PAC Rank Elicitation through Adaptive Sampling of Stochastic Pairwise Preferences

Róbert Busa-Fekete, Balázs Szörényi, Eyke Hüllermeier

PAC Rank Elicitation through Adaptive Sampling of Stochastic Pairwise Preferences

Róbert Busa-Fekete *
Balázs Szörényi†
Eyke Hüllermeier ‡

November 20, 2015

Abstract

We introduce the problem of PAC rank elicitation, which consists of sorting a given set of options based on adaptive sampling of stochastic pairwise preferences. More specifically, we assume the existence of a *ranking procedure*, such as Copeland's method, that determines an underlying *target order* of the options. The goal is to predict a ranking that is sufficiently close to this target order with high probability, where closeness is measured in terms of a suitable distance measure. We instantiate this setting with combinations of two different distance measures and ranking procedures. For these instantiations, we devise efficient strategies for sampling pairwise preferences and analyze the corresponding sample complexity. We also present first experiments to illustrate the practical performance of our methods.

Introduction

Exploiting revealed (pairwise) preferences to learn a ranking (total order) over a set of options is a challenging problem with many practical applications. For example, think of crowd-sourcing services like the Amazon Mechanical Turk, where simple questions such as pairwise comparisons between decision alternatives are asked to a group of annotators. The task is to approximate an underlying target ranking on the basis of these pairwise comparisons, which are possibly noisy and partially inconsistent [Chen et al., 2013]. Another application worth mentioning is the ranking of XBox gamers based on their pairwise online duels; the ranking system of XBox is called TrueSkill™[Guo et al., 2012].

In this paper, we focus on a problem that we call *PAC rank elicitation*. In the setting of this problem, we consider a finite set of options $\mathcal{O} = \{o_1, \ldots, o_K\}$, on which a weighted relation $\mathbf{Y} = (y_{i,j})_{1 \leq i,j \leq K}$ is defined. As will be explained in more detail later on, this relation specifies the probability of observing preferences $o_j \prec o_i$, suggesting that, in a single comparison of two options o_i and o_j, the former was liked more than the latter. Furthermore, we assume the existence of a *ranking procedure* $\mathcal{R}$ that determines an underlying *target (strict) order* $\prec^*$ of the options $\mathcal{O}$ based on $\mathbf{Y}$.

In rank elicitation, we assume that $\mathcal{R}$ is given whereas $\mathbf{Y}$ is not known. Instead, information about $\mathbf{Y}$ can only be obtained through (adaptive) sampling of pairwise preferences. The goal, then, is to quickly gather enough information so as to enable the prediction of a ranking that is sufficiently close to the target order $\prec^*$ with high probability. We shall describe this rank elicitation setting more formally and, moreover, instantiate it with combinations of two different distance measures and two ranking procedures for determining the target order. For these instantiations, we devise efficient sampling strategies and analyze them in terms of expected sample complexity. Finally, we also present an experimental study, prior to concluding the paper.

*MTA-SZTE Research Group on Artificial Intelligence, Hungary; busarobi@inf.u-szeged.hu
†INRIA Lille, SequeL project, France; MTA-SZTE Research Group on Artificial Intelligence, Hungary; szorenyi@inf.u-szeged.hu
‡University of Paderborn, Germany; eyke@upb.de

Related work

Ranking based on sampling pairwise relations has a long history in the literature [Braverman and Mossel, 2008, Braverman and Mossel, 2009, Eriksson, 2013, Feige et al., 1994]. Existing algorithms for *noisy sorting* typically solve this problem with sample complexity $O(K \log K)$. However, these algorithms make strong assumptions: the target relation is a total order, and the comparisons are representative of that order (if o_i precedes o_j, then $\mathbb{P}(o_i \prec o_j) > 1/2$).

Pure exploration algorithms for the stochastic multi-armed bandit problem sample the arms a certain number of times (not necessarily known in advance), and then output a recommendation, such as the best arm or the m best arms [Bubeck et al., 2009, Even-Dar et al., 2002, Bubeck et al., 2013, Gabillon et al., 2011, Cappé et al., 2012]. While our algorithm can be viewed as a pure exploration strategy, too, we do not assume that *numerical* feedback can be generated for *individual* options; instead, our feedback is *qualitative* and refers to *pairs* of options.

Seen from this point of view, our approach is closer to the dueling bandits problem introduced by [Yue et al., 2012], where feedback is provided in the form of noisy comparisons between option. However, apart from making strong structural assumptions (namely strong stochastic transitivity and stochastic triangle inequality), their problem of cumulative regret minimization is of an exploration-exploitation nature.

The kind of feedback assumed in our rank elicitation setup is in fact the one considered by [Busa-Fekete et al., 2013] and [Urvoy et al., 2013], who both solve the top-k subset selection (or EXPLORE-k) problem: Find the k best options with respect to a target ranking based on sampling pairwise preferences. Interestingly, rank elicitation can be seen as solving the top-k problem for all $k \in [K]$ simultaneously, and indeed, our approach builds on this connection. Our starting point is the recent paper [Kalyanakrishnan et al., 2012], which introduces a PAC-bandit algorithm for the top-k problem in the stochastic multi-armed bandit environment (i.e., based on numerical feedback, not pairwise preferences).

In the formulation of [Kalyanakrishnan et al., 2012], an algorithm is an (ϵ, m, δ)-PAC bandit algorithm if it selects the m best options (those with the highest expected value) under the PAC-bandit conditions [Even-Dar et al., 2002]. The concrete algorithm they propose is based on the widely-known UCB index-based multi-armed bandit method [Auer et al., 2002]. Our theoretical analysis partly relies on their results, using an expected sample complexity and a high probability bound for the worst case sample complexity. In fact, although our setup is based on preferences, we aim at a similar kind of sample complexity result.

Problem setting and terminology

PAC rank elicitation setup

Our point of departure are pairwise preferences over the set of options $\mathcal{O} = \{o_1, \ldots, o_K\}$. More specifically, we allow three possible outcomes of a single pairwise comparison between o_i and o_j, namely (strict) preference for o_i, (strict) preference for o_j, and incomparability/indifference. These outcomes are denoted by $o_i \succ o_j$, $o_i \prec o_j$, and $o_i \perp o_j$, respectively. In our setting, 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 $o_j \prec o_i$, 0 if $o_i \prec o_j$, and $1/2$ otherwise. Thus, the case $o_i \perp o_j$ is handled by giving half a point to both options. Essentially, this means that these outcomes are treated in a neutral way by the ranking procedures.

