Top-k Selection based on Adaptive Sampling of Noisy Preferences

Róbert Busa-Fekete[1,2] busarobi@inf.u-szeged.hu
Balázs Szörényi[2,3] szorenyi@inf.u-szeged.hu
Paul Weng[4] paul.weng@lip6.fr
Weiwei Cheng[1] cheng@mathematik.uni-marburg.de
Eyke Hüllermeier[1] eyke@mathematik.uni-marburg.de

[1]Mathematics and Computer Science, University of Marburg, Hans-Meerwein-Str., 35032 Marburg, Germany
[2]Research Group on Artificial Intelligence, Hungarian Academy of Sciences and University of Szeged, Hungary
[3]INRIA Lille - Nord Europe, SequeL project, 40 avenue Halley, 59650 Villeneuve d'Ascq, France
[4]Pierre and Marie Curie University (UPMC), 4 place Jussieu, 75005 Paris, France

Abstract

We consider the problem of reliably selecting an optimal subset of fixed size from a given set of choice alternatives, based on noisy information about the quality of these alternatives. Problems of similar kind have been tackled by means of adaptive sampling schemes called racing algorithms. However, in contrast to existing approaches, we do not assume that each alternative is characterized by a real-valued random variable, and that samples are taken from the corresponding distributions. Instead, we only assume that alternatives can be compared in terms of pairwise preferences. We propose and formally analyze a general preference-based racing algorithm that we instantiate with three specific ranking procedures and corresponding sampling schemes. Experiments with real and synthetic data are presented to show the efficiency of our approach.

1. Introduction

Consider the problem of selecting the best κ out of K random variables with high probability on the basis of finite samples, assuming that random variables are ranked based on their expected value. A natural way of approaching this problem is to apply an adaptive sampling strategy, called *racing algorithm*, which makes use of confidence intervals derived from the concentration property of the mean estimate (Hoeffding, 1963). This formal setup was first considered by Maron & Moore (1994) and is now used in many practical applications, such as model selection (Maron & Moore, 1997), large-scale learning (Mnih et al., 2008) and policy search in MDPs (Heidrich-Meisner & Igel, 2009).

Motivated by recent work on learning from qualitative or implicit feedback, including preference learning in general (Fürnkranz & Hüllermeier, 2011) and preference-based reinforcement learning in particular (Akrour et al., 2011; Cheng et al., 2011), we introduce and analyze a *preference-based* generalization of the *value-based* setting of the above selection problem, subsequently denoted TKS (short for Top-k Selection) problem: Instead of assuming that the decision alternatives or *options* $\mathcal{O} = \{o_1, \ldots, o_K\}$ are characterized by real values (namely expectations of random variables) and that samples provide information about these values, we only assume that the options can be compared in a pairwise manner. Thus, a sample essentially informs about pairwise preferences, i.e., whether or not an option o_i might be preferred to another one o_j (written $o_i \succ o_j$).

An important observation is that, in this setting, the original goal of finding the top-κ options is no longer well-defined, simply because pairwise comparisons can be cyclic. Therefore, to make the specification of our problem complete, we add a *ranking procedure* that turns a pairwise preference relation into a complete preorder of the options $\mathcal{O}$. The goal is then to find the top-κ options according to that order. More concretely, we consider Copeland's ranking (binary voting), the sum of expectations (weighted voting) and the random walk ranking (PageRank) as *target rankings*. For each of these ranking models, we devise proper sampling strategies that constitute the core of our preference-based racing algorithm.

Proceedings of the 30th International Conference on Machine Learning, Atlanta, Georgia, USA, 2013. JMLR: W&CP volume 28. Copyright 2013 by the author(s).

After detailing the problem setting in Section 2, we introduce a general preference-based racing algorithm in Section 3 and analyze sampling strategies for different ranking methods in Section 4. In Section 5, a first experimental study with sports data is presented, and in Section 6, we consider a special case of our setting that is close to the original value-based one. Related work is discussed in Section 7.

2. Problem Setting and Terminology

In this section, we first recapitulate the original value-based setting of the TKS problem and then introduce our preference-based generalization.

2.1. Value-based TKS

Consider a set of decision alternatives or options $\mathcal{O} = \{o_1, \ldots, o_K\}$, where each option o_i is associated with a random variable X_i. Let $F_1, \ldots, F_K$ denote the (unknown) distribution functions of $X_1, \ldots, X_K$, respectively, and $\mu_i = \int x dF_i(x)$ the corresponding expected values (supposed to be finite).

The TKS task consists of selecting, with a predefined confidence $1-\delta$, the $\kappa < K$ options with highest expectations. In other words, one seeks an index set $I \subseteq [K] = \{1, \ldots, K\}$ of cardinality κ maximizing $\sum_{i \in I} \mu_i$, which is formally equivalent to the following optimization problem:

$$\operatorname*{argmax}_{I \subseteq [K]:\ |I|=\kappa} \sum_{i \in I} \sum_{j \neq i} \mathbb{I}\{\mu_j < \mu_i\} \ , \tag{1}$$

where $\mathbb{I}\{\cdot\}$ is the indicator function which is 1 if its argument is true and 0 otherwise. This selection problem must be solved on the basis of random samples drawn from $X_1, \ldots, X_K$.

2.2. Preference-based TKS

Our point of departure is pairwise preferences over the set $\mathcal{O}$ of options. In the most general case, one typically allows four possible outcomes of a single pairwise comparison between o_i and o_j, namely (strict) preference for o_i, (strict) preference for o_j, indifference and incomparability. They are denoted by $o_i \succ o_j$, $o_i \prec o_j$, $o_i \sim o_j$ and $o_i \perp o_j$, respectively.

To make ranking procedures applicable, these pairwise outcomes need to be turned into numerical scores. We consider the outcome of a comparison between o_i and o_j as a random variable $Y_{i,j}$ which assumes the value 1 if $o_i \succ o_j$, 0 if $o_i \prec o_j$, and 1/2 otherwise. Thus, indifference and incomparability are handled in the same way, namely by giving half a point to both options. Essentially, this means that these outcomes are treated in a neutral way.

