

Characterization of Scoring Rules with Distances: Application to the Clustering of Rankings *

Paolo Viappiani

- 1) Sorbonne Universités, UPMC Univ Paris 06, UMR 7606, LIP6
 - 2) CNRS, UMR 7606, LIP6, 4 Place Jussieu, 75005 Paris, France
- paolo.viappiani@lip6.fr

Abstract

Positional scoring rules are often used for rank aggregation. In this work we study how scoring rules can be formulated as the minimization of some distance measures between rankings, and we also consider a new family of aggregation methods, called biased scoring rules. This work extends a previous known observation connecting Borda count with the minimization of the sum of the Spearman distances (calculated with respect to a set of input rankings). In particular we consider generalizations of the Spearman distance that can give different weights to items and positions; we also handle the case of incomplete rank data.

This has applications in the clustering of rank data, where two main steps need to be performed: aggregating rankings of the same cluster into a representative ranking (the cluster's centroid) and assigning each ranking to its closest centroid. Using the proper combination of scoring rules (for aggregation) and distances (for assignment), it is possible to perform clustering in a computationally efficient way and as well account for specific desired behaviors (give more weight to top positions, bias the centroids in favor of particular items).

1 Introduction

It is often the case that data is available in the form of rankings (ordered lists of elements that express a preference order), for instance, this is the case of preference information obtained in electronic commerce applications. Clustering is useful in the sense that by producing a (small) number of aggregated rankings we are able to provide a meaningful qualitative description for the entire population. In this work we focus on distance-based methods for clustering that are attractive because they do not make specific assumptions (contrary to probabilistic methods [Critchlow *et al.*, 1991; Lu and Boutilier, 2011] that assume a specific generative model for the rankings).

A distance-based clustering method partitions the elements into clusters, so that the within-cluster distance is minimized;

each cluster can then be associated with a representative element of the partition (the centroid). When considering rankings as elements to be clustered, the issue is to define a meaningful sound distance measure that can be used, as several alternative possibilities exist (for instance, Kendall tau, Spearman or footrule). As noted by a number of researchers [Sculley, 2007; Kumar and Vassilvitskii, 2010], in many applications the distance measure should account for the fact that we may wish to weigh more top positions in the rankings and/or give more importance to specific items.

A related problem is ranking aggregation. Several methods for aggregation have been proposed, notably in the social choice theory community. In particular, in positional scoring rules, each item gets some points from each ranking according to its position. When a distance measure is defined, a natural aggregation method is to look for the ranking that minimizes the sum of the distances with all input rankings.

An interesting question is whether an aggregation method has a corresponding distance that they minimize implicitly: if so, we then say that such distance characterizes the given aggregation method. In this work we show how scoring rules (including some new variants, biased scoring rules) can be characterized. This has practical implications; it allows to perform clustering in a computationally efficient way. Moreover, many of our results involve new generalizations of Spearman distance that account for the importance associated to positions and to items, allowing for greater flexibility.

2 Distance-based Clustering

We have a set of n items or objects and a set of m users or agents. A ranking π is a permutation on the set of available objects. Formally π is a function from $\{1, \dots, n\}$ to $\{1, \dots, n\}$ associating each item with its position (rank). As usual, the set of possible rankings is denoted as S_n . A ranking can be expressed alternatively in an explicit form of a tuple $\langle \pi^{-1}(1), \dots, \pi^{-1}(n) \rangle$, with $\pi^{-1}(r)$ being the r -th most preferred item; for example $\langle 2, 1, 3 \rangle$ is the ranking for which item 2 is the most preferred, then item 1 is preferred, and finally item 3 is the least preferred (this corresponds to $\pi(1) = 2, \pi(2) = 1$ and $\pi(3) = 3$).

We consider to have a number of rankings π_1, \dots, π_m , associated with different users, and we want to partition them into different clusters. Let $f: i \rightarrow r$ assign rankings to clusters and d be a distance. Distance-based clustering is the problem of,

*Supported by ANR-14-CE24-0007-01 "CoCoRiCo-CoDec".

given a number m of rankings, partitioning them in k clusters (or classes) in order to minimize the within-cluster sum of distances with respect to the “central” rankings $\bar{\pi}_1^*, \dots, \bar{\pi}_k^*$ of each cluster.

$$(f^*, \bar{\pi}_1^*, \dots, \bar{\pi}_k^*) = \arg \min_{f, \bar{\pi}_1, \dots, \bar{\pi}_k} \sum_{z=1}^k \sum_{j:f(j)=z} d(\bar{\pi}_z, \pi_j) \quad (1)$$

The problem of clustering is frequently tackled in the literature with an iterative algorithm that proceeds in two steps. In the *assignment* step, each observation is assigned to the cluster whose “mean” yields the least within-cluster distance. In the *update* step, we calculate the new means to be the centroids of the observations in the new clusters. When items are vectors and the Euclidean distance is used, the problem is that of k-means clustering and the iterative algorithm described above is often called as well k-means¹.

Following [Kamishima *et al.*, 2005] we adopt the same idea for clustering a set of rankings. We maintain a set of centroids (initialized randomly) and we associate each ranking with the cluster whose centroid is closest, according to a suitable distance d on rankings. We then recompute the centroids for each of the clusters. We alternate between these two actions until the clusters do not change anymore.