The expected values $y_{i,j} = \mathbb{E}[Y_{i,j}]$ can be summarized in the relation $\mathbf{Y} = [y_{i,j}] \in [0,1]^{K \times K}$. A natural idea to define a pairwise preference relation $\prec$ on $\mathcal{O}$ is to "binarize" $\mathbf{Y}$: $o_i \prec o_j$ if and only if $y_{i,j} < y_{j,i}$. This relation, however, may contain preferential cycles and, therefore, may not define a proper order relation. In decision making, this problem is commonly avoided by using a ranking procedure $\mathcal{R}$ (concrete choices of $\mathcal{R}$ will be discussed in the next section) that turns $\mathbf{Y}$ into a strict order relation $\prec^{\mathcal{R}}$ of the options $\mathcal{O}$. Formally, a ranking procedure $\mathcal{R}$ is a map $[0,1]^{K \times K} \to \mathcal{S}_{\mathcal{O}}$, where $\mathcal{S}_{\mathcal{O}}$ denotes the set of strict orders on $\mathcal{O}$. We denote the strict order produced by the ranking procedure $\mathcal{R}$ on the basis of $\mathbf{Y}$ by $\prec^{\mathcal{R}}_{\mathbf{Y}}$, or simply by $\prec^{\mathcal{R}}$ if $\mathbf{Y}$ is clear from the context.

The task in PAC rank elicitation is to approximate $\prec^{\mathcal{R}}$ without knowing the $y_{i,j}$. Instead, relevant information can only be obtained *through sampling pairwise comparisons from the underlying distribution*. Thus, we assume that options can be compared in a pairwise manner, and that a single sample essentially informs about a pairwise

preference between two options o_i and o_j. The goal is to devise a *sampling strategy* that keeps the size of the sample (the sample complexity) as small as possible while producing an estimation $\prec$ that is "good" in a PAC sense: $\prec$ is supposed to be sufficiently "close" to $\prec^{\mathcal{R}}$ with high probability. Actually, our algorithms even produce a total order as a prediction, i.e., $\prec$ is a ranking that can be represented by a permutation τ of order K, where τ_i denotes the rank of option o_i in the order (with smaller ranks indicating higher preference, i.e., $o_i \prec o_j$ if $\tau_i > \tau_j$).

To formalize the notion of "closeness", we make use of appropriate distance measures that compare a (predicted) permutation τ with a (target) strict order $\prec$. In particular, we adopt the following two measures: The *number of discordant pairs* (NDP), which is closely connected to Kendall's rank correlation [Kendall, 1955], and can be expressed in terms of the indicator function $\mathbb{I}\{\cdot\}$ as follows:

$$d_{\mathcal{K}}(\tau, \prec) = \sum_{i=1}^{K} \sum_{j \neq i} \mathbb{I}\{\tau_j > \tau_i\} \mathbb{I}\{o_i \prec o_j\}.$$

The *maximum rank difference* (MRD) is defined as the maximum difference between the rank of an object o_i according to τ and $\prec$, respectively. More specifically, since $\prec$ is a partial but not necessarily total order, we compare τ to the set $\mathcal{L}^{\prec}$ of its linear extensions[2]:

$$d_{\mathcal{M}}(\tau, \prec) = \min_{\tau' \in \mathcal{L}^{\prec}} \max_{1 \leq i \leq K} |\tau_i - \tau'_i|.$$

Our setup allows for small approximation errors, formalized by a tolerance parameter $\rho \in \mathbb{N}^+$.[3] We call an algorithm $\mathscr{A}$ a (ρ, δ)-*PAC rank elicitation algorithm* with respect to a ranking procedure $\mathcal{R}$ and rank distance d, if it returns a ranking τ for which $d(\tau, \prec^{\mathcal{R}}) < \rho$ with probability at least $1 - \delta$.

Ranking procedures

In the following, we introduce two instantiations of the ranking procedure $\mathcal{R}$, namely Copeland's ranking (binary voting) and the sum of expectations (weighted voting). To define the former, let $d_i = \#\{k \in [K] \,|\, 1/2 < y_{i,k}\}$ denote the number of options that are "beaten" by o_i. Copeland's ranking (CO) is then defined as follows [Moulin, 1988]: $o_i \prec^{\mathrm{CO}} o_j$ if and only if $d_i < d_j$. The sum of expectations (SE) ranking is a "soft" version of CO: $o_i \prec^{\mathrm{SE}} o_j$ if and only if

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

Since $\mathcal{R}$ is mapping the continuous space $[0,1]^{K \times K}$ to the discrete space $\mathcal{S}_{\mathcal{O}}$, ranking is a "non-smooth" operation. In the case of the Copeland order $\prec^{\mathrm{CO}}$, for example, a minimal change of a value $y_{i,j} \approx \frac{1}{2}$ may strongly influence $\prec^{\mathrm{CO}}$. Consequently, the number of samples needed to assure (with high probability) a certain approximation quality may become arbitrarily large. A similar problem arises for $\prec^{\mathrm{SE}}$ as a target order if some of the individual scores y_i are very close or equal to each other.

As a practical (yet meaningful) solution to this problem, we propose to make the relations $\prec^{\mathrm{CO}}$ and $\prec^{\mathrm{SE}}$ a bit more "partial" by imposing stronger requirements on the strict order. To this end, let $d_i^* = \#\{k \,|\, 1/2 + \epsilon < y_{i,k}, i \neq k\}$ denote the number of options that are beaten by o_i with a margin $\epsilon > 0$, and let $s_i^* = \#\{k : |1/2 - y_{i,k}| \leq \epsilon,\, i \neq k\}$. Then, we define the ϵ-insensitive Copeland relation as follows: $o_i \prec^{\mathrm{CO}_\epsilon} o_j$ if and only if $d_i^* + s_i^* < d_j^*$. Likewise, in the case of $\prec^{\mathrm{SE}}$, we neglect small differences of the y_i and define the ϵ-insensitive sum of expectations relation as follows: $o_i \prec^{\mathrm{SE}_\epsilon} o_j$ if and only if $y_i + \epsilon < y_j$.