Based on a set of realizations $\{y^1_{i,j}, \ldots, y^n_{i,j}\}$ of $Y_{i,j}$, assumed to be independent, the expected value $y_{i,j} = \mathbb{E}[Y_{i,j}]$ of $Y_{i,j}$ can be estimated by the mean

$$\bar{y}_{i,j} = \frac{1}{n} \sum_{\ell=1}^{n} y^\ell_{i,j} \ . \tag{2}$$

A ranking procedure $\mathcal{A}$ (concrete choices of $\mathcal{A}$ will be discussed in the next section) produces a complete preorder $\preceq^{\mathcal{A}}$ of the options $\mathcal{O}$ on the basis of the relation $\mathbf{Y} = [y_{i,j}]_{K \times K} \in [0,1]^{K \times K}$. In analogy to (1), our preference-based TKS task can then be defined as selecting a subset $I \subset [K]$ such that

$$\operatorname*{argmax}_{I \subseteq [K]:\ |I|=\kappa} \sum_{i \in I} \sum_{j \neq i} \mathbb{I}\{o_j \prec^{\mathcal{A}} o_i\} \ , \tag{3}$$

where $\prec^{\mathcal{A}}$ denotes the strict part of $\preceq^{\mathcal{A}}$. More specifically, the optimality of the selected subset should be guaranteed with probability at least $1-\delta$.

2.3. Ranking Procedures

In the following, we introduce three instantiations of the ranking procedure $\mathcal{A}$, starting with Copeland's ranking (CO); it is defined as follows (Moulin, 1988): $o_i \prec^{\mathrm{CO}} o_j$ if and only if $d_i < d_j$, where $d_i = \#\{k \in [K] \,|\, 1/2 < y_{i,k}\}$. The interpretation of this relation is very simple: An option o_i is preferred to o_j whenever o_i "beats" more options than o_j does.

The sum of expectations (SE) ranking is a "soft" version of CO: $o_i \prec^{\mathrm{SE}} o_j$ if and only if

$$y_i = \frac{1}{K-1} \sum_{k \neq i} y_{i,k} < \frac{1}{K-1} \sum_{k \neq j} y_{j,k} = y_j \ . \tag{4}$$

The idea of the random walk (RW) ranking is to handle the matrix $\mathbf{Y}$ as a transition matrix of a Markov chain and order the options based on its stationary distribution. More precisely, RW first transforms $\mathbf{Y}$ into the stochastic matrix $\mathbf{S} = [s_{i,j}]_{K \times K}$ where $s_{i,j} = y_{i,j} / \sum_{\ell=1}^{K} y_{\ell,i}$. Then, it determines the stationary distribution $(v_1, \ldots, v_K)$ for this matrix (i.e., the eigenvector corresponding to the largest eigenvalue 1). Finally, the options are sorted according to these probabilities: $o_i \prec^{\mathrm{RW}} o_j$ iff $v_i < v_j$.

The RW ranking is directly motivated by the PageRank algorithm (Brin & Page, 1998), which has been well studied in social choice theory (Altman & Tennenholtz, 2008; Brandt & Fischer, 2007) and rank aggregation (Negahban et al., 2012), and which is

Algorithm 1 PBR$(Y_{1,1}, \ldots, Y_{K,K}, \kappa, n_{\max}, \delta)$

1: $B = D = \emptyset$ ▷ Set of selected and discarded options
2: $A = \{(i,j) |\ i \neq j, 1 \leq i,j \leq K\}$
3: ▷ Set of all pairs of options still racing
4: **for** $i,j = 1 \to K$ **do** $n_{i,j} = 0$ ▷ Initialization
5: **while** $(\forall\ i\ \forall j,\ (n_{i,j} \leq n_{\max})) \wedge (|A| > 0)$ **do**
6: **for all** $(i,j) \in A$ **do**
7: $n_{i,j} = n_{i,j} + 1$
8: $y_{i,j}^{n_{i,j}} \sim Y_{i,j}$ ▷ Draw a random sample
9: Update $\bar{\mathbf{Y}} = [\bar{y}_{i,j}]_{K \times K}$ with the new samples
10: according to (2)
11: **for** $i,j = 1 \to K$ **do**
12: ▷ Update confidence bounds, $\mathbf{C}, \mathbf{U}, \mathbf{L}$
13: $c_{i,j} = \sqrt{\frac{1}{2n_{i,j}} \log \frac{2K^2 n_{\max}}{\delta}}$
14: ▷ Hoeffding bound
15: $u_{i,j} = \bar{y}_{i,j} + c_{i,j}\ ,\ \ell_{i,j} = \bar{y}_{i,j} - c_{i,j}$
16: $(A, B) = \mathbf{SSCO}(A, \bar{\mathbf{Y}}, K, \kappa, \mathbf{U}, \mathbf{L})$
17: ▷ Sampling strategy for $\prec^{\mathrm{CO}}$
18: $(A, B, D) = \mathbf{SSSE}(A, \bar{\mathbf{Y}}, K, \kappa, \mathbf{U}, \mathbf{L}, D)$
19: ▷ Sampling strategy for $\prec^{\mathrm{SE}}$
20: $(A, B) = \mathbf{SSRW}(\bar{\mathbf{Y}}, K, \kappa, \mathbf{C})$
21: ▷ Sampling strategy for $\prec^{\mathrm{RW}}$
22: **return** B

widely used in many application fields (Brin & Page, 1998; Kocsor et al., 2008).

3. Preference-based Racing Algorithm

The original racing algorithm for the value-based TKS problem is an iterative sampling method. In each iteration, it either selects a subset of options to be sampled, or it terminates and returns a κ-sized subset of options as a (probable) solution to (1).

In this section, we introduce a general preference-based racing (PBR) algorithm that provides the basic statistics needed to solve the selection problem (3), notably estimates of the $y_{i,j}$ and corresponding confidence intervals. It contains a subroutine that implements sampling strategies for the different ranking models described in Section 2.3.

The pseudocode of PBR is shown in Algorithm 1. The set A contains all pairs of options that still need to be sampled; it is initialized with all $K^2 - K$ pairs of indices. The set B contains the indices of the current top-κ solution. The algorithm samples those $Y_{i,j}$ with $(i,j) \in A$ (lines 6–8). Then, it maintains the $\bar{y}_{i,j}$ given in (2) for each pair of options in lines (9–10). We denote the confidence interval of $\bar{y}_{i,j}$ by $[u_{i,j}, \ell_{i,j}]$. To compute confidence intervals, we apply the Hoeffding bound (Hoeffding, 1963) for a sum of random variables in the usual way (see (Mnih et al., 2008) for example).[1]

After the confidence intervals are calculated, one of the sampling strategies implemented as a subroutine is called. Since each sampling strategy can decide to select or discard pairs of options at any time, the confidence level δ has to be divided by $K^2 n_{\max}$ (line 13); this will be explained in more detail below.