Algorithm 1: Distance-based clustering of rankings.

Data: π_1, \dots, π_m (population of rankings given by m users), k (number of clusters)

Randomly initialize the centroids $\bar{\pi}_1, \dots, \bar{\pi}_k$;

while there are changes in cluster assignments **do**

Assign each ranking π to the cluster whose distance to the centroid is lower

$f(\pi_i) := \arg \min_{r=1, \dots, k} d(\pi_i, \bar{\pi}^r) \quad i = 1, \dots, m$;

Find the centroid of each cluster $\bar{\pi}^r :=$

$\arg \min_{\pi \in S_n} \sum_{\pi_t: f(\pi_t)=r} d(\pi, \pi_t) \quad r = 1, \dots, k$;

end

A well known fact is that Algorithm 1 converges to a local optimum. This is easily proved by showing that both main steps of the algorithm cannot increase the total distances of each data point to the centroid of its cluster.²

Algorithm 1 requires the specification of a suitable distance measure. Rankings are particular “objects”, and there are many different ways to define a distance between two rankings; this is discussed in Section 3.2. Note however, that rank aggregation is often treated as a separate problem, especially in social choice literature, notably in voting methods (an aggregate ranking is obtained without considering an underlying distance measure).

The main contribution of this paper is to study the connection between rank aggregation using scoring rules and ranking distances. We will propose new distances that allow to assign different degrees of importance to positions and to items.

¹Also sometimes called Lloyd’s algorithm.

²Since the distance-based clustering approach returns a local optimum, the algorithm may be run for a number of times (typically 10 or 20) and store the clustering assignment associated with the lowest sum of distances to the closest centroid.

Since these distances are easy to aggregate (in our terminology, distances that *characterize* a scoring rule), clustering can be computed very efficiently.

3 Distances and Aggregation

There are a number of ways that can be used to aggregate several rankings into a single one. Some aggregation rules are devised from social choice: the Condorcet method, sorting the items by their Borda score or a generic scoring rule. There are as well many commonly used distance measures for rankings, such as Kendall-tau, footrule or Spearman.

From a theoretical point of view, the interest is to study if, for a given common aggregation rule (such as plurality), there is a distance measure that it is (implicitly) minimized. In this paper we establish a connection between scoring rules (often used in social choice) and their associated distance measures.

3.1 Aggregation with Scoring Rules

An aggregation rule is a mapping $g(\sigma_1, \dots, \sigma_m)$ from a set of input rankings $\sigma_1, \dots, \sigma_m$ to a single “best” ranking summarizing the whole population. We allow g to return more than one ranking, to be considered equally good.

Many ways of aggregating rankings arise from the field of social choice, where one needs to make a decision for a group of people, aggregating several (usually different) preferences, expressed as a vote in a ballot. Here we focus on rank aggregation using scoring rules.

A scoring rule associates each position $r \in \{1, \dots, n\}$ with a score $w(r)$; scores are weakly decreasing $w(1) \geq \dots \geq w(n)$. Items are evaluated by summing up the score they are awarded in each ranking $v(i) = \sum_{u=1}^m w(\pi_u(i))$. In order to form an aggregate ranking, items are sorted according to their total score $v(i)$: the ranking π_{SR}^* obtained by a scoring rule is such that $\pi_{SR}^*(i) < \pi_{SR}^*(j)$ if $v(i) > v(j)$ (when ties exist in the overall score, tie-breaking is needed).

Borda count (or Borda rule) is a particular type of scoring rule considering weights defined as $w(r) = n - r + 1$ (the item ranked first gets a score of n points, an item in the second position gets $n - 1$, and so on). We denote with π_{Borda}^* the ranking obtained by following Borda rule. Borda weights are such that Borda counts for element i are $v(i) = \sum_{u=1}^m n - \sigma_u(i) + 1 = m(n + 1) - \sum_{u=1}^m \sigma_u(i)$. The optimal ranking π_{Borda}^* according to Borda count is such that i precedes j in π_{Borda}^* , formally expressed as $\pi_{Borda}^*(i) < \pi_{Borda}^*(j)$ if $\sum_{u=1}^m \sigma_u(i) < \sum_{u=1}^m \sigma_u(j)$ (an item i is ranked better than another item j if the sum of its ranks is lower)

Plurality (sorting items by the number of times that they are ranked first) can be represented as a scoring rule with weights $(1, 0, \dots, 0)$; *veto*, sorting items in decreasing order with respect to the number of times they are ranked in the last position, is represented by weights $(1, \dots, 1, 0)$, and top- k , that can be modelled as a scoring rule with a weight of 1 in the first k positions and then 0.