These ϵ-insensitive extensions are interval (and hence strict) orders, that is, they are obtained by characterizing each option o_i by the interval $[d_i^*, d_i^* + s_i^*]$ and sorting intervals according to $[a, b] \prec [a', b']$ iff $b < a'$. It is readily shown that $\prec^{\mathrm{CO}_\epsilon} \subseteq \prec^{\mathrm{CO}_{\epsilon'}} \subseteq \prec^{\mathrm{CO}}$ for $\epsilon > \epsilon'$, with equality $\prec^{\mathrm{CO}_0} \equiv \prec^{\mathrm{CO}}$ if $y_{i,j} \neq 1/2$ for all $i \neq j \in [K]$ (and similarly for SE). Subsequently, ϵ will be taken as a parameter that controls the strictness of the order relations, and thereby the difficulty of the (ρ, δ)-rank elicitation task.

[2] $\tau \in \mathcal{L}^{\prec}$ iff $\forall\, i, j \in [K] :\ (o_i \prec o_j) \Rightarrow (\tau_j < \tau_i)$

[3] Note that our distance measures assume values in $\mathbb{N}_0$ and are not normalized. Although a normalization to $[0, 1]$ could easily be done, it would unnecessarily complicate the description of the algorithms and their analysis.

A general rank elicitation algorithm

In this section, we introduce a general rank elicitation framework (RANKEL) that provides the basic statistics needed to solve the PAC rank elicitation problem, notably estimates of the pairwise probabilities $y_{i,j}$ and the number of samples drawn from $Y_{i,j}$ so far. It contains a subroutine that implements sampling strategies for the different distance measures and ϵ-insensitive ranking models.

Our general framework 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 (line 3). In each iteration, the algorithm samples those $Y_{i,j}$ with $(i,j) \in A$ (lines 7) and maintains the estimates $\bar{\mathbf{Y}} = [\bar{y}_{i,j}]_{K\times K}$, where $\bar{y}_{i,j} = \frac{1}{n_{i,j}} \sum_{\ell=1}^{n_{i,j}} y_{i,j}^{\ell}$ is the mean of the $n_{i,j}$ samples drawn from $Y_{i,j}$ so far. These numbers are maintained by the algorithm, too, and are stored in the matrix $\mathbf{N} = [n_{i,j}]_{K\times K}$. The sampling strategy subroutine returns the indices of option pairs to be sampled. If A is empty, then RANKEL stops and returns a ranking τ over $\mathcal{O}$, which is calculated based on $\bar{\mathbf{Y}}$ (line 15). The sampling strategy depends on the ranking procedure and the distance measure used. We shall describe its concrete implementations in subsequent sections.

Algorithm 1 RANKEL $(Y_{1,1}, \ldots, Y_{K,K}, \rho, \delta, \epsilon)$

1: **for** $i, j = 1 \to K$ **do** ▷ Initialization
2: $\bar{y}_{i,j} = 0$, $n_{i,j} = 0$
3: $A = \{(i,j) | i \neq j, 1 \leq i, j \leq K\}$
4: $t = 0$
5: **repeat**
6: **for** $(i,j) \in A$ **do**
7: $y \sim Y_{i,j}$ ▷ Draw a random sample
8: $n_{i,j} = n_{i,j} + 1$
9: ▷ Keep track the number of samples drawn for each $Y_{i,j}$
10: Update $\bar{y}_{i,j}$ with y
11: ▷ $\bar{\mathbf{Y}} = [\bar{y}_{i,j}]_{K\times K} \approx \mathbf{Y} = [y_{i,j}]_{K\times K}$
12: $t = t + 1$
13: $A = \text{SAMPLINGSTRATEGY}(\bar{\mathbf{Y}}, \mathbf{N}, \delta, \epsilon, t, \rho)$
14: **until** $0 < |A|$
15: $\tau = \text{GETESTIMATEDRANKING}(\bar{\mathbf{Y}}, \mathbf{N}, \delta, \epsilon, t)$ ▷ Calculate a ranking based on $\bar{\mathbf{Y}}$ by using $\mathcal{R}$
16: **return** τ

We refer to our algorithm as $\text{RANKEL}_d^{\mathcal{R}}$, depending on which ranking procedure $\mathcal{R}$ (ϵ-insensitive Copeland (CO_ϵ) or sum of expectations (SE_ϵ)) and which distance measure d ($d_{\mathcal{K}}$ or $d_{\mathcal{M}}$) are used. For example, $\text{RANKEL}_{d_{\mathcal{K}}}^{\text{CO}_\epsilon}$ denotes the instance of our algorithm that seeks to find a ranking close to the ϵ-insensitive Copeland order in terms of $d_{\mathcal{K}}$.

Sampling strategies

The case of ϵ-insensitive Copeland

In the following, we denote the estimate of $y_{i,j} = \mathbb{E}(Y_{i,j})$ at time step t by $\bar{y}_{i,j}^t$, and the number of samples taken from $Y_{i,j}$ up to that time step by $n_{i,j}^t$ (omitting the time index if not needed). We start the description of our sampling strategy by determining reasonable confidence intervals for the $\bar{y}_{i,j}^t$ values.[4]

Lemma 1. *For any sampling strategy in line 13 of Algorithm 1,* $\sum_{i=1}^{K} \sum_{j\neq i} \sum_{t=1}^{\infty} \mathbb{P}(A_{i,j}^t) \leq \delta$, *where* $A_{i,j}^t = \left\{ y_{i,j} \notin \left[\bar{y}_{i,j}^t - c(n_{i,j}^t, t, \delta), \bar{y}_{i,j}^t + c(n_{i,j}^t, t, \delta) \right] \right\}$ *with* $c(n,t,\delta) = \sqrt{\frac{1}{2n} \ln\left(\frac{5K^2t^4}{4\delta}\right)}$.

[4] Due to space limitations, all proofs are omitted.