The sampling strategies determine which pairs of options have to be sampled in the subsequent iteration. There are three subroutines (**SSCO**, **SSSE**, **SSRW**) in lines 16–21 of PBR that implement, respectively, the sampling strategies for our three ranking models, namely Copeland's (CO), sum of expectation (SE) and random walk (RW). The concrete implementation of the subroutines is detailed in the next section. We refer to the different versions of our preference-based racing algorithm as PBR$-\{$CO, SE, RW$\}$, depending on which sampling strategy is used.

4. Sampling Strategies

4.1. Copeland's Ranking ($\prec^{\mathrm{CO}}$)

The preference relation specified by the matrix $\mathbf{Y}$ is obviously reciprocal, i.e., $y_{i,j} = 1 - y_{j,i}$ for $i \neq j$. Therefore, when using $\prec^{\mathrm{CO}}$ for ranking, the optimization task (3) can be reformulated as follows:

$$\operatorname*{argmax}_{I \subseteq [K]:\ |I| = \kappa} \sum_{i \in I} \sum_{j \neq i} \mathbb{I}\{y_{i,j} > 1/2\} \tag{5}$$

Procedure 2 implements a sampling strategy that optimizes (5). First, for each o_i, we compute the number z_i of options that are worse with sufficiently high probability—that is, for which $u_{i,j} < 1/2$, $j \neq i$ (line 2). Similarly, for each option o_i, we also compute the number w_i of options o_j that are preferred to it with sufficiently high probability—that is, for which $\ell_{i,j} > 1/2$ (line 3). Note that, for each i, there are always at most $K - z_i$ options that can be better. Therefore, if $|\{j \,|\, K - z_j < w_i\}| > K - \kappa$, then i is a member of the solution set I of (5) with high probability (see line 4). The indices of these options are collected in C. Based on a similar argument, options can also be discarded (line 5); their indices are collected in D.

[1] The empirical Bernstein bound (Audibert et al., 2007) could be applied, too, but its application is only advantageous if the support of the random variables is much bigger than their variances (Mnih et al., 2008). Since the support of $Y_{i,j}$ is $[0,1]$, it will not provide tighter bounds in our applications.

Procedure 2 **SSCO**$(A, \bar{\mathbf{Y}}, K, \kappa, \mathbf{U}, \mathbf{L})$

1: **for** $i = 1 \to K$ **do**
2: $\quad z_i = |\{j | u_{i,j} < 1/2 \wedge i \neq j\}|$
3: $\quad w_i = |\{j | \ell_{i,j} > 1/2 \wedge i \neq j\}|$
4: $C = \left\{i : K - \kappa < \left|\{j | K - z_j < w_i\}\right|\right\}$ ▷ Select
5: $D = \left\{i : \kappa < \left|\{j | K - w_j < z_i\}\right|\right\}$ ▷ Discard
6: **for** $(i, j) \in A$ **do**
7: $\quad$ **if** $(i, j \in C \cup D) \vee (1/2 \notin [\ell_{i,j}, u_{i,j}])$ **then**
8: $\quad\quad A = A \setminus (i, j)$ ▷ Stop updating $\bar{y}_{i,j}$
9: $B =$ the top-κ options for which the corresponding rows of $\bar{\mathbf{Y}}$ with most entries above $1/2$
10: **return** (A, B)

In order to update A (the set of $Y_{i,j}$ still racing), we note that, for those options whose indices are in $C \cup D$, it is already decided with high probability whether or not they belong to I. Therefore, if the indices of two options o_i and o_j both belong to $C \cup D$, then $Y_{i,j}$ does not need to be sampled any more, and thus the index pair (i, j) can be excluded from A. Additionally, if $1/2 \notin [\ell_{i,j}, u_{i,j}]$, then the pairwise relation of o_i and o_j is known with sufficiently high probability, so (i, j) can again be excluded from A. These filter steps are implemented in line 7.

Despite important differences between the value-based and the preference-based racing approach, the expected number of samples taken by the latter can be upper-bounded in much the same way as Even-Dar et al. (2002) did for the former.[2]

Theorem 1. *Let $\mathcal{O} = \{o_1, \ldots, o_K\}$ be a set of options such that $\Delta_{i,j} = y_{i,j} - 1/2 \neq 0$ for all $i, j \in [K]$. The expected number of pairwise comparison taken by* PBR-CO *is bounded by*

$$\sum_{i=1}^{K} \sum_{j \neq i} \left\lceil \frac{1}{2\Delta_{i,j}^2} \log \frac{2K^2 n_{\max}}{\delta} \right\rceil .$$

Moreover, the probability that no optimal solution of (6) is found by PBR-CO *is at most δ if $n_{i,j} \leq n_{\max}$ for all $i, j \in [K]$.*

4.2. Sum of Expectations ($\prec^{\text{SE}}$) Ranking

For the SE ranking model, the problem (3) can be written equivalently as

$$\operatorname*{argmax}_{I \subseteq [K]:\ |I| = \kappa} \sum_{i \in I} \sum_{j \neq i} \mathbb{I}\{y_j < y_i\} \ , \tag{6}$$

[2]Due to space limitations, all proofs are moved to the supplementary material.