We propose a new aggregation method, that we call *biased scoring rule* parametrized by two vectors z_1, \dots, z_n (between 0 and 1, a multiplicative bias distorting each item’s score) and ϕ_1, \dots, ϕ_n (representing an additive bias associated to each item). Each item i receives a contribution $z_i w(\sigma_u(i))$ to its

score for each ranking σ_u plus an overall bonus ϕ_i :

$$v_{BSR}(i) = \phi_i + z_i \cdot v_{SR}(i) = z_i = \phi_i + z_i \sum_{u=1}^m w(\sigma_u(i)) \quad (2)$$

This allows to “tweak” a scoring role in order to give some advantage to some items, or penalize others. Obviously when $z_i = 1$ for all i , and ϕ_i is the same for all items, aggregation with v_{BSR} reduces to an unbiased scoring rule. If $w(r) = n - r + 1$, then we have a *biased Borda count*.

3.2 Distance Measures for Rankings

Distance measures characterize how different two rankings are; different distances might pose more strength on specific aspects: penalizing the displacements in different ways. For a given distance measure d the total distance from a given ranking π to a set of rankings $\sigma_1, \dots, \sigma_m$ is $D(\pi; \sigma_1, \dots, \sigma_m) = \sum_{u=1}^m d(\pi, \sigma_u)$. A distance function between rankings naturally leads to a way to generating an aggregate ranking; the ranking minimizing this score is chosen as aggregated ranking for the population: $\pi^* = \arg \min_{\pi} D(\pi, \sigma_1, \dots, \sigma_m)$.

We recall hereafter the usual definition of metric and common generalizations relaxing some of the properties. Note that, while the term distance is often used as a synonym for metric, in the following, we use the former to loosely mean any function that quantifies the difference between elements (rankings in our case), and we explicitly state which distances are metric.

Definition 1. A function $d : X \times X \rightarrow \mathbb{R}$ is a metric on X iff it satisfies the following properties:

- $d(x, y) \geq 0$ (non negativity),
- $d(x, y) = 0$ iff $x = y$ (identity of the indiscernibles),
- $d(x, y) = d(y, x)$ (symmetry) and
- $d(x, y) + d(y, z) \geq d(x, z)$ (triangular inequality).

Moreover we have the following relaxations:

- A pseudometric d relaxes the identity of the indiscernibles ($d(x, x) = 0$ but it may hold $d(x, y) = 0$ for $y \neq x$);
- A quasimetric relaxes symmetry;
- A semimetric relaxes the triangular inequality;
- A function satisfying non-negativity and $d(x, x) = 0$ is a premetric.

We are interested in distance measures on rankings; ranking distances d are defined on the S_n (the set of permutation of n elements). Common distance measures for rankings are Kendall tau (counting the number of disagreements in terms of pairs between two rankings), footrule³ (measuring the total displacement of all elements) and Spearman.

In this paper we focus on *Spearman distance*, because as will be discussed below in Section 4, it is connected to the aggregation using Borda count. The Spearman distance is defined

³Also known as Spearman’s footrule.

as taking the squares of the differences:

$$d_S(\pi, \sigma) = \sum_{j=1}^n [\pi(j) - \sigma(j)]^2. \quad (3)$$

An interesting observation, that we will use several times in our proofs, is that Spearman can be expressed as follows:

$$d_S(\pi, \sigma) = \frac{n(n+1)(2n+1)}{3} - 2 \sum_{i=1}^n \pi(i)\sigma(i). \quad (4)$$

While footrule and Kendall-tau distance are metrics, Spearman distance does not satisfy the triangular inequality and is a semimetric. It can be turned into a metric if we take the root of sum of the squares of the differences between positions.

The traditional definition of Spearman distance treats all positions in the same way. Following [Dwork *et al.*, 2001; Kumar and Vassilvitskii, 2010], in order to allow to put more emphasis on some ranks we define a generalization of Spearman distance, that we call *positional Spearman*, giving different weights to rank positions, computed as

$$d_{PS}(\pi, \sigma) = \sum_{i=1}^n [w(\pi(i)) - w(\sigma(i))]^2 \quad (5)$$

parametrized by a vector w . We will often use the observation that d_{PS} can be rewritten as

$$d_{PS}(\pi, \sigma) = Z_n^w - 2 \sum_{i=1}^n w(\pi(i)) w(\sigma(i)) \quad (6)$$

where $Z_n^w = 2 \sum_{r=1}^n w(r)^2$ depends only on the weights w and the number of items n .

Observation 1. d_{PS} is a semimetric if the weights w are strictly decreasing; otherwise it is a pseudo-semimetric.

Consider, for instance, $w = (3, 2, 2, 1)$, two rankings that differ only by the fact that items on the second and third position are inverted are associated with null distance according to w : $d_{PS}(\langle 1, 2, 3, 4 \rangle, \langle 1, 3, 2, 4 \rangle) = 0$.

4 Connection between Aggregation Methods and Distance Minimization

The general aggregation problem is that of finding the ranking (permutation of items) that minimizes a given distance measure with respect to several other given rankings $\sigma_1, \dots, \sigma_m$ given.

$$\pi^* = \arg \min_{\pi} D(\pi; \sigma_1, \dots, \sigma_m) = \arg \min_{\pi} \sum_{j=1}^m d(\pi, \sigma_j). \quad (7)$$

Definition 2. A distance function $d(\pi, \sigma)$ on rankings characterizes a ranking aggregation $g(\sigma_1, \dots, \sigma_m)$ iff it holds

$$\arg \min_{\pi \in S_n} D(\pi; \sigma_1, \dots, \sigma_m) = g(\sigma_1, \dots, \sigma_m)$$

with $D(\pi; \sigma_1, \dots, \sigma_m) = \sum_{u=1}^m d(\pi, \sigma_u)$.

