Online Supplement for “Axiomatic Aggregation of Incomplete Rankings”

Erick Moreno-Centeno, Adolfo R Escobedo

Department of Industrial and Systems Engineering, Texas A&M University

emc@tamu.edu

adolfoescobedo@tamu.edu

A Further Discussion of Alternative Axiomatic Distances

A.1 Bogart’s (1973) Distance

Example: Consider the universe of objects $V = \{1, 2, 3, 4\}$, and the incomplete rankings $\mathbf{a} = (2, \bullet, 2, 1)$, $\mathbf{b} = (2, \bullet, 3, 1)$, $\mathbf{c} = (3, \bullet, 2, 1)$ and $\mathbf{d} = (\bullet, 2, \bullet, 1)$. To calculate Bogart’s distances, $P^{\mathbf{a}} = \{(1, 1), (2, 2), (3, 3), (4, 4), (1, 3), (3, 1), (4, 1), (4, 3)\}$ is the partial order representation of $\mathbf{a}$, for example. Bogart’s distances between the given rankings are $d_B(\mathbf{a}, \mathbf{b}) = 1$, $d_B(\mathbf{a}, \mathbf{c}) = 1$, $d_B(\mathbf{a}, \mathbf{d}) = 5$, $d_B(\mathbf{b}, \mathbf{c}) = 2$, $d_B(\mathbf{b}, \mathbf{d}) = 4$, and $d_B(\mathbf{c}, \mathbf{d}) = 4$.

Notice $d_B(\mathbf{b}, \mathbf{d}) = d_B(\mathbf{c}, \mathbf{d}) < d_B(\mathbf{a}, \mathbf{d})$. Therefore, Bogart’s distance indicates a higher level of agreement between $\mathbf{b}$ (alternatively, $\mathbf{c}$) and $\mathbf{d}$ than between $\mathbf{a}$ and $\mathbf{d}$. We believe this is incorrect. In particular, the only difference between $\mathbf{b}$ (alternatively, $\mathbf{c}$) and $\mathbf{a}$ is $\mathbf{a}$ ties objects 1 and 3, while $\mathbf{b}$ prefers 1 over 3 (alternatively, $\mathbf{c}$ prefers 3 over 1); on the other hand $\mathbf{d}$ does not rank objects 1 and 3. Hence, we believe the level of agreement between $\mathbf{b}$ (alternatively, $\mathbf{c}$) and $\mathbf{d}$ should be equal to the level of agreement between $\mathbf{a}$ and $\mathbf{d}$.

In general, the behavior of Bogart’s distance, can be generalized as follows: Suppose a given incomplete ranking, $\mathbf{d}$, does not compare objects i and j . The incidence matrix $I(\mathbf{d})$ will have $I(\mathbf{d})_{ij} = 0$ and $I(\mathbf{d})_{ji} = 0$. The incidence matrix $I(\mathbf{a})$ of some arbitrary ranking $\mathbf{a}$ tying objects i and j will have $I(\mathbf{a})_{ij} = 1$ and $I(\mathbf{a})_{ji} = 1$. The incidence matrix $I(\mathbf{b})$ of some arbitrary ranking $\mathbf{b}$ preferring i over j will have $I(\mathbf{b})_{ij} = 1$ and $I(\mathbf{b})_{ji} = 0$. Finally, the incidence matrix $I(\mathbf{c})$ of some arbitrary ranking $\mathbf{c}$ preferring j over i will have $I(\mathbf{c})_{ij} = 1$ and $I(\mathbf{c})_{ji} = 0$. Therefore, with respect

to the pair of objects i and j , ranking $\mathbf{b}$ (alternatively, $\mathbf{c}$) and ranking $\mathbf{d}$ are closer than ranking $\mathbf{a}$ and ranking $\mathbf{d}$.

A.2 Cook et al.’s (1986) Distance

Example: Consider the universe of objects $V = \{1, 2, 3, 4\}$ and the incomplete rankings $\mathbf{a} = (2, \bullet, 2, 1)$, $\mathbf{b} = (2, \bullet, 3, 1)$, $\mathbf{c} = (3, \bullet, 2, 1)$ and $\mathbf{d} = (\bullet, 2, \bullet, 1)$. Cook et al.’s distance between the each pair of the above rankings are $d_B(\mathbf{a}, \mathbf{b}) = 1$, $d_B(\mathbf{a}, \mathbf{c}) = 1$, $d_B(\mathbf{a}, \mathbf{d}) = 8.5$, $d_B(\mathbf{b}, \mathbf{c}) = 2$, $d_B(\mathbf{b}, \mathbf{d}) = 9.5$, and $d_B(\mathbf{c}, \mathbf{d}) = 9.5$.