Procedure 3 **SSSE**$(A, \bar{\mathbf{Y}}, K, \kappa, \mathbf{U}, \mathbf{L}, D)$

1: $G = \{i : i \text{ appearing in } A\}$ ▷ Active options
2: $\widetilde{B} = \{1, \ldots, K\} \setminus (G \cup D)$ ▷ Already selected
3: **for all** $i \in G$ **do**
4: $\quad \ell_i = \frac{1}{K-1} \sum_{j \in G \setminus \{i\}} \ell_{i,j}$
5: $\quad u_i = \frac{1}{K-1} \sum_{j \in G \setminus \{i\}} u_{i,j}$
6: $\widetilde{K} = |G|, \widetilde{\kappa} = \kappa - |\widetilde{B}|$ ▷ Reduced problem
7: $\widetilde{B} = \widetilde{B} \cup \left\{i : \widetilde{K} - \widetilde{\kappa} < \left|\{j \in G : u_j < \ell_i\}\right|\right\}$
8: $D = D \cup \left\{i : \widetilde{\kappa} < \left|\{j \in G : u_i < \ell_j\}\right|\right\}$
9: **for** $(i, j) \in A$ **do**
10: $\quad$ **if** $(i \in \widetilde{B} \cup D)$ **then**
11: $\quad\quad A = A \setminus (i, j)$ ▷ Stop updating $\bar{y}_{i,j}$
12: **for** $i = 1 \to K$ **do** $\bar{y}_i = \frac{1}{K-1} \sum_{j \neq i} \bar{y}_{i,j}$
13: $B =$ the top-κ options with the highest $\bar{y}_i$ values
14: **return** (A, B, D)

with y_i as in (4). The naive implementation would be to sample each random variable until the confidence intervals of the estimates $\bar{y}_i = \frac{1}{K-1} \sum_{j \neq i} \bar{y}_{i,j}$ are non-overlapping. Note, however, that if the upper confidence bound of $\bar{y}_i$ calculated as $u_i = \frac{1}{K-1} \sum_{j \neq i} u_{i,j}$ is smaller than $K - \kappa$ lower bounds $\ell_{i'} = \frac{1}{K-1} \sum_{j \neq i'} \ell_{i',j}$, then the pairwise comparisons with respect to option o_i do not need to be sampled anymore; instead, o_i can be excluded from the solution set of (6) with high probability. Therefore, o_i can be discarded, and we can continue the run of PBR-SE with parameters $K - 1$ and κ (line 6). We use the set D to keep track of the discarded options. An analogous rule can be devised for the selection of options. The pseudocode of the PBR-SE sampling strategy is shown in Procedure 3.

We can also upper-bound the expected number of samples taken by PBR-SE. In fact, this setup is very close to the value-based one, since a single real value $\bar{y}_i$ is assigned to each option.

Theorem 2. *Let $\mathcal{O} = \{o_1, \ldots, o_K\}$ be a set of options. Assume $o_i \prec^{\text{SE}} o_j$ iff $i < j$ without loss of generality and $y_i \neq y_j$ for all $1 \leq i \neq j \leq K$. Let $b_i = \left\lceil \left(\frac{4}{y_i - y_{K-\kappa+1}} \right)^2 \log \frac{2K^2 n_{\max}}{\delta} \right\rceil$ for $i \in [K - \kappa]$ and $b_j = \left\lceil \left(\frac{4}{y_j - y_{K-\kappa}} \right)^2 \log \frac{2K^2 n_{\max}}{\delta} \right\rceil$ for $j = K - \kappa + 1, \ldots, K$. Then, whenever $n_{\max} \geq b_{K-\kappa} = b_{K-\kappa+1}$,* PBR-SE *terminates after $\sum_{i \neq j} b_i = \sum_{i=1}^{K-\kappa} (K - 1) b_i + \sum_{j=K-\kappa+1}^{K} (K-1) b_j$ pairwise comparisons and outputs the optimal solution with probability at least $(1 - \delta)$.*

4.3. Random Walk ($\prec^{\mathrm{RW}}$) Ranking

We start the description of the RW sampling strategy with computing confidence intervals for the elements of a stochastic matrix $\bar{\mathbf{S}} = [\bar{s}_{i,j}]_{K\times K}$ calculated as $\bar{s}_{i,j} = \frac{\bar{y}_{i,j}}{\sum_\ell \bar{y}_{\ell,i}}$, assuming that we know confidence bounds $c_{i,j}$ for a given confidence level δ for each element of the matrix $\bar{\mathbf{Y}} = [\bar{y}_{i,j}]_{K\times K}$. Aslam & Decatur (1998) provide simple bounds for propagating error via some basic operations (see Lemma 1-2). Using their results, a direct calculation yields that $s_{i,j} \in [\bar{s}_{i,j} - \mathfrak{c}_{i,j}, \bar{s}_{i,j} + \mathfrak{c}_{i,j}]$ where $\mathbf{S} = [s_{i,j}]_{K\times K}$ is the stochastic matrix calculated as $s_{i,j} = \frac{y_{i,j}}{\sum_\ell y_{\ell,i}}$ and

$$\mathfrak{c}_{i,j} = \frac{K}{3} \max_k c_{i,k} \sum_\ell \bar{y}_{\ell,i} \tag{7}$$

with probability at least $1 - K\delta$ (since we assumed that the confidence term is δ and each $y_{i,j}$ in the i^{th} row of matrix $\mathbf{Y}$ must be within the confidence interval of $\bar{y}_{i,j}$ to meet (7)). Note that the components of a particular row of matrix $\mathfrak{C} = [\mathfrak{c}_{i,j}]_{K\times K}$ are equal to each other, therefore $\|\mathfrak{C}\|_1 = \max_i \sum_j |\mathfrak{c}_{i,j}| = \frac{K}{3} \max_{i,k} c_{i,k} \sum_\ell \bar{y}_{\ell,i}$.

As a next step, we use the result of Funderlic & Meyer (1986) on the updating of Markov chains.

Theorem 3 (Funderlic&Meyer, 1986)**.** *Let* $\mathbf{S}$ *and* $\mathbf{S}'$ *be the transition matrices of two irreducible Markov chains whose stationary distributions are* $\mathbf{v} = (v_1, \cdots, v_K)$ *and* $\mathbf{v}' = (v'_1, \cdots, v'_K)$*, respectively. Moreover, define the difference matrix of the transition matrices as* $\mathbf{E} = \mathbf{S} - \mathbf{S}'$*. Then, the following inequality holds:*

$$\|\mathbf{v} - \mathbf{v}'\|_{\max} \leq \|\mathbf{E}\|_1 \|\mathbf{A}^{\#}\|_{\max} \ , \tag{8}$$

where $\mathbf{A}^{\#} = \left[a^{\#}_{i,j}\right]_{K\times K} = \left(I - \mathbf{S} + \mathbf{1}\mathbf{v}^T\right)^{-1} - \mathbf{1}\mathbf{v}^T$.