Table 1: Distance measures and associated scoring functions.

Aggregation method	Distance measure	Properties
Plurality	d_{PL}	premetric
Top-k	d_{TK}	premetric
Veto	d_V	premetric
Borda	Spearman	semimetric
Scoring rule (strictly decreasing weights)	positional Spearman	semimetric
Biased scoring rule	item-weighting pos. Spearman	non negativity

In case the aggregation function returns multiple rankings⁴, then all such rankings should achieve the same minimal value $D^* = \min_{\pi} D(\pi; \sigma_1, \dots, \sigma_m)$, and, conversely, if there are several rankings associated with minimum sum-of-distances D^* , these must be returned by g .

We now establish⁵ theoretical connections between distance minimization and scoring rules. In particular, we establish connections between the newly proposed positional Spearman distance (see above in Section 3.2) and aggregation using scoring rules, and between a distance giving different weights to items and aggregation using biased scoring rules (defined above in Section 3.1).

First of all, rank aggregation with Borda count and minimization of Spearman produce the same aggregated ranking.

Proposition 1. [Marden, 1996]

The Spearman distance characterizes the Borda rule: $\pi_{Borda}^* = \arg \min_{\pi \in S_n} \sum_{u=1}^m d_S(\pi, \sigma_u)$.

We have noted before that Spearman distance, as defined in Equation 3, is not a metric, as triangular inequalities does not hold. One might wonder if it is possible to “tweak” the Spearman distance to find a metric characterizing Borda. We prove, however, that this is not possible.

Proposition 2. There is no metric characterizing Borda rule.

Proof : Let d be a semimetric between rankings. We prove that if d characterizes Borda rule, then it cannot satisfy the triangular inequality. Consider the following population of rankings: $\sigma_1 = \langle 1, 2, 3 \rangle$ and $\sigma_2 = \langle 3, 1, 2 \rangle$. Application of Borda gives the following scores to items: $v(1) = 5, v(2) = 3, v(3) = 4$ yielding the optimal ranking $\pi^* = \langle 1, 3, 2 \rangle$. If d characterizes Borda rule, then it must hold that $D(\pi^*; \sigma_1, \sigma_2) < D(\pi; \sigma_1, \sigma_2)$, for any $\pi \in S_n, \pi \neq \pi^*$. In particular, π^* must compare favorably with respect to σ_1 : $\sum_{u=1,2} d(\pi^*, \sigma_u) < \sum_{u=1,2} d(\sigma_1, \sigma_u)$. Since d is a semimetric, we must have $d(\pi^*, \sigma_1) = d(\sigma_1, \pi^*)$ and $d(\sigma_1, \sigma_1) = 0$; it then follows that $d(\sigma_1, \pi^*) + d(\pi^*, \sigma_2) < d(\sigma_1, \sigma_2)$. ■

We now extend this result to scoring rules assigning ar-

⁴For example, due to ties in the score obtained with Borda count.

⁵We provide proofs for selected statements. Proofs for all propositions will be provided in a longer version of the paper.

bitrary weights to positions. We derive a novel connection between scoring rules and our proposed positional Spearman distance. Note, however, that the scoring vector must (in addition of being non increasing) not assign the same weights to more than one position.

Proposition 3. Assume a scoring rule with strictly decreasing weights $w(r) > w(s)$ if $r < s$ with $r, s \in \{1, \dots, n\}$. The positional Spearman distance with weights w characterizes the scoring rule with the same weights; $\pi_{SR}^* = \arg \min_{\pi \in S_n} \sum_{u=1}^m d_{PS}(\pi, \sigma_u)$.

Proof : The optimal ranking π_{SR}^* wrt a scoring rule with weights w is such that i precedes j in π^* iff $v(i) \geq v(j)$, or equivalently iff $\sum_{u=1}^m w(\sigma_u(i)) \geq \sum_{u=1}^m w(\sigma_u(j))$. The ranking with minimum total distance wrt positional Spearman is (we use Equation 6):

$$\arg \min_{\pi} \sum_{u=1}^m d_{PS}(\pi, \sigma_u) = \arg \max_{\pi} \sum_{i=1}^n w(\pi(i)) \sum_{u=1}^m w(\sigma_j(i)). \quad (8)$$

The π maximizing the expression above is such that $w(\pi(i)) \geq w(\pi(j))$ if $\sum_{u=1}^m w(\sigma_u(i)) \geq \sum_{u=1}^m w(\sigma_u(j))$. If w is strictly decreasing this gives π_{SR}^* . ■

The previous observations only holds for scoring rules with distinct weights. This means that the association fails, notably, for plurality, veto and top-k, that can be represented, respectively, as scoring rules with weights $(1, 0, \dots, 0)$, $(0, \dots, 0, 1)$ and with a weight vector with 1 in the first k positions and then 0s everywhere. We therefore look for an alternative characterization for these rules, in order to define some distance measures that they implicitly minimize.