From now on, we will concisely write $c^t_{i,j}$ for $c(n^t_{i,j}, t, \delta)$ and $C^t_{i,j}$ for the confidence interval $\left[\bar{y}^t_{i,j} - c^t_{i,j}, \bar{y}^t_{i,j} + c^t_{i,j}\right]$. Now, one can calculate a lower bound of d^*_i based on $\bar{\mathbf{Y}}^t$ and $\mathbf{N}^t$. First, let us define $d^t_i = \#D^t_i$, where

$$D^t_i = \left\{ j \,|\, 1/2 - \epsilon < \bar{y}^t_{i,j} - c^t_{i,j}, j \neq i \right\} \ .$$

Put in words, d^t_i denotes the number of options that are already known to be beaten by o_i. Similarly, we define the number of "undecided" pairwise preferences for an option o_i as $u^t_i = \#U^t_i$, where

$$U^t_i = \left\{ j \,|\, [1/2 - \epsilon, 1/2 + \epsilon] \subseteq C^t_{i,j}, j \neq i \right\} \ .$$

Based on d^t_i and u^t_i, we define a ranking τ^t over $\mathcal{O}$ by sorting the options o_i in increasing order according to d^t_i, and in case of a tie ($d^t_i = d^t_j$) according to the sum $d^t_i + u^t_i$. The following corollary upper-bounds the NDP and MRP distances between τ^t and the underlying order $\prec^{\mathrm{CO}_\epsilon}$ based on only empirical estimates.

Corollary 2. *Using the notation introduced above, let*

$$\mathbb{I}^t_{i,j} = \mathbb{I}\left\{ (d^t_i < d^t_j + u^t_j) \wedge (d^t_j < d^t_i + u^t_i) \right\}$$

for all $1 \leq i \neq j \leq K$*. Then for any time step t, and for any sampling strategy,* $d_{\mathcal{K}}(\tau^t, \prec^{\mathrm{CO}_\epsilon}) \leq \frac{1}{2} \sum_{i=1}^{K} \sum_{j \neq i} \mathbb{I}^t_{i,j}$ *holds with probability at least* $1 - \delta$*, and* $d_{\mathcal{M}}(\tau^t, \prec^{\mathrm{CO}_\epsilon}) \leq \max_{1 \leq i \leq K} \sum_{j \neq i} \mathbb{I}^t_{i,j}$ *holds again with probability at least* $1 - \delta$*.*

Corollary 2 implies that sampling can be stopped as soon as

$$\sum_{i=1}^{K} \sum_{j \neq i} \mathbb{I}^t_{i,j} < \rho \quad \text{and} \quad \max_{1 \leq i \leq K} \sum_{j \neq i} \mathbb{I}^t_{i,j} < \rho \tag{2}$$

in the case of NDP and MRD, respectively. Moreover, it suggests a simple greedy strategy for sampling, namely to sample those pairwise preferences that promise a maximal decrease of the respective upper bound in (2). For NDP, this comes down to sampling all undecided pairs of objects ($\cup_i U^t_i$), although this strategy can still be improved: If the rank of an object o_i can be determined based on the samples seen so far ($\mathbb{I}^t_{i,j} = 0$ for all $j \in [K]$), then there is no need to sample any more pairwise preference involving o_i. Formally, the set of object pairs to be sampled can thus be written

$$\widetilde{A}^t_{\mathcal{K}} = \left\{ (i,j) \,|\, (j \in U^t_i) \wedge \exists\, j' : \, (\mathbb{I}^t_{i,j'} = 1) \right\}.$$

Further considering the stopping rule in (2), the pairwise preferences to be sampled by $\textsc{RankEl}^{\mathrm{CO}_\epsilon}_{d_{\mathcal{K}}}$ in iteration t is given by

$$A^t_{\mathcal{K}} = \begin{cases} \widetilde{A}^t_{\mathcal{K}} & \text{if } \rho \leq \sum_{i=1}^{K} \sum_{j \neq i} \mathbb{I}^t_{i,j} \\ \emptyset & \text{otherwise} \end{cases} \ . \tag{3}$$

In the case of the MRD distance, the goal is to decrease the upper bound on $d_{\mathcal{M}}(\tau^t, \prec^{\mathrm{CO}})$. Correspondingly, the greedy strategy samples the set of pairs

$$\widetilde{A}^t_{\mathcal{M}} = \left\{ (i,j) \,|\, (j \in U^t_i) \wedge \rho \leq \sum_{j' \neq i} \mathbb{I}^t_{i,j'} \right\} \ .$$

Thus, again considering the stopping rule in (2), we can formally write the set of pairs to be sampled by $\textsc{RankEl}^{\mathrm{CO}_\epsilon}_{d_{\mathcal{M}}}$ in iteration t as follows:

$$A^t_{\mathcal{M}} = \begin{cases} \widetilde{A}^t_{\mathcal{M}} & \text{if } \rho \leq \max_{1 \leq i \leq K} \sum_{j \neq i} \mathbb{I}^t_{i,j} \\ \emptyset & \text{otherwise} \end{cases} \tag{4}$$

As a last step, the RankEl algorithm calls a subroutine to calculate the estimated ranking. According to Corollary 2, τ^t is a suitable choice, because its distance to $\prec^{\mathrm{CO}_\epsilon}$ is smaller than ρ with probability at least $1 - \delta$.

The case of ϵ-insensitive sum of expectations

The SE ranking procedure assigns a real number $y_i = \frac{1}{K-1}\sum_{k\neq i} y_{i,k}$ to every option o_i. Based on the pairwise estimates $\bar{y}^t_{i,1}, \ldots, \bar{y}^t_{i,K}$, an estimate for y_i can simply be obtained as $\bar{y}^t_i = \frac{1}{K-1}\sum_{k\neq i} \bar{y}^t_{i,k}$. Similarly to Lemma 1, one can determine a reasonable confidence interval for the $\bar{y}^t_i$ values.

Lemma 3. *Let $c(n,t,\delta)$ be the function defined in Lemma 1. Then, for any sampling strategy in line 13 of Algorithm 1 that ensures $n^t_{i,1} = n^t_{i,2} = \cdots = n^t_{i,K}$ for any $1 \leq i \leq K$, it holds that $\sum_{i=1}^{K}\sum_{t=1}^{\infty}\mathbb{P}(B^t_i) \leq \delta$, where $B^t_i = \{y_i \notin [\bar{y}^t_i - c(n^t_i,t,\delta), \bar{y}^t_i + c(n^t_i,t,\delta)]\}$ and $n^t_i = \sum_{k\neq i} n^t_{i,k}$.*

From now on, we will concisely write c^t_i for $c(n^t_i,t,\delta)$ and C^t_i for the confidence interval $[\bar{y}^t_i - c^t_i, \bar{y}^t_i + c^t_i]$. Given the above estimates, the most natural way to define a ranking σ^t on $\mathcal{O}$ is to sort the options o_i in increasing order according to their scores $\bar{y}^t_i$ (again breaking ties at random). The following corollary upper-bounds the rank distances between σ^t thus defined and $\prec^{\mathrm{SE}_\epsilon}$ in terms of the overlapping confidence intervals of $\bar{y}^t_1, \ldots, \bar{y}^t_K$.

