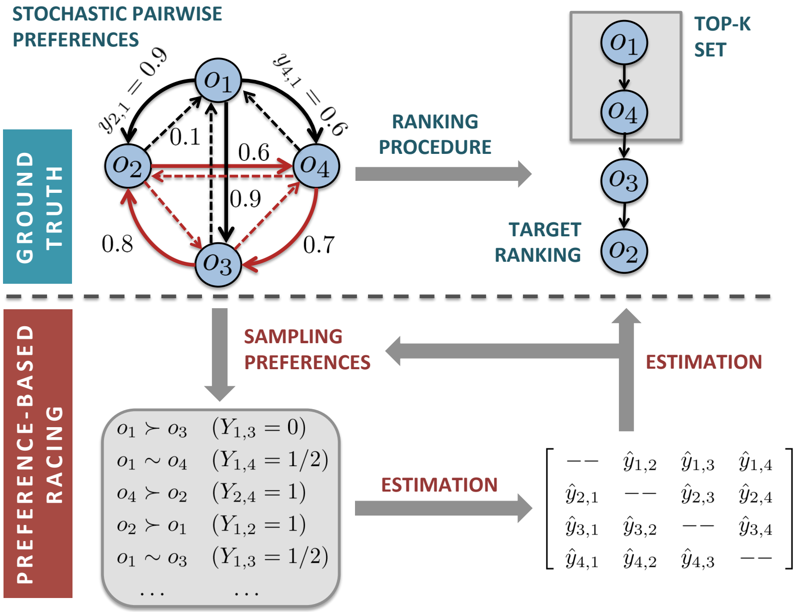


Preference-based Top-k Selection (TKS)

- 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)
- $y_{i,j} = \mathbb{E}\left[Y_{i,j}\right]$
  - 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): $o_i \prec^{\mathrm{CO}} o_j$ if and only if $d_i < d_j$, where

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

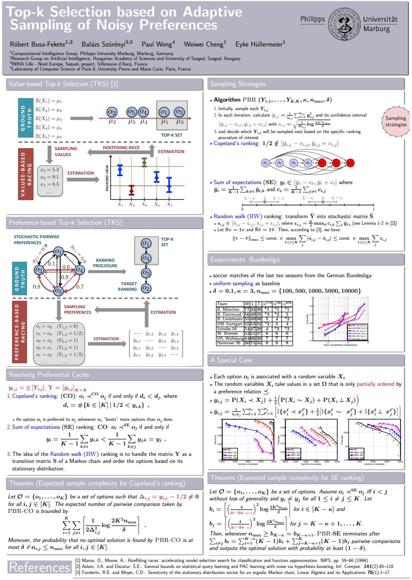
   - An option o_i is preferred to o_j whenever o_i "beats" more options than o_j does.

2. Sum of expectations (SE) ranking: $o_i \prec^{\mathrm{SE}} o_j$ 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
Philipps Universität Marburg
Róbert Busa-Fekete[1,2] Balázs Szörényi[2,3] Paul Weng[4] Weiwei Cheng[1] Eyke Hüllermeier[1]
[1]Computational Intelligence Group, Philipps University Marburg, Marburg, Germany
[2]Research Group on Artificial Intelligence, Hungarian Academy of Sciences and University of Szeged, Szeged, Hungary
[3]INRIA Lille - Nord Europe, SequeL project, Villeneuve d'Ascq, France
[4]Laboratory of Computer Science of Paris 6, University Pierre and Marie Curie, Paris, France
Value-based Top-k Selection (TKS) [1]
GROUND TRUTH
TOP-K SET
VALUE-BASED RACING
SAMPLING VALUES
HOEFFDING-RACE
ESTIMATION
Preference-based Top-k Selection (TKS)
STOCHASTIC PAIRWISE PREFERENCES
RANKING PROCEDURE
TARGET RANKING
PREFERENCE-BASED RACING
SAMPLING PREFERENCES
Resolving Preferential Cycles
Theorem (Expected sample complexity for Copeland's ranking)
Sampling Strategies
Sampling strategies
Experiments: Bundesliga
soccer matches of the last ten seasons from the German Bundesliga
uniform sampling as baseline
A Special Case
Theorem (Expected sample complexity for SE ranking)
References
[1] Maron, O., Moore, A.: Hoeffding races: accelerating model selection search for classification and function approximation. NIPS, pp. 59–66 (1994)
[2] Aslam, J.A. and Decatur, S.E.: General bounds on statistical query learning and PAC learning with noise via hypothesis boosting, Inf. Comput. 141(2):85–118
[3] Funderlic, R.E. and Meyer, C.D.: Sensitivity of the stationary distribution vector for an ergodic Markov chain, Linear Algebra and its Applications 76(1):1–17