Using plurality in our framework for clustering intuitively means to put together rankings based on the first preferred item. If the number of clusters is lower than the number of distinct items placed first in any ranking, aggregation will be made by ordering items according to the number of “votes” (number of rankings placing an item first); when assigning rankings to clusters, a ranking with item i in the first position will be assigned to the cluster whose centroid put item i in the highest position. The following premetric d_{PL} captures this behavior

$$d_{PL}(\pi, \sigma) = \pi(\sigma^{-1}(1)) - 1. \quad (9)$$

Note that d_{PL} is not symmetric. Furthermore, for a given π there are many π' such that $d(\pi, \pi') = 0$; in fact, any σ such that $\sigma^{-1}(1) = \pi^{-1}(1)$: for example $d_{PL}(\langle 1, 2, 3 \rangle, \langle 1, 3, 2 \rangle) = 0$. Therefore d_{PL} is neither a metric nor a semimetric. It holds $d_{PL}(\pi, \pi) = 0$ for any π and $d_{PL}(\pi, \sigma) \geq 0$ for any π, σ , but not the triangular inequality; thus d_{PL} is a premetric.

We now present our result about the characterization of plurality: we prove that the ranking obtained by aggregating the rankings $\sigma_1, \dots, \sigma_m$ using plurality minimizes the sum of distances $D_{PL}(\pi; \sigma_1, \dots, \sigma_m) = \sum_{t=1}^m d_{PL}(\pi, \sigma_t)$.

Proposition 4. The distance d_{PL} characterizes plurality as a method for aggregation of rankings.

Proof : $\alpha_i^1 = |\{\sigma : \sigma(i) = 1\}|$ is the number of input rankings among $\sigma_1, \dots, \sigma_m$ in which item i is ranked first. It follows

$\sum_{t=1}^m d_{PL}(\pi, \sigma_t) = \sum_{i=1}^n \alpha_i^1 \pi(i) - m$. Therefore we have

$$\pi^* = \arg \min_{\pi \in S_n} \sum_{t=1}^m d_{PL}(\sigma_t, \pi) = \arg \min_{\pi \in S_n} \sum_{i=1}^n \alpha_i^1 \pi(i).$$

The permutation π^* is such that $\pi^*(i) < \pi^*(j)$ (i ranked before j) if $\alpha_i^1 > \alpha_j^1$; this is exactly the result of aggregation when using plurality. ■

One can wonder if there is another distance that can characterize plurality, with additional properties such as symmetry. In fact, we show that this is not possible.

Proposition 5. *There is no semimetric and no quasi metric (hence there is no metric) characterizing plurality.*

Proof: Consider a population of two rankings $\sigma_1 = \langle 1, 2, 3 \rangle$ and $\sigma_2 = \langle 2, 3, 1 \rangle$. According to plurality, the best rankings obtained by aggregating σ_1 and σ_2 are σ_1 itself and the ranking $\langle 2, 1, 3 \rangle$. Now, consider a premetric d and assume that it characterizes plurality. Since σ_2 is not a optimal ranking according to plurality, the sum of the distances between σ_1 and the population must be strictly lower than the sum of the distances from σ_2

$$d(\sigma_1, \sigma_1) + d(\sigma_1, \sigma_2) < d(\sigma_2, \sigma_2) + d(\sigma_2, \sigma_1)$$

from which (since $d(\pi, \pi) = 0$, d being a premetric) it follows $d(\sigma_1, \sigma_2) < d(\sigma_2, \sigma_1)$, hence any d characterizing plurality cannot be symmetric. ■

For veto, we can define a premetric, analogous to the one we defined for plurality, but that looks for the position of the lowest ranked items. We similarly characterize aggregation with respect to the *top-k* elements.

Proposition 6. *The pseudo-distance d_{VT} characterizes the veto rule.*

$$d_{VT}(\pi, \sigma) = n - \pi(\sigma^{-1}(n)). \quad (10)$$

Proposition 7. *The following premetric⁶ characterizes the top-k aggregation rule.*

$$d_{topk}(\pi, \sigma) = \sum_{r=1}^k \pi(\sigma^{-1}(r)) - \frac{n(n+1)}{2} \quad (11)$$

Note that d_{topk} is the same as d_{PL} when $k = 1$. From the fact that top-k aggregation subsumes plurality, and from Proposition 5, it immediately follows that there is no semimetric, no quasi metric and no metric, characterizing top-k.

In order to characterize biased scoring rules, we introduce another kind of generalization of Spearman, allowing to give different weights z_i, ϕ_i to items. *Item-weighting positional Spearman* is defined (Compare with Equation 4 and 6) as:

$$d_{IPS}(\pi, \sigma) = Z_n^w - 2 \sum_{i=1}^n w(\pi(i)) [z_i w(\sigma(i)) + \phi_i] \quad (12)$$

where $Z_n^w = 2 \sum_{i=1}^n w(i)$ depends only on w and n . The role of the z_i, ϕ_i is to weigh more the important items. Note that $d_{IPS}(\pi, \sigma) = d_{PS}(\pi, \sigma)$ if all weights z_i are set to 1 and all ϕ_i to zero.