Corollary 4. *Under the condition of Lemma 3, $d_{\mathcal{K}}(\sigma^t, \prec^{\mathrm{SE}_\epsilon}) \leq \frac{1}{2}\sum_{i=1}^{K}\sum_{j\neq i}\mathbb{O}^t_{i,j}$ holds with probability at least $1-\delta$ for any time step t, where $\mathbb{O}^t_{i,j} = \mathbb{I}\left\{|C^t_i \cap C^t_j| > \epsilon\right\}$ indicates that the confidence intervals of $\bar{y}^t_i$ and $\bar{y}^t_j$ are overlapping by more than ϵ. Moreover, $d_{\mathcal{M}}\left(\sigma^t, \prec^{\mathrm{SE}_\epsilon}\right) \leq \max_{1\leq i\leq K}\sum_{j\neq i}\mathbb{O}^t_{i,j}$ is again valid with probability at least $1-\delta$.*

Based on Corollary 4, one can devise greedy sampling strategies that gradually decrease the upper bound of the distances between the current ranking and $\prec^{\mathrm{SE}_\epsilon}$ with respect to $d_{\mathcal{K}}$ or $d_{\mathcal{M}}$, similar to the one described in the previous section for ϵ-sensitive Copeland procedure.

The ranking eventually returned by RANKEL (Algorithm 1, line 15) is simply the one introduced above, namely the permutation that sorts the options o_i according to their scores $\bar{y}_i$.

Complexity analysis

From Propositions 2 and 4, it is immediate that all instantiations of our RANKEL algorithm ($\mathrm{RANKEL}^{\mathrm{CO}_\epsilon}_{d_{\mathcal{K}}}$, $\mathrm{RANKEL}^{\mathrm{CO}_\epsilon}_{d_{\mathcal{M}}}$, $\mathrm{RANKEL}^{\mathrm{SE}_\epsilon}_{d_{\mathcal{K}}}$, $\mathrm{RANKEL}^{\mathrm{SE}_\epsilon}_{d_{\mathcal{M}}}$) are correct, and hence they are all (ρ,δ)-PAC rank elicitation algorithms. In this section, we analyze $\mathrm{RANKEL}^{\mathrm{CO}_\epsilon}_{d_{\mathcal{M}}}$ and calculate an upper bound for its expected sample complexity. In our preference-based setup, the sample complexity of an algorithm is the expected number of pairwise comparisons drawn for a given instance of the rank elicitation problem.

The technique we shall use for analyzing $\mathrm{RANKEL}^{\mathrm{CO}_\epsilon}_{d_{\mathcal{M}}}$ can be applied for $\mathrm{RANKEL}^{\mathrm{SE}_\epsilon}_{d_{\mathcal{M}}}$, too. It cannot be used, however, to characterize the complexity of the rank elicitation task in the case of the $d_{\mathcal{K}}$ distance (see Lemma 6), whence we leave the analysis of $\mathrm{RANKEL}^{\mathrm{CO}_\epsilon}_{d_{\mathcal{K}}}$ and $\mathrm{RANKEL}^{\mathrm{SE}_\epsilon}_{d_{\mathcal{K}}}$ as an open problem.

Expected sample complexity of $\mathrm{RANKEL}^{\mathrm{CO}_\epsilon}_{d_{\mathcal{M}}}$

Step 1: The following lemma upper-bounds the probability of an estimate $\bar{y}^t_{i,j}$ being significantly bigger than $1/2$ while $y_{i,j} < 1/2$ and vice versa. More specifically, it shows that the error probability decreases with the number of iterations t as fast as $O(1/t^3)$, a fact that will be useful in our sample complexity analysis later on.

Lemma 5. *Let $\mathcal{E}^t_{i,j}$ denote the event that either $\bar{y}^t_{i,j} - c^t_{i,j} > 1/2 - \epsilon$ and $y_{i,j} < 1/2 - \epsilon$ or $\bar{y}^t_{i,j} + c^t_{i,j} < 1/2 + \epsilon$ and $y_{i,j} > 1/2 + \epsilon$. Then $\mathrm{RANKEL}^{\mathrm{CO}_\epsilon}_{d}$ satisfies $\sum_{i=1}^{K}\sum_{j\neq i}\mathbb{P}\left[\mathcal{E}^t_{i,j}\right] < \frac{4\delta}{5t^3}$.*

Step 2: An interesting property of our problem setting, which distinguishes it from related ones such as top-k and best arm identification, is that it does not only incorporate an ϵ-tolerance on the level of pairwise probability estimates ($y_{i,j}$ values), but also relaxes the required accuracy of the solution along another dimension, namely the proximity of the predicted ranking and the target order. More precisely, the algorithm receives a parameter ρ, and has to guarantee with high confidence that the ranking τ it outputs is at most of distance ρ from some ranking in $\mathcal{L}^{\prec^{\mathrm{CO}_\epsilon}_{\mathbf{Y}}}$.

Unfortunately, one cannot directly determine the smallest distance between a given τ and $\mathcal{L}^{\prec^{\mathrm{CO}_\epsilon}_{\mathbf{Y}}}$ without knowing the entries of $\mathbf{Y}$ with high accuracy. Instead, an indirect method has to be used in order to bound the sample complexity. To this end, denote by $(\mathbf{Y})_r$ the set of matrices that are obtained from $\mathbf{Y}$ as follows

$$\begin{aligned}(\mathbf{Y})_r = \{\tilde{\mathbf{Y}} \,|\, &\tilde{y}_{i,j} < 1/2 \text{ if } y_{i,j} < 1/2 - \epsilon \text{ and} \\ &\tilde{y}_{i,j} > 1/2 \text{ if } y_{i,j} > 1/2 + \epsilon \text{ where} \\ &(i,j) \in A' \subset A, |A \setminus A'| = r\}\end{aligned}$$

where $A = \{(i,j) \,|\, i \neq j, 1 \leq i,j \leq K\}$ is the set of all off-diagonal index pairs.