In the PBR framework (Algorithm 1), we gradually decrease the confidence intervals of the entries of the matrix $\bar{\mathbf{Y}}$, thus getting more precise estimates for $\mathbf{Y}$. Let us denote the stochastic matrices derived from $\bar{\mathbf{Y}}$ and $\mathbf{Y}$ by $\bar{\mathbf{S}}$ and $\mathbf{S}$, respectively, and their principal eigenvectors (that belong to the eigenvalue 1) by $\bar{\mathbf{v}} = (\bar{v}_1, \cdots, \bar{v}_K)$ and $\mathbf{v} = (v_1, \cdots, v_K)$. Moreover, let $\mathfrak{C}$ be the matrix that contains the confidence intervals of $\bar{\mathbf{S}}$ as defined in (7). Applying Theorem 3,[3] we have $\|\mathbf{v} - \bar{\mathbf{v}}\|_{\max} \leq \|\mathbf{S} - \bar{\mathbf{S}}\|_1 \|\bar{\mathbf{A}}^{\#}\|_{\max}$, where $\bar{\mathbf{A}}^{\#} = (I - \bar{\mathbf{S}} + \mathbf{1}\bar{\mathbf{v}}^T)^{-1} - \mathbf{1}\bar{\mathbf{v}}^T$. Moreover, we have $\|\mathbf{S} - \bar{\mathbf{S}}\|_1 \leq \|\mathfrak{C}\|_1$ with probability at least $1 - K^2\delta$, since this inequality requires all $s_{i,j}$ to be within the confidence interval given in (7) and, therefore all $y_{i,j}$ must be within the confidence interval of $\bar{y}_{i,j}$.

Summarizing what we found so far, we have

$$\begin{aligned} \|\mathbf{v} - \bar{\mathbf{v}}\|_{\max} &\leq \|\mathbf{S} - \bar{\mathbf{S}}\|_1 \|\bar{\mathbf{A}}^{\#}\|_{\max} \\ &\leq \|\mathfrak{C}\|_1 \|\bar{\mathbf{A}}^{\#}\|_{\max} \end{aligned} \tag{9}$$

This upper bound suggests the minimization of $\|\mathfrak{C}\|_1$. What remains to be shown, however, is that $\|\bar{\mathbf{A}}^{\#}\|_{\max}$ is bounded. In PBR, we gradually estimate $\mathbf{Y}$, thereby obtaining a series of estimates $\bar{\mathbf{Y}}^{(1)}, \ldots, \bar{\mathbf{Y}}^{(n)}$. Now, it is easy to see that if $\bar{\mathbf{Y}}^{(n)}$ converges componentwise to $\mathbf{Y}$, then $\|\bar{\mathbf{A}}^{(n)\#}\|_{\max} \to \|\mathbf{A}^{\#}\|_{\max}$. Moreover, based on (Seneta, 1992) Eq. (7), $\|\mathbf{A}^{\#}\|_{\max}$ is bounded from above for a stochastic matrix $\mathbf{S}$. In order to have a sample complexity analysis for PBR-RW, we would also need to know the rate of convergence of the series $\|\bar{\mathbf{A}}^{(n)\#}\|_{\max}$, which is a quite difficult question.

The inequality (9) suggests a simple sampling strategy: Since the goal is to decrease $\|\mathfrak{C}\|_1 = \frac{K}{3} \max_{i,j} c_{i,j} \sum_\ell \bar{y}_{\ell,i}$, select the pairs of random variables $(i, j) = \operatorname{argmax}_{i,j} c_{i,j} \sum_\ell \bar{y}_{\ell,i}$ for sampling.

Recall our original optimization task, namely to select a subset of options as follows:

$$\operatorname*{argmax}_{I\subseteq[K]:\ |I|=\kappa} \sum_{i\in I} \sum_{j\neq i} \mathbb{I}\{v_j < v_i\} \tag{10}$$

Let σ be the sorting permutation that puts the elements of $\bar{\mathbf{v}}$ in a descending order. Now, if $|\bar{v}_{\sigma(\kappa)} - \bar{v}_{\sigma(\kappa+1)}| > 2\|\mathfrak{C}\|_1 \|\bar{\mathbf{A}}^{\#}\|_{\max}$ is fulfilled, then we can stop sampling, since $|v_i - \bar{v}_i| \leq \|\mathfrak{C}\|_1 \|\bar{\mathbf{A}}^{\#}\|_{\max}$ for $1 \leq i \leq K$ with probability $1 - K^2\delta$; therefore, the confidence term has to be divided by K^2. The pseudo-code of RW sampling strategy is shown in Procedure 4.

5. Experiments with Soccer Data

In this experiment, we applied our preference-based racing method to sports data. We collected the scores of all soccer matches of the last ten seasons from the German Bundesliga. Our goal was to find those three teams that performed best during that time. We restricted to the 8 teams that participated in each Bundesliga season between 2002 to 2012. Table 1 lists the names of these teams and the number of their overall wins (W), losses (L) and ties (T).

Each pair of teams met 20 times. For teams o_i and o_j, we denote the outcome of these matches

[3] Here, we assume that matrix $\bar{\mathbf{S}}$ defines an irreducible Markov chain, but in practice we revised $\bar{\mathbf{S}}$ as $\bar{\mathbf{S}}' = \alpha\bar{\mathbf{S}} + (1-\alpha)/K\mathbf{1}\mathbf{1}^T$ where $0 < \alpha < 1$. We used $\alpha = 0.98$ (for more details on random perturbation of stochastic matrices, see (Langville & Meyer, 2004)).

Procedure 4 SSRW($\bar{\mathbf{Y}}, K, \kappa, \mathbf{C}$)

Convert $\bar{\mathbf{Y}}$ to be stochastic matrix $\bar{\mathbf{S}}$, and calculate $\mathfrak{C}$ based on Eq. (7)
Calculate the eigenvector $\bar{\mathbf{v}}$ of $\bar{\mathbf{S}}$ which belongs to the largest eigenvalue ($= 1$)
Calculate $\bar{\mathbf{A}}^{\#} = \left(I - \bar{\mathbf{S}} + \mathbf{1}\bar{\mathbf{v}}^T\right)^{-1} - \mathbf{1}\bar{\mathbf{v}}^T$
Take the κth and $\kappa+1$th biggest elements of $\bar{\mathbf{v}}$ that are denoted by a and b
if $|a-b| > 2\|\mathfrak{C}\|_1 \|\bar{\mathbf{A}}^{\#}\|_{\max}$ **then** $A = \emptyset$
else $A = \{\text{argmax}_{i,j}\, c_{i,j} \sum_\ell \bar{y}_{\ell,i}\}$
$B =$ the top-κ options for which the elements of $\bar{\mathbf{v}}$ are largest
return (A, B)

by $y^1_{i,j}, \ldots, y^{20}_{i,j}$, and we take the corresponding frequency distribution as the (ground-truth) probability distribution of $Y_{i,j}$. The rankings of the teams with respect to $\prec^{\text{CO}}$, $\prec^{\text{SE}}$ and $\prec^{\text{RW}}$, computed from the expectations $y_{i,j} = \mathbb{E}[Y_{i,j}]$, are also shown in Table 1. While the team of Munich (Bayern München) dominates the Bundesliga regardless of the ranking model, the follow-up positions may vary depending on which method is chosen.