We now establish the connection between d_{IPS} and the biased scoring rules presented before.

⁶The constant addend $-\frac{n(n+1)}{2}$ is used in order to satisfy $d(\pi, \pi) = 0$, i.e. to obtain a premetric.

Proposition 8. *The Item-weighted Positional Spearman distance d_{IPS} with strictly decreasing weights (w_1, \dots, w_n) , arbitrary (z_1, \dots, z_n) , and $(\phi_1/m, \dots, \phi_n/m)$ characterizes a biased scoring rule with (w_1, \dots, w_n) , (z_1, \dots, z_n) and (ϕ_1, \dots, ϕ_n) ; with m being the number of input rankings.*

Note that d_{IPS} is not even a premetric, as $d_{IPS}(\pi, \pi)$ can (and will often) yield a value different than zero. d_{IPS} is symmetric only if the bias terms ϕ_i are zero; it is non-negative assuming the z_i bounded by 1 and the ϕ_i non negative. A much “nicer” distance function (in the case $\phi_i = 0$) is the following $\hat{d}_{IPS}(\pi, \sigma) = \sum_{i=1}^n z_i [w(\pi(i)) - w(\sigma(i))]^2$ that is a semimetric; notice that $\hat{d}_{IPS}(\pi, \sigma) = d_{IPS}(\pi, \sigma) - d_{IPS}(\pi, \pi) - d_{IPS}(\sigma, \sigma)$. However, we could not find a characterization for this distance, and its optimization seems to be rather hard.

The theoretical results are summarized in Table 1.

5 Incomplete Rankings

In realistic cases it will be often the case that only a subset of items are ranked. Formally a partial ranking σ is defined on a restricted domain $\mathcal{D}(\sigma)$ (the items involved in the partial ranking), with $l = |\mathcal{D}(\sigma)|$ be the length. Given a partial ranking σ , we use $S_n(\sigma)$ for the set of consistent full rankings.

In our setting, aggregation produces a full ranking, while the distance measures how far a partial ranking is from each centroid. Given a partial ranking, we consider a probability distribution over the rankings that are consistent with such partial ranking. According to order statistics, and as noticed by [Kamishima and Akaho, 2009], given a partial ranking σ , under the uniform distribution the expected rank of an item i is $\frac{n+1}{l+1} \sigma(i)$ if $i \in \mathcal{D}(\sigma)$. If i is not ranked in σ ($i \notin \mathcal{D}(\sigma)$), its expected rank is $\frac{n+1}{2}$. *Expected Borda count* ranks items by increasing order of their sum of expected ranks.

The expected Spearman distance is naturally defined in the case of an uniform prior as (there are $\frac{n!}{l!}$ complete rankings consistent with a partial ranking of length l):

$$\begin{aligned} d_{E(S)}(\pi, \sigma) &= \sum_{\hat{\sigma} \in S_n(\sigma)} \frac{l!}{n!} d_S(\pi, \hat{\sigma}) = \\ &= C_n - 2 \frac{l!}{n!} \sum_{\hat{\sigma} \in S_n(\sigma)} \sum_{i=1}^n \pi(i) \hat{\sigma}(i). \end{aligned}$$

Proposition 9. [Kamishima and Akaho, 2009]

$d_{E(S)}$ characterizes expected Borda count.

We now generalize scoring rules and the positional Spearman distance to partial rankings. Assuming an uniform distribution over the complete orders consistent with an observed partial order σ , the expectation $E[v(i)]$ of the score of an item i is obtained by averaging over the score obtained by i in all complete consistent orders; $E[v_{SR}(i)] = \sum_{\hat{\sigma} \in S_n(\sigma)} \frac{l!}{n!} w(\hat{\sigma}(i))$. The *expected scoring rule* ranks items according to $E[v(i)]$; similarly we can define the *expected biased scoring rule* (refer to Equation 2), with $E[v_{BSR}(i)] = \phi_i + z_i E[v_{SR}(i)]$.

We introduce the *expected positional Spearman* distance $d_{E(PS)}(\pi, \sigma) = \frac{l!}{n!} \sum_{\hat{\sigma} \in S_n(\sigma)} d_{PS}(\pi, \hat{\sigma})$ and $d_{E(IPS)}$ in an analogous way: $d_{E(IPS)}(\pi, \sigma) = \frac{l!}{n!} \sum_{\hat{\sigma} \in S_n(\sigma)} d_{IPS}(\pi, \hat{\sigma})$.

Proposition 10.

- $d_{E(PS)}$ characterizes the expected scoring rule,
- $d_{E(IPS)}$ characterizes the expected biased scoring rule.

While expected Borda can be computed efficiently, its associated distance $d_{E(S)}$ is computationally expensive to compute (and the same is true for $d_{E(PS)}$ and $d_{E(IPS)}$), as one would need to consider the set of all full rankings consistent with partial one. As efficient heuristics to compute $d(\pi, \sigma)$, where one or both between π and σ are partial, we can consider, following [Kamishima and Akaho, 2009], to compute the classic Spearman distance d_S on the full rankings obtained by considering only items in $\mathcal{D}(\pi) \cap \mathcal{D}(\sigma)$ (we call this *reduced Spearman*) or to substitute, when dealing with partial rankings, ranks with expected ranks in Eq. 3 (*Spearman ER*). Finally we consider an approximation based on sampling a set of consistent full ranking (*Sampling Spearman*).