Now, if all but at most r entries in $\bar{\mathbf{Y}}^t$ are known to be either bigger than $1/2+\epsilon$ or smaller than $1/2-\epsilon$ with sufficiently high confidence (i.e., if all but at most r pairs (i,j) satisfy $j \notin U_i^t$), then $\bar{\mathbf{Y}}^t \in (\mathbf{Y})_r$ with high probability. Moreover, note that no algorithm can safely terminate as long as no ranking τ exists that satisfies both that it is consistent with the current information (i.e., $\tau \in \mathcal{L}^{\prec^{\mathrm{CO}_\epsilon}_{\bar{\mathbf{Y}}}}$), and that it is of distance at most ρ from any possible strict order—that is formally

$$\max_{\mathbf{Y}' : \bar{\mathbf{Y}}^t \in (\mathbf{Y}')_r} d_{\mathcal{M}}(\tau, \prec^{\mathrm{CO}_\epsilon}_{\mathbf{Y}'}) \leq \rho \ .$$

Accordingly, one should define the variation of distance $d_{\mathcal{M}}$ around $\mathbf{Y}$ at radius r as

$$v^{\mathrm{CO}_\epsilon}_{d_{\mathcal{M}}}(r, \mathbf{Y}) = \max_{\tilde{\mathbf{Y}} \in (\mathbf{Y})_r} \min_{\tau \in \mathcal{L}^{\prec^{\mathrm{CO}_\epsilon}_{\tilde{\mathbf{Y}}}}} \max_{\mathbf{Y}' : \tilde{\mathbf{Y}} \in (\mathbf{Y}')_r} d_{\mathcal{M}}\left(\tau, \prec^{\mathrm{CO}\epsilon}_{\mathbf{Y}'}\right)$$

The next result shows that the ranking output by $\textsc{RankEl}^{\mathrm{CO}_\epsilon}_{d_{\mathcal{M}}}$ is always within this distance ($v^{\mathrm{CO}_\epsilon}_{d_{\mathcal{M}}}(r, \mathbf{Y})$) and thus, it is indeed a reasonable definition.

Lemma 6. *Assume that $A^t = \cap_{i=1}^K \cap_{j \neq i} A^t_{i,j}$ holds, where $A^t_{i,j}$ denotes the event defined in Lemma 1. Let τ denote some ranking that satisfies $\tau_i > \tau_j$ whenever $(d_i^t < d_j^t)$ or $(d_i^t = d_j^t) \wedge (d_i^t + u_i^t < d_j^t + u_j^t)$ holds for some $t > 0$. Then $d_{\mathcal{M}}(\tau, \prec^{\mathrm{CO}_\epsilon}) \leq \max_i I_i^t$, where $I_i^t = \sum_{j \neq i} \mathbb{I}^t_{i,j} = \#\{j : (d_i^t < d_j^t + u_j^t) \wedge (d_j^t < d_i^t + u_i^t)\}$. Moreover, $\max_i I_i^t \leq 2 v^{\mathrm{CO}_\epsilon}_{d_{\mathcal{M}}}(r^t, \mathbf{Y})$, where $r^t = \sum_{i=1}^k |U_i^t|$ is the number of pairwise preferences which cannot yet be decided with high probability.*

Remark 7. *Lemma 6 establishes the existence of a fast and easy method for computing the largest MRD distance possible, given some $\bar{\mathbf{Y}}$ and r. Needless to say, having an approximation with similar properties (at least for an approximation of the largest distance) for the NDP measure would be quite desirable. However, as it is not clear how such a result can be obtained (if at all), determining the complexity of this task is left as an open problem.*

Remark 8. *Lemma 6 assumes A^t to hold for a particular $t > 0$. This lemma can be restated so that it holds for any $t > 0$ with probability at least $1 - \delta$, since, according to Lemma 1, $\sum_{i=1}^K \sum_{j \neq i} \sum_{t=1}^\infty \mathbb{P}(A^t_{i,j}) \leq \delta$.*

Step 3: We will use $\Delta_{i,j} = |1/2 - y_{i,j}|$ as a complexity measure of the rank elicitation task. Furthermore, let $\Delta_{(r)}$ denote the r-th smallest value among $\Delta_{i,j}$ for all distinct $i, j \in [K]$. The next lemma upper-bounds (building on Lemma 6) the probability that $\textsc{RankEl}^{\mathrm{CO}_\epsilon}_{d_{\mathcal{M}}}$ does not terminate at iteration t.

Lemma 9. *With $A^t_{\mathcal{M}}$ the set of pairs $\textsc{RankEl}^{\mathrm{CO}_\epsilon}_{d_{\mathcal{M}}}$ samples in round t, it holds that*

$$\begin{aligned}&\mathbb{P}\left\{A^t_{\mathcal{M}} \neq \emptyset \wedge \forall (i,j) : (\Delta_{i,j} \geq \Delta_{(r_1)}) \Rightarrow (n^t_{i,j} > 2b^t_{i,j})\right\} \\ &\qquad \leq \frac{3\delta}{10K^2t^4} \sum_{r=1}^{K^2 - r_1} \frac{1}{\left(\Delta_{(r)} + \epsilon\right)^2} \ ,\end{aligned}$$

where $b^t_{i,j} = \left\lceil \frac{1}{2(\Delta_{i,j}+\epsilon)^2} \ln\left(\frac{5K^2t^4}{4\delta}\right) \right\rceil$ and $r_1 = 2 \operatorname{argmax}\left\{r \in [K^2] | v^{\mathrm{CO}_\epsilon}_{d_{\mathcal{M}}}(r, \mathbf{Y}) < \rho\right\}$.

Step 4: Using Lemmas 5 and 9, one can calculate an upper bound for the expected sample complexity of $\textsc{RankEl}^{\mathrm{CO}_\epsilon}_{d_{\mathcal{M}}}$.

Theorem 10. *Using the notation introduced in Lemma 9, the expected sample complexity for* $\textsc{RankEl}_{d_{\mathcal{M}}}^{\mathrm{CO}_\epsilon}$ *is* $O\left(R_1 \log\left(\frac{R_1}{\delta}\right)\right)$*, where* $R_1 = \sum_{r=1}^{K^2-r_1}\left(\Delta_{(r)}+\epsilon\right)^{-2}$.