Notice $d_B(\mathbf{b}, \mathbf{d}) = d_B(\mathbf{c}, \mathbf{d}) > d_B(\mathbf{a}, \mathbf{d})$ (This is the opposite situation than with Bogart’s (1973) distance, as seen in §A.1). Therefore, using Cook et al.’s distance, one might conclude there is a lower level of agreement between $\mathbf{b}$ (alternatively, $\mathbf{c}$) and $\mathbf{d}$ than between $\mathbf{a}$ and $\mathbf{d}$. We believe this is incorrect. In particular, the only difference between $\mathbf{b}$ (alternatively, $\mathbf{c}$) and $\mathbf{a}$ is $\mathbf{a}$ ties objects 1 and 3, while $\mathbf{b}$ prefers 1 over 3 (alternatively, $\mathbf{c}$ prefers 3 over 1). Hence, we believe the level of agreement between $\mathbf{b}$ (alternatively, $\mathbf{c}$) and $\mathbf{d}$ should be equal to the level of agreement between $\mathbf{a}$ and $\mathbf{d}$.

In general, the behavior of Cook et al.’s distance can be generalized similarly to how the behavior of Bogart’s distance was generalized in Section A.1. The difference is the median ranking obtained by using Cook et al.’s distance will tend to be a ranking containing an artificially high amount of ties.

B Review of Restrictive and Unrelated Ranking Aggregation Methods

B.1 Top- k List Aggregation

Within the field of machine learning, there has been extensive research on the top- k list (ranking) aggregation problem: given $m = |V|$ objects and a set of evaluations listing the top- k ranked choices of these objects, where typically $k \ll m$, find a ranking that best agrees with the input rankings - see, for instance, (Fagin et al. 2003) and (Ailon 2010). Whereas many such methods are distance-based, they treat incomplete rankings differently compared to the current work. In

top- k rankings, the $m - k$ objects not included in an individual ranking are designated as tied for position $k + 1$, that is, they are all indifferently worse. The proposed method, on the other hand, does not make any assumptions about the unranked objects in an individual ranking. In fact, two main reasons it allows incomplete rankings as inputs is to prevent judges from gaming the consensus ranking and to limit the noise associated with over-ranking fatigue. Moreover, the input evaluations of the proposed axiomatic distance may be rankings of distinct subsets of V , with possibly different sizes, by design. In top- k ranking aggregation schemes it may happen that two top- k lists have no elements in common, but this is a rare occurrence in most cases; by definition, top- k lists must be of equal size (Klementiev et al. 2008). Therefore, based on their treatment of incomplete rankings, top- k ranking aggregation approaches attempt to solve a more specific instance of the ranking aggregation problem, namely one in which a judge’s incomplete evaluation implies her extended evaluation of the entire object set. Having said this, however, because of its flexibility and broader scope, the axiomatic distance-based method discussed in the accompanying manuscript could conceivably be applied to solve the top- k ranking problem as well.

B.2 Sequence Mining

Data mining techniques have also been applied to ranking aggregation. More specifically, the goal in what are known as sequence mining methods is to extract the largest ordered subsequence (i.e., ranking) from a group of ordered sequences of maximal size $m = |V|$ such that a majority of the input sequences agrees with the subsequence and only a minority disagrees with it - two sample papers on this topic are (Chen and Cheng 2009, 2010). While some of these methods allow incomplete rankings as inputs and do not make assumptions about unranked items in individual evaluations, they typically produce incomplete aggregate rankings. This is because whenever judge groupings with enough members rank an object subset $V' \subset V$ in contradicting order, the objects in V' are left to be renegotiated. Furthermore, the aggregate ranking obtained via these techniques is by design the most popular "sub-ranking" and, as such, does not lead to solutions that represent the best compromise between all evaluations. The proposed method, in contrast, provides a complete aggregate ranking, defined as the median evaluation in Section 3.1 of the accompanying manuscript, that represents the best middle ground mathematically between all the input evaluations. Due to these distinctions, it is evident that sequence mining methods and the proposed axiomatic method

solve different problems. In particular, unlike sequence mining methods, the proposed method performs only one aggregation, and it does not require judges to switch their preference relations to achieve a complete final ranking; any necessary compromises are implicit in the mathematical characterization of a consensus ranking.

B.3 Consensus Clustering

The consensus clustering problem can be stated as follows: given a number of clusterings, find the single clustering that agrees the most with the input clusterings - see, for instance, (Ailon et al. 2008) and (Filkov and Skiena 2004). The key distinction between consensus clustering and ranking aggregation lies in how each defines disagreement. In consensus clustering, disagreement between two clusterings $\mathcal{C}$ and $\mathcal{C}'$ arises when a pair of objects (u, v) in the same cluster $C \in \mathcal{C}$ are in separate clusters $C', C'' \in \mathcal{C}'$ (Gionis et al. 2007). There is no ordering relationship between clusters and, thus, the same disagreement between $\mathcal{C}$ and $\mathcal{C}'$ would result if $u \in C'$ and $v \in C'''$, where $C' \neq C'' \neq C'''$ and $C''' \in \mathcal{C}'$. On the other hand, depending on the distance function used, the disagreement between rankings $\mathbf{a}$ and $\mathbf{b}$ is measured according to the difference in magnitude in the positions occupied by an object $u \in V$ in $\mathbf{a}$ and $\mathbf{b}$ or in the relative relationship between u and another object $v \in V$ in $\mathbf{a}$ and $\mathbf{b}$ (i.e., the rankings do not agree when u is preferred over v in $\mathbf{a}$, but u is not preferred over v in $\mathbf{b}$). Therefore, ranking aggregation and consensus clustering are significantly different problems.