We run our racing algorithm on the outcomes of all matches by sampling from the distributions of the $Y_{i,j}$ (i.e., we sampled from each set of 20 scores with replacement). PBR was parametrized by $\delta = 0.1, \kappa = 3, n_{\max} = \{100, 500, 1000, 5000, 10000\}$. Figure 1 shows the empirical sample complexity versus accuracy of different runs averaged out over 100 runs. As a baseline, we also run the PBR algorithm with uniform sampling meaning that in each iteration we sampled all pairwise comparisons. The accuracy of a run is 1 if all top-κ teams were found, otherwise 0. As we increase $n_{\max}$, the accuracy converges to $1-\delta$. This experiment confirms that our preference-based racing algorithm can indeed recover the top-κ options with a confidence at least $1-\delta$ provided $n_{\max}$ is large enough. Moreover, by using the sampling strategies introduced in Section 4, PBR can achieve an accuracy similar to the uniform sampling for an empirical sample complexity that is an order of magnitude smaller (if again $n_{\max}$ is large enough).

6. A Special Case

In this section, we consider a setting that is in a sense in-between the value-based and the preference-based one. Like in the former, each option o_i is associated with a random variable X_i; thus, it is possible to evaluate individual options, not only to compare pairs of options. However, the random variables X_i take values in a set Ω that is only partially ordered by a preference relation $\preceq$. Thus, like in the preference-based setting, two options are not necessarily comparable in terms of their sampled values. Obviously, the value-based TKS setup described in Section 2.1 is a special case with $\Omega = \mathbb{R}$ and $\preceq$ the standard $\leq$ relation on the reals.

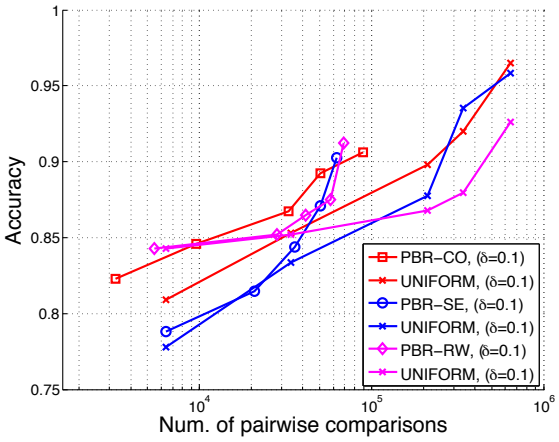


Figure 1. The accuracy of different racing methods versus empirical sample complexity. The algorithms were run with $n_{\max} = \{100, 500, 1000, 5000, 10000\}$. The lowest empirical sample complexity is achieved by setting $n_{\max} = 100$, and the sample complexity grows with $n_{\max}$.

Coming back to our preference-based setting, the pairwise relation $y_{i,j}$ between options can now be written as

$$\mathbf{P}(X_i \prec X_j) + \frac{1}{2}\Big(\mathbf{P}(X_i \sim X_j) + \mathbf{P}(X_i \perp X_j)\Big) \ .$$

Table 1. The 8 Bundesliga teams considered and their scores achieved in the last 10 years. In the last three columns, their ranks are shown according to the different ranking models ($\prec^{\text{CO}}$, $\prec^{\text{SE}}$ and $\prec^{\text{RW}}$). The stars indicate that a team is among the top three.

Team	W	L	T	$\prec^{\text{CO}}$	$\prec^{\text{SE}}$	$\prec^{\text{RW}}$
B. München	77	33	30	*1	*1	*1
B. Dortmund	56	49	35	*3	*2	5
B. Leverkusen	55	49	36	5	4	*2
VfB Stuttgart	55	53	32	*2	5	4
Schalke 04	54	47	39	4	*3	*3
W. Bremen	52	51	37	6	6	6
VfL Wolfsburg	44	66	30	7	7	7
Hannover 96	30	75	35	8	8	8

It can be estimated on the basis of random samples $\mathbf{X}_i = \{x_i^1, \ldots, x_i^{n_i}\}$ and $\mathbf{X}_j = \{x_j^1, \ldots, x_j^{n_j}\}$ drawn from $\mathbf{P}_{X_i}$ and $\mathbf{P}_{X_j}$, respectively, as follows:

$$\bar{y}_{i,j} = \frac{1}{n_i n_j} \sum_{\ell=1}^{n_i} \sum_{\ell'=1}^{n_j} \Big[\mathbb{I}\{x_i^\ell \prec x_j^{\ell'}\} + \frac{1}{2}\big[\mathbb{I}\{x_i^\ell \sim x_j^{\ell'}\} + \mathbb{I}\{x_i^\ell \perp x_j^{\ell'}\}\big]\Big] \tag{11}$$

This estimate is known as *Mann-Whitney U-statistic* (also known as the *Wilcoxon 2-sample statistic*) and belongs to the family of two-sample U-statistics. Apart from $\bar{y}_{i,j}$ being an unbiased estimator of $y_{i,j}$, (11) exhibits concentration properties resembling those of the sum of independent random variables.

Theorem 4 ((Hoeffding, 1963), §5b)**.** [4] *For any $\epsilon > 0$, using the notations introduced above,*

$$\mathbf{P}\left(|y_{i,j} - \bar{y}_{i,j}| \geq \epsilon\right) \leq 2\exp(-2\min(n_i, n_j)\epsilon^2) \ .$$

Based on this concentration result, one can obtain a confidence interval for $\bar{y}_{i,j}$ as follows: for any $0 < \delta < 1$, the interval $[\bar{y}_{i,j} - c_{i,j}, \bar{y}_{i,j} + c_{i,j}]$ contains $y_{i,j}$ with probability at least $1 - \delta$ where $c_{i,j} = \sqrt{\frac{1}{2\min(n_i,n_j)} \ln \frac{2}{\delta}}$.