Proof sketch: First, it can be shown that $\textsc{RankEl}_{d_{\mathcal{M}}}^{\mathrm{CO}_\epsilon}$ terminates before iteration $T \in O\left(R_1 \log\left(\frac{R_1}{\delta}\right)\right)$ if enough samples are drawn from each $Y_{i,j}$ ($n_{i,j}^t > 2b_{i,j}^t$ according to Lemma 9) and no error occurs for any of the $\bar{y}_{i,j}^t$ (Lemma 5). Consequently, after iteration T, the probability of an error along with the probability of the non-termination of the algorithm (if enough samples are drawn) upper-bounds the number of iterations taken by $\textsc{RankEl}_{d_{\mathcal{M}}}^{\mathrm{CO}_\epsilon}$ after T. This probability can be upper-bounded by $4/3\pi^2\delta$ for iterations $> T$ based on Lemmas 5 and 9. □

The expected sample complexity bound given in Theorem 10 is similar in spirit to the one given for LUCB1 in the framework of stochastic multi-armed bandits [Kalyanakrishnan et al., 2012], but the complexity measure of the rank elicitation task is essentially of different nature.

Expected sample complexity of $\textsc{RankEl}_{d_{\mathcal{M}}}^{\mathrm{SE}_\epsilon}$

The sample complexity analysis of $\textsc{RankEl}_{d_{\mathcal{M}}}^{\mathrm{SE}_\epsilon}$ is very similar to the one we carried out for the ϵ-insensitive Copeland ranking, although the complexity measure of the rank elicitation task in this case can be given as follows: let $\lambda_{i,j} = |y_i - y_j|$, and furthermore, let $\lambda_{(r)}$ denote the r-th smallest value among $\lambda_{i,j}$ for all distinct $i, j \in [K]$. Now, the expected sample complexity of $\textsc{RankEl}_{d_{\mathcal{M}}}^{\mathrm{SE}_\epsilon}$ can be upper-bounded in terms of $\Lambda_1 = \sum_{r=1}^{K^2-\ell_1}\left(\lambda_{(r)}+\epsilon\right)^{-2}$ (similarly to Theorem 10) where $\ell_1 = 2\,\mathrm{argmax}\left\{r \in [K^2] | v_{d_{\mathcal{M}}}^{\mathrm{SE}_\epsilon}(r, \mathbf{Y}) < \rho\right\}$. We omit the technical details, since the analysis is straightforward based on the previous section and [Kalyanakrishnan et al., 2012].

Experiments

To illustrate our PAC rank elicitation method, we applied it to sports data, namely the soccer matches of the last ten seasons of the German Bundesliga. Our goal was to learn the corresponding Copeland or SE ranking. We restricted to the 8 teams that participated in each Bundesliga season between 2002 to 2012. Each pair of teams o_i and o_j met 20 times; we denote the outcomes of these matches by $y_{i,j}^1, \ldots, y_{i,j}^{20}$ and take the corresponding frequency distribution as the (ground-truth) probability distribution of $Y_{i,j}$. The matrix $\mathbf{Y}$ thus obtained is shown in Figure 1(a).

As a baseline, we run the RankEl algorithm with uniform sampling, meaning that all pairwise comparisons are sampled in each iteration. The accuracy of a run is 1 if $d(\tau, \prec^{\mathcal{R}}) \leq \rho$ for the ranking τ that was produced, and 0 otherwise. The relative empirical sample complexity achieved by RankEl with respect to the uniform sampling is shown in Table 1(b) for various parameter settings. Our results confirm that RANKEL has a significantly smaller empirical sample complexity than uniform sampling (while providing the same guarantees in terms of approximation quality).

Conclusion and future work

We introduced a PAC rank elicitation problem and proposed an algorithm for solving this task, that is, for eliciting a ranking that is close to the underlying target order with high probability. Our algorithm consistently outperforms the uniform sampling strategy that was taken as a baseline. Moreover, it scales gracefully with the parameters ϵ and ρ that specify, respectively, the strictness of the target order and the sought quality of approximation to that order.

There is still a number of theoretical questions to be addressed in future work, as well as interesting variants of our setting. First, as mentioned in Remark 7, the sample complexity for $\textsc{RankEl}_{d_{\mathcal{K}}}^{\mathrm{SE}_\epsilon}$ and $\textsc{RankEl}_{d_{\mathcal{K}}}^{\overline{\mathrm{SE}}_\epsilon}$ is still an open question. Second, noting that the $Y_{i,j}$ are trinomial random variables for which a Clopper-Pearson-type high probability confidence bound exists [Chafaï and Concordet, 2009], there is hope to significantly improve our

B. Munchen	0.5	0.7	0.55	0.575	0.75	0.55	0.775	0.7	[7,7] [4,7] [0.73,0.75] [0.73,0.83]
B. Dortmund	0.3	0.5	0.55	0.475	0.425	0.525	0.6	0.675	[4,4] [1,6] [0.58,0.60] [0.58,0.68]
B. Leverkusen	0.45	0.45	0.5	0.425	0.55	0.55	0.65	0.6	[4,4] [1,7] [0.60,0.62] [0.60,0.70]
VfB Stuttgart	0.425	0.525	0.575	0.5	0.4	0.6	0.5	0.65	[4,5] [1,7] [0.60,0.62] [0.60,0.70]
Schalke 04	0.25	0.575	0.45	0.6	0.5	0.45	0.65	0.675	[4,4] [2,6] [0.59,0.61] [0.59,0.69]
W. Bremen	0.45	0.475	0.45	0.4	0.55	0.5	0.55	0.65	[3,3] [1,7] [0.58,0.60] [0.58,0.68]
VfL Wolfsburg	0.225	0.4	0.35	0.5	0.35	0.45	0.5	0.675	[1,2] [1,4] [0.49,0.51] [0.49,0.59]
Hannover 96	0.3	0.325	0.4	0.35	0.325	0.35	0.325	0.5	[0,0] [0,1] [0.41,0.43] [0.41,0.51]

(a) Matrix $\mathbf{Y}$ for Bundesliga data, and the intervals for the interval orders $\prec^{\mathrm{CO}_{0.02}}$, $\prec^{\mathrm{CO}_{0.1}}$, $\prec^{\mathrm{SE}_{0.02}}$ and $\prec^{\mathrm{SE}_{0.1}}$, respectively