C Borda Count Variation Definitions

In this section, we define the three different scoring functions whose corresponding aggregate rankings are evaluated in Comparison 4 of Section 5.4 of the accompanying manuscript. For this purpose, let $U_{\mathbf{a}} \subset V$ be the subset of unranked objects in ranking vector $\mathbf{a}$; let $\pi_{\mathbf{a}}(i)$ be the ranking assigned to object i in $\mathbf{a}$; let $\pi_{\mathbf{a}}^{\max}$ and $\pi_{\mathbf{a}}^{\min}$ be the maximum and minimum rankings specified in $\mathbf{a}$, respectively; lastly, let $\sigma_{\mathbf{a}}^j(i)$ be the score assigned to object i in $\mathbf{a}$ by Borda Count variation j , where $1 \leq i \leq |V| = m$ and $1 \leq j \leq 3$.

The scoring functions of the featured Borda Count variations score each object in an individual ranking as follows: σ^1 assigns a typical Borda score to ranked items (i.e. count the number of

ranked items with equal or lower rank than the current object, including the object itself) and a null score to unranked items; σ^2 assigns a slightly modified Borda score to ranked items (i.e. increase the typical Borda score by the number of unranked items) and it assigns the midpoint of the missing Borda scores (i.e., if the individual ranking were complete) to unranked items; σ^3 assigns the best and worst possible typical Borda scores of a complete ranking of size m to the highest and lowest ranked items in the individual ranking, respectively, it gives corresponding scaled scores to remaining ranked objects, and it assigns the midpoint Borda score of a complete without-ties ranking of size m to unranked objects. The mathematical definition of the scoring functions associated with three Borda Count variations is as follows:

$$\sigma_a^1(i) = \begin{cases} 0 & \text{if } i \in U_a \\ m - |U| - \pi(i) + 1 & \text{otherwise} \end{cases} \quad \sigma_a^2(i) = \begin{cases} \frac{|U|+1}{2} & \text{if } i \in U_a \\ m - \pi(i) + 1 & \text{otherwise} \end{cases}$$

$$\sigma_a^3(i) = \begin{cases} \frac{1}{2}(m+1) & \text{if } \pi_a^{\min} = \pi_a^{\max} \text{ or } i \in U_a \\ m & \text{if } \pi(i) = \pi_a^{\min} \\ 1 & \text{if } \pi(i) = \pi_a^{\max} \\ m - \frac{(\pi(i)-1)(m-1)}{m-|U|-1} & \text{otherwise} \end{cases}$$

Each of the above scoring functions ultimately implies an assumption regarding the preference of unranked objects in an individual ranking: σ^1 and σ^2 assume they rank the worst and σ^3 assumes they rank in the middle. Similarly, all ranking aggregation approaches that assign scores to unranked items imply (intentionally or otherwise) certain assumptions about these objects based on the score they assign them.

References

- Ailon, Nir. 2010. Aggregation of partial rankings, p-ratings and top-m lists. *Algorithmica* **57**(2) 284–300.
- Ailon, Nir, Moses Charikar, Alantha Newman. 2008. Aggregating inconsistent information: ranking and clustering. *Journal of the ACM (JACM)* **55**(5) 23.
- Chen, Yen-Liang, Li-Chen Cheng. 2009. Mining maximum consensus sequences from group ranking data. *European Journal of Operational Research* **198**(1) 241–251.
- Chen, Yen-Liang, Li-Chen Cheng. 2010. An approach to group ranking decisions in a dynamic environment. *Decision support systems* **48**(4) 622–634.

- Cook, W. D., M. Kress, L. M. Seiford. 1986. Information and preference in partial orders: a bimatrix representation. *Psychometrika* **51** 197–207.
- Fagin, Ronald, Ravi Kumar, D Sivakumar. 2003. Comparing top k lists. *SIAM Journal on Discrete Mathematics* **17**(1) 134–160.
- Filkov, Vladimir, Steven Skiena. 2004. Integrating microarray data by consensus clustering. *International Journal on Artificial Intelligence Tools* **13**(04) 863–880.
- Gionis, Aristides, Heikki Mannila, Panayiotis Tsaparas. 2007. Clustering aggregation. *ACM Transactions on Knowledge Discovery from Data (TKDD)* **1**(1) 4.
- Klementiev, Alexandre, Dan Roth, Kevin Small. 2008. Unsupervised rank aggregation with distance-based models. *Proceedings of the 25th international conference on Machine learning*. ACM, 472–479.