We can readily adapt the PBR framework to this special setup: In each iteration of PBR, those random variables have to be sampled whose indices appear in A, i.e., those X_i with $(i,j) \in A$ or $(j,i) \in A$. Then, by comparing the random samples with respect to $\preceq$, one can calculate $\bar{y}_{i,j}$ according to (11). Finally, the confidence intervals for the $\bar{y}_{i,j}$ can be obtained based on Theorem 4 (for pseudo-code see Appendix B.1).

6.1. Results on Synthetic Data

Recall that the setup described above is more general than the original value-based one and, therefore, that the PBR framework is more widely applicable than the value-based Hoeffding race (HR).[5] Nevertheless, it is interesting to compare their empirical sample complexity in the standard numerical setting, where both algorithms can be used.

We considered three test scenarios. In the first, each random variable X_i follows a normal distribution $\mathcal{N}((k/2)m_i, c_i)$, where $m_i \sim U[0,1]$, $c_i \sim U[0,1]$, $k \in \mathbb{N}^+$; in the second, each X_i obeys a uniform distribution $U[0, d_i]$, where $d_i \sim U[0, 10k]$ and $k \in \mathbb{N}^+$; in the third, each X_i obeys a Bernoulli distribution $Bern(1/2) + d_i$, where $d_i \sim U[0, k/5]$ and $k \in \mathbb{N}^+$. In every scenario, the goal is to rank the distributions by their means. Note that the complexity of the TKS problem is controlled by the parameter k, with a higher k indicating a less complex task; we varied k between 1 and 10. Besides, we used the parameters $K = 10$, $\kappa = 5$, $n_{\max} = 300$, $\delta = 0.05$.

Strictly speaking, HR is not applicable in the first scenario, since the support of a normal distribution is not bounded; we used $R = 8$ as an upper bound, thus conceding to HR a small probability for a mistake[6]. For Bernoulli and uniform distributions, the bounds of the supports can be readily determined.

Figure 2 shows the number of random samples drawn by the racing algorithms versus precision (percentage of true top-κ variables among the predicted top-κ). PBR-CO, PBR-SE and PBR-RW achieve a significantly lower sample complexity than HR, whereas its accuracy is on a par or better in most cases in the first two test scenarios. While this may appear surprising at first sight, it can be explained by the fact that the Wilcoxon 2-sample statistic is *efficient* (Serfling, 1980).

In the Bernoulli case, one may wonder why the sample complexity of PBR-CO hardly changes with k (see the red point cloud in Figure 2(c)). This can be explained by the fact that the two sample U-statistic $\bar{Y}$ in (11) does not depend on the magnitude of the drift d_i (as long as it is smaller than 1).

7. Related Work

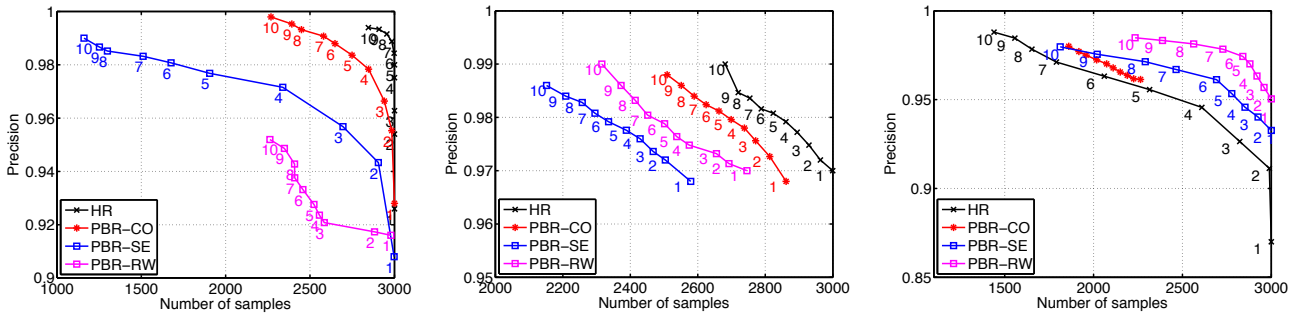
The racing setup and the Hoeffding race algorithm were first considered by Maron & Moore (1994; 1997) in the context of model selection. Mnih et al. (2008) improved the HR algorithm by using the empirical Bernstein bound instead of the Hoeffding bound. In this way, the variance information of the mean estimates could be incorporated in the calculation of confidence intervals.

In the context of multi-armed bandits, Even-Dar et al. (2002) introduced a slightly different setup, where an ϵ-optimal random variable has to be chosen with probability at least $1-\delta$; here, ϵ-optimality of X_i means that $\mu_i + \epsilon \geq \max_{j\in[K]} \mu_j$. Those algorithms solving this problem are called (ϵ, δ)-PAC

[4] Although $\bar{y}_{i,j}$ is a sum of $n_i n_j$ random values here, these values are combinations of only $n_i + n_j$ independent values. This is why the convergence rate is not better than the usual one for a sum of n independent variables.

[5] For a detailed description and implementation of this algorithm, see (Heidrich-Meisner & Igel, 2009).

[6] The probability that all samples remain inside the range is larger than 0.99 for $K = 10$ and $n_{\max} = 300$.



(a) I. Normal distributions (b) II. Uniform distributions (c) III. Bernoulli distributions

Figure 2. The accuracy is plotted against the empirical sample complexities for the Hoeffding race algorithm (HR) and PBR, with the complexity parameter k shown below the markers. Each result is the average of 1000 repetitions.

bandit algorithms. The authors propose such an algorithm and prove an upper bound on the expected sample complexity. In this paper, we borrowed their technique and used it in the complexity analysis of PBR-CO and PBR-SE.

Recently, Kalyanakrishnan et al. (2012) introduced a PAC-bandit algorithm for TKS which is based on the widely-known UCB index-based multi-armed bandit method (Auer et al., 2002). In their formalization, an algorithm is an (ϵ, m, δ)-PAC bandit algorithm that selects the m best random variables under the PAC-bandit conditions. According to their definition, a racing algorithm is a $(0, \kappa, \delta)$-PAC algorithm. They could prove a high probability bound for the worst case sample complexity instead of the expected sample complexity. It is an interesting question whether their slack variable technique can be applied in our setup.