$\prec^*$	$d(.,.)$	ρ	ϵ	Improvement (%)
$\prec^{\mathrm{CO}}$	$d_{\mathcal{K}}$	3	0.02	25.3 ± 0.4
$\prec^{\mathrm{CO}}$	$d_{\mathcal{M}}$	3	0.02	24.0 ± 0.4
$\prec^{\mathrm{SE}}$	$d_{\mathcal{K}}$	3	0.02	21.9 ± 0.2
$\prec^{\mathrm{SE}}$	$d_{\mathcal{M}}$	3	0.02	23.1 ± 0.2
$\prec^{\mathrm{CO}}$	$d_{\mathcal{K}}$	3	0.1	43.6 ± 0.7
$\prec^{\mathrm{CO}}$	$d_{\mathcal{M}}$	3	0.1	43.9 ± 0.7
$\prec^{\mathrm{SE}}$	$d_{\mathcal{K}}$	3	0.1	24.7 ± 0.1
$\prec^{\mathrm{SE}}$	$d_{\mathcal{M}}$	3	0.1	23.5 ± 0.2
$\prec^{\mathrm{CO}}$	$d_{\mathcal{K}}$	5	0.1	49.1 ± 0.6
$\prec^{\mathrm{CO}}$	$d_{\mathcal{M}}$	5	0.1	64.3 ± 0.8
$\prec^{\mathrm{SE}}$	$d_{\mathcal{K}}$	5	0.1	25.4 ± 0.2
$\prec^{\mathrm{SE}}$	$d_{\mathcal{M}}$	5	0.1	31.8 ± 0.4

(b) Improvement in empirical sample complexity

Figure 1: The top panel (1(a)) shows the matrix $\mathbf{Y}$ for the Bundesliga data, and the $[d_i^*, d_i^* + s_i^*]$ intervals for $\prec^{\mathrm{CO}_{0.02}}$ and $\prec^{\mathrm{CO}_{0.1}}$, and the $[y_i, y_i + \epsilon]$ intervals for $\prec^{\mathrm{SE}_{0.02}}$ and $\prec^{\mathrm{SE}_{0.1}}$, respectively. The bottom panel (1(b)) shows the reduction of the empirical sample complexity achieved by RANKEL for various parameter settings, taking the complexity of uniform sampling as 100%. Mean and standard deviation of the improvement were obtained by averaging over 100 repetitions. The confidence parameter δ was set to 0.1 for each run; accordingly, the average accuracy was significantly above $1 - \delta = 0.9$ in each case.

bound on expected sample complexity. Third, based on [Kalyanakrishnan et al., 2012], a high probability bound for the sample complexity might be devised instead of the expected complexity bound. Last but not least, there are other interesting ranking procedures $\mathcal{R}$ and distance measures that can be used to instantiate our setting.

Acknowledgments

This work was supported by the German Research Foundation (DFG) as part of the Priority Programme 1527.

References

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

[Braverman and Mossel, 2008] Braverman, M. and Mossel, E. (2008). Noisy sorting without resampling. In *Proceedings of the nineteenth annual ACM-SIAM Symposium on Discrete algorithms*, pages 268–276.

[Braverman and Mossel, 2009] Braverman, M. and Mossel, E. (2009). Sorting from noisy information. *CoRR*, abs/0910.1191.

[Bubeck et al., 2009] Bubeck, S., Munos, R., and Stoltz, G. (2009). Pure exploration in multi-armed bandits problems. In *Proceedings of the 20th international conference on Algorithmic learning theory*, ALT'09, pages 23–37, Berlin, Heidelberg. Springer-Verlag.

[Bubeck et al., 2013] Bubeck, S., Wang, T., and Viswanathan, N. (2013). Multiple identifications in multi-armed bandits. In *Proceedings of The 30th International Conference on Machine Learning*, pages 258–265.

[Busa-Fekete et al., 2013] Busa-Fekete, R., Szörényi, B., Weng, P., Cheng, W., and Hüllermeier, E. (2013). Top-k selection based on adaptive sampling of noisy preferences. In *Proceedings of the 30th International Conference on Machine Learning, JMLR W&CP*, volume 28.

[Cappé et al., 2012] Cappé, O., Garivier, A., Maillard, O.-A., Munos, R., and Stoltz, G. (2012). Kullback-Leibler upper confidence bounds for optimal sequential allocation. *Submitted to the Annals of Statistics*.

[Chafaï and Concordet, 2009] Chafaï, D. and Concordet, D. (2009). Confidence regions for the multinomial parameter with small sample size. *Journal of the American Statistical Association*, 104(487):1071–1079.

[Chen et al., 2013] Chen, X., Bennett, P. N., Collins-Thompson, K., and Horvitz, E. (2013). Pairwise ranking aggregation in a crowdsourced setting. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 193–202.

[Eriksson, 2013] Eriksson, B. (2013). Learning to Top-K search using pairwise comparisons. *Journal of Machine Learning Research - Proceedings Track*, 31:265–273.

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

[Feige et al., 1994] Feige, U., Raghavan, P., Peleg, D., and Upfal, E. (1994). Computing with noisy information. *SIAM J. Comput.*, 23(5):1001–1018.

[Gabillon et al., 2011] Gabillon, V., Ghavamzadeh, M., Lazaric, A., and Bubeck, S. (2011). Multi-bandit best arm identification. In Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F., and Weinberger, K., editors, *Advances in Neural Information Processing Systems 24*, pages 2222–2230. MIT.

[Guo et al., 2012] Guo, S., Sanner, S., Graepel, T., and Buntine, W. (2012). Score-based bayesian skill learning. In *European Conference on Machine Learning*, pages 1–16, Bristol, UK.

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

[Kalyanakrishnan, 2011] Kalyanakrishnan, S. (2011). *Learning Methods for Sequential Decision Making with Imperfect Representations*. PhD thesis, University of Texas at Austin.

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

[Kendall, 1955] Kendall, M. (1955). *Rank correlation methods*. Charles Griffin, London.

[Moulin, 1988] Moulin, H. (1988). *Axioms of cooperative decision making*. Cambridge University Press.

[Urvoy et al., 2013] Urvoy, T., Clerot, F., Féraud, R., and Naamane, S. (2013). Generic exploration and K-armed voting bandits. In *Proceedings of the 30th International Conference on Machine Learning, JMLR W&CP*, volume 28, pages 91–99.

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