6 Application to Clustering

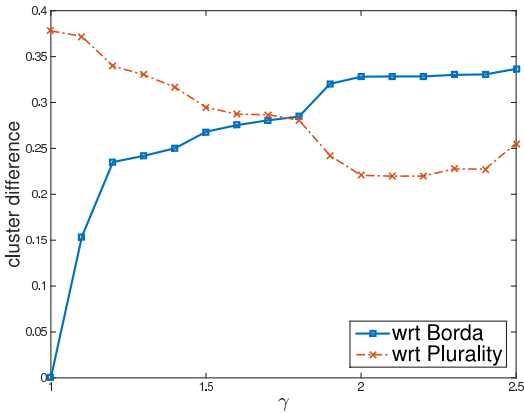


Figure 1: Comparison between clusters obtained with various scoring rules and those obtained with Borda and Plurality (γ controls the steepness of the weights used by positional Spearman; 5 clusters).

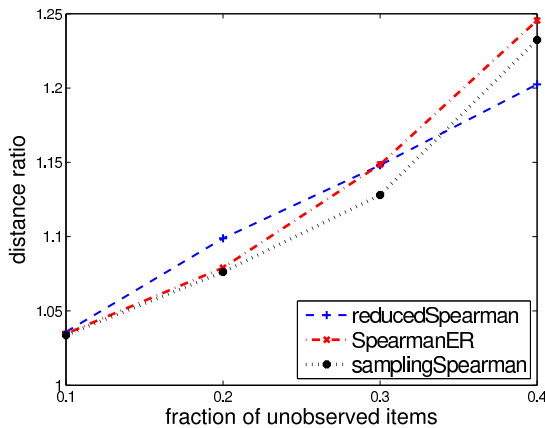


Figure 2: Performance of distance-based clustering with partial rankings (3 clusters; 10 runs).

We perform a number of experiments in order to illustrate the use of distance-based clustering in a variety of settings. Algorithm 1 is performed using a scoring rule as aggregation method; the appropriate distance is used to assign rankings to clusters. In the experiments below we consider the *sushi dataset*⁷; 5000 users have been asked to rank a set of items (sushis) [Kamishima *et al.*, 2005] from the most to the least preferred. Note that the number of clusters is viewed as an input parameter by the clustering algorithm. We choose different values in different experiments for illustration purposes (in each experiment, all methods are given the same input). The problem of assessing the best number of clusters is a different and orthogonal problem.

Full rank data. The weight vector w can have a large impact on the clustering obtained using d_{PS} . We consider convex sequences as weights w such that the difference between consecutive weights is a geometric series; i.e. $w_i - w_{i+1} = \gamma \cdot (w_{i+1} - w_{i+2})$ where factor γ control the steepness of the sequence; for $\gamma = 1$ we have Borda. Figure 1 shows the sensitivity of clustering to the weight w used by positional Spearman, for different values of γ (with 5 clusters): we evaluate the difference of the clustering obtained with that would be obtained with Borda (straight line) and Plurality (dotted line); the difference between clusterings is measured as one minus the Rand index. When γ increases the clusters becomes increasingly similar to that obtained by Plurality and less to Borda ($\gamma \approx 1.8$ is a sort of “middle ground”). It is also interesting to note that, at least for this dataset, “steeper” scoring rules tends to have a larger fraction of the population concentrated in one single cluster: at the extreme, when using plurality and d_{PL} around 50% of the rankings are assigned to the same cluster; with Borda the fraction is only about 27%.

Incomplete rank data. In the second experiment, we first perform clustering with Borda/Spearman, and this is considered as reference (our “ground truth”). We then randomly remove a number of items from each ranking (obtaining partial rankings), and perform clustering using expected Borda and (as the exact probabilistic method is not feasible) three distances on partial rankings (reduced Spearman, spearman with Expected Ranks, and sampling Spearman). As evaluation measure, we compute the ratio between the distance of each point to its assigned centroid *calculated with Spearman using the full rankings* and that of the reference. We observed that the heuristic methods works almost as well as the more demanding approach based on sampling (Figure 2).

Recommendation under uncertain availability. We consider the following problem: we need to assign users to $k = 5$ clusters; each item (sushi) has a constant i.i.d. probability of being available (0.5 in our tests). Each user’s preferences are encoded by her ranking; the “utility” is represented by a scoring rule⁸ (assumed the same for all users). In each cen-

⁷ Available at <http://www.kamishima.net/sushi/>.

⁸ In this test, w is a geometric sequence: $w_r = 10 \cdot a^{r-1}$ (r being the position in the ranking; $a = 0.5$).

triod the top item among those available is recommended to each user in the cluster. We need to assign users to clusters before knowing which items will be available. Clustering with the combination scoring rule and positional Spearman (both using w) achieves the highest value of per-user-averaged expected social welfare (5.47), scoring rule w together with classic Spearman yields 4.41 and Borda/Spearman 4.40; aggregation without clustering would yield a dismal 3.58.