Yue et al. (2012) introduce a multi-armed bandit setup where feedback is provided in the form of noisy comparisons between options, just like in our approach. In their setup, however, they are aiming at a small cumulative regret, where the reward of a pairwise comparison of o_i and o_j is $\max\{\Delta_{i*,i}, \Delta_{i*,j}\}$ whereas ours is a pure exploration approach. To ensure the existence of the best option o_{i*}, strong assumptions are made on the distributions of the comparisons, such as strong stochastic transitivity and stochastic triangle inequality.

In "noisy sorting" (Braverman & Mossel, 2008), noisy pairwise preferences are sampled like in our case, but it is assumed that there is a total order over the objects. That is why the algorithms proposed for this setup require in general less pairwise comparisons in expectation ($O(K \log K)$) than ours.

8. Conclusion and Future Work

We introduced a generalization of the problem of top-k selection under uncertainty, which is based on *comparing* pairs of options in a *qualitative* instead of *evaluating* single options in a *quantitative* way. To tackle this problem, we proposed a general framework in the form of a preference-based racing algorithm along with three concrete instantiations, using different methods for ranking options based on pairwise comparisons. Our algorithms were analyzed formally, and their effectiveness was shown in experimental studies on real and synthetic data.

For future work, there are still a number of theoretical questions to be addressed, as well as interesting variants of our setting. For example, inspired by (Kalyanakrishnan et al., 2012), we plan to consider a variant that seeks to find a ranking that is close to the reference ranking (such as $\prec^{\mathrm{CO}}$) in terms of a given rank distance, thereby distinguishing between correct and incorrect solutions in a more gradual manner than the (binary) top-k criterion.

Moreover, there are several interesting applications of our preference-based TKS setup. Concretely, we are currently working on an application in preference-based reinforcement learning, namely a preference-based variant of evolutionary direct policy search as proposed by Heidrich-Meisner & Igel (2009).

Acknowledgments

This work was supported by the German Research Foundation (DFG) as part of the Priority Programme 1527, and by the ANR-10-BLAN-0215 grant of the French National Research Agency.

References

Akrour, R., Schoenauer, M., and Sebag, M. Preference-based policy learning. In *Proceedings ECMLPKDD 2011, European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, pp. 12–27, 2011.

Altman, A. and Tennenholtz, M. Axiomatic foundations for ranking systems. *Journal of Artificial Intelligence Research*, 31(1):473–495, 2008.

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, 1998.

Audibert, J.Y., Munos, R., and Szepesvári, C. Tuning bandit algorithms in stochastic environments. In *Proceedings of the Algorithmic Learning Theory*, pp. 150–165, 2007.

Auer, P., Cesa-Bianchi, N., and Fischer, P. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47:235–256, 2002.

Brandt, F. and Fischer, F. Pagerank as a weak tournament solution. In *Proceedings of the 3rd international conference on Internet and Network Economics*, pp. 300–305, 2007.

Braverman, Mark and Mossel, Elchanan. Noisy sorting without resampling. In *Proceedings of the nineteenth annual ACM-SIAM Symposium on Discrete algorithms*, pp. 268–276, 2008.

Brin, S. and Page, L. The anatomy of a large-scale hypertextual web search engine. *Computer Networks*, 30 (1-7):107–117, 1998.

Cheng, W., Fürnkranz, J., Hüllermeier, E., and Park, S.H. Preference-based policy iteration: Leveraging preference learning for reinforcement learning. In *Proceedings ECMLPKDD 2011, European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, pp. 414–429, 2011.

Even-Dar, E., Mannor, S., and Mansour, Y. PAC bounds for multi-armed bandit and markov decision processes. In *Proceedings of the 15th Annual Conference on Computational Learning Theory*, pp. 255–270, 2002.

Funderlic, R.E. and Meyer, C.D. Sensitivity of the stationary distribution vector for an ergodic markov chain. *Linear Algebra and its Applications*, 76:1–17, 1986.

Fürnkranz, J. and Hüllermeier, E. (eds.). *Preference Learning*. Springer-Verlag, 2011.

Heidrich-Meisner, V. and Igel, C. Hoeffding and Bernstein races for selecting policies in evolutionary direct policy search. In *Proceedings of the 26th International Conference on Machine Learning*, pp. 401–408, 2009.

Hoeffding, W. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.

Kalyanakrishnan, S., Tewari, A., Auer, P., and Stone, P. Pac subset selection in stochastic multi-armed bandits. In *Proceedings of the Twenty-ninth International Conference on Machine Learning (ICML 2012)*, pp. 655–662, 2012.

Kocsor, A., Busa-Fekete, R., and Pongor, S. Protein classification based on propagation on unrooted binary trees. *Protein and Peptide Letters*, 15(5):428–34, 2008.

Langville, A. N and Meyer, C. D. Deeper inside pagerank. *Internet Mathematics*, 1(3):335–380, 2004.

Maron, O. and Moore, A.W. Hoeffding races: accelerating model selection search for classification and function approximation. In *Advances in Neural Information Processing Systems*, pp. 59–66, 1994.

Maron, O. and Moore, A.W. The racing algorithm: Model selection for lazy learners. *Artificial Intelligence Review*, 5(1):193–225, 1997.

Mnih, V., Szepesvári, C., and Audibert, J.Y. Empirical Bernstein stopping. In *Proceedings of the 25th international conference on Machine learning*, pp. 672–679, 2008.

Moulin, H. *Axioms of cooperative decision making*. Cambridge University Press, 1988.

Negahban, S., Oh, S., and Shah, D. Iterative ranking from pairwise comparisons. In *Advances in Neural Information Processing Systems*, pp. 2483–2491, 2012.

Seneta, E. Sensitivity of finite markov chains under perturbation. *Statistics & probability letters*, 17(2):163–168, 1992.

Serfling, R.J. *Approximation theorems of mathematical statistics*, volume 34. Wiley Online Library, 1980.

Yue, Y., Broder, J., Kleinberg, R., and Joachims, T. The k-armed dueling bandits problem. *Journal of Computer and System Sciences*, 78(5):1538–1556, 2012.