We now modify the setting associating each item with a cost, so that the utility a recommendation i to a user is $w(\sigma(i))$ (where σ is the user's ranking) minus its cost. We perform clustering on a number of instances, with randomly generated costs associated to items (results are averaged over several runs). In this setting, with 2 clusters, the biased scoring rule (with weights w for positions, multiplicative biases set to zero, $z_i = 0$, and the additive biases ϕ set to the cost with negative sign), together with its characterizing distance d_{IPS} achieves 2.38 of expected social welfare; biased scoring rule with d_{PS} achieves 2.31, the unbiased scoring rule (using w) yields 2.20 and aggregation without clustering 1.3.

Computation time Clustering with distance measures associated to scoring rules is particularly efficient. By exploiting our characterization of scoring rules, the aggregation step (computing the centroid of a cluster) becomes very efficient: running time is $O(n \log n)$ with respect to the number of items (we need to count the scores associated to each item and then to sort them). In our tests, clustering using plurality was fastest (less than 1 second) and biased scoring rules are the slowest (convergence obtained in around 3s with 2 clusters and 8s for 5); a case aside is Sampling Spearman that is approximately 10 times slower⁹.

7 Conclusion

We provided a taxonomy of distance measures that characterize scoring rules as aggregation method. We extended the result about the connection between Borda rule and minimization of Spearman distances to scoring rules; we also consider the case of plurality, veto and top-k. We introduced a new family of aggregation rules, called biased scoring rules, giving an advantage to specific items, and show how they can be characterized. We provided experimental tests showing how these distance measures can be applied in clustering.

Axiomatic treatment of the median ranking from a point of view of social choice is given in [Barthélemy and Monjardet, 1981]. The idea of looking aggregation techniques in terms of minimization of distances is known as *distance rationalizability* in social choice theory [Elkind *et al.*, 2009]. The difference is that here we are interested in aggregations that produce a ranking, while in social choice the aggregation produces one or more winners (the elected candidates).

Because of its connection with the Condorcet property, Kemeny aggregation (minimization of Kendall tau) is often advocated. Since Kemeny aggregation is NP-hard problem, several researchers proposed heuristic methods, including the

Markov chains of [Dwork *et al.*, 2001]. Other researchers also considered extending common distances in some ways. In [Sculley, 2007] methods of ranking aggregation are extended in order to exploit similarity information between ranked items. The generalized distance functions presented in [Kumar and Vassilvitskii, 2010] are a rich generalization of footrule and Kendall whose expressivity is similar to the distances proposed here. Here, we focus on Spearman distances, because of the connection with scoring rules, clustering is very efficient to compute; this contrasts to Kemeny aggregation that can be computed exactly only in trivial cases.

References

- [Barthélemy and Monjardet, 1981] Jean-Pierre Barthélemy and Bernard Monjardet. The median procedure in cluster analysis and social choice theory. *Mathematical Social Sciences*, 1(3):235–267, 1981.
- [Critchlow *et al.*, 1991] Douglas E Critchlow, Michael A Fligner, and Joseph S Verducci. Probability models on rankings. *Journal of Mathematical Psychology*, 35(3):294–318, 1991.
- [Dwork *et al.*, 2001] Cynthia Dwork, Ravi Kumar, Moni Naor, and D. Sivakumar. Rank aggregation methods for the web. In Vincent Y. Shen, Nobuo Saito, Michael R. Lyu, and Mary Ellen Zurko, editors, *WWW*, pages 613–622. ACM, 2001.
- [Elkind *et al.*, 2009] Edith Elkind, Piotr Faliszewski, and Arkadii M. Slinko. On distance rationalizability of some voting rules. In Aviad Heifetz, editor, *TARK*, pages 108–117, 2009.
- [Kamishima and Akaho, 2009] Toshihiro Kamishima and Shotaro Akaho. Efficient clustering for orders. In Djamel A. Zighed, Shusaku Tsumoto, Zbigniew W. Ras, and Hakim Hacid, editors, *Mining Complex Data*, volume 165 of *Studies in Computational Intelligence*, pages 261–279. Springer, 2009.
- [Kamishima *et al.*, 2005] Toshihiro Kamishima, Hideto Kazawa, and Shotaro Akaho. Supervised ordering - an empirical survey. In *ICDM*, pages 673–676. IEEE Computer Society, 2005.
- [Kumar and Vassilvitskii, 2010] Ravi Kumar and Sergei Vassilvitskii. Generalized distances between rankings. In Michael Rappa, Paul Jones, Juliana Freire, and Soumen Chakrabarti, editors, *WWW*, pages 571–580. ACM, 2010.
- [Lu and Boutilier, 2011] Tyler Lu and Craig Boutilier. Learning Mallows Models with Pairwise Preferences. In Lise Getoor and Tobias Scheffer, editors, *ICML*, pages 145–152. Omnipress, 2011.
- [Marden, 1996] John I Marden. *Analyzing and modeling rank data*. CRC Press, 1996.
- [Sculley, 2007] D. Sculley. Rank Aggregation for Similar Items. In *SDM*. SIAM, 2007.

⁹Unoptimized implementation in MATLAB with off-the-shelf computing equipment.