Minimising the Rank Aggregation Error

(Extended Abstract)

Mathijs M. de Weerdt
Delft University of Technology, Netherlands
m.m.deweerdt@tudelft.nl

Enrico H. Gerding
University of Southampton, UK
eg@ecs.soton.ac.uk

Sebastian Stein
University of Southampton, UK
ss2@ecs.soton.ac.uk

ABSTRACT

Rank aggregation is the problem of generating an overall ranking from a set of individual votes which is as close as possible to the (unknown) correct ranking. The challenge is that votes are often both noisy and incomplete. Existing work focuses on the most likely ranking for a particular noise model. Instead, we focus on minimising the error, i.e., the expected distance between the aggregated ranking and the correct one. We show that this results in different rankings, and we show how to compute local improvements of rankings to reduce the error. Extensive experiments on both synthetic data based on Mallows' model and real data show that Copeland has a smaller error than the Kemeny rule, while the latter is the maximum likelihood estimator.

Keywords

Economic paradigms: Social Choice Theory

1. INTRODUCTION

Marquis de Condorcet [6] said that voting or rank aggregation may be regarded as a way of uncovering the ground truth. There are many practical examples of rank aggregation, including websites that produce rankings of restaurants, books and movies based on crowdsourced contributions from their users, and scientific communities that use votes from their members to select which project proposals to fund or which papers to accept [2]. In these settings votes are not only noisy, but also incomplete since typically only a subset of the alternatives is ranked by any single individual.

To find a ranking close to the ground truth, most current work assumes a probabilistic noise model such as Mallows [4, 7], and then aims to find the aggregate ranking with the highest likelihood of being the true ranking. For Mallows' model, it has been shown that the Kemeny rule is the maximum likelihood estimator (MLE) [7]. However, in most settings the aim should be to find a ranking that gives the best results when used in subsequent decision making. When votes are noisy and incomplete, many rankings may have a likelihood of similar magnitude, and the probability of any of these being the true ranking is small. In addition, the likelihood does not account for the distance to the true ranking when the chosen ranking is not the true one. Therefore, we argue that, instead, the aim should be to minimise the *expected distance* of an aggregate ranking to the true ranking, which we term the *error*. In contrast to MLEs, to date, it is still unknown what voting rules perform best regarding this objective.

Appears in: *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016), J. Thangarajah, K. Tuyls, C. Jonker, S. Marsella (eds.), May 9–13, 2016, Singapore.*

2. MODEL AND THEORETICAL RESULTS

Let $A = \{1, 2, \ldots, m\}$ denote a set of alternatives and $N = \{1, 2, \ldots, n\}$ the set of agents. We define a vote by agent k as a linear order over an $A_k \subseteq A$, denoted by $\sigma_k : A_k \to \{1, 2, \ldots, |A_k|\}$. Here, $\sigma_k(i)$ defines the rank of alternative i (lower is better). We also use $i \succ_{\sigma_k} j$ to denote $\sigma_k(i) < \sigma_k(j)$. Furthermore, we use D to denote the set of all votes.

The Kendall-tau distance K counts the pairs of alternatives that are differently ordered by σ than by τ.

$$K(\sigma, \tau) = |\{\{i, j\} \subseteq A : i \succ_\sigma j \text{ and } i \prec_\tau j\}| \qquad (1)$$

We assume noise according to the well-known Mallows' model for a probability $p > 0.5$. Given this model, the likelihood of observing the votes D given that the true ranking is τ is given by:

$$\mathcal{L}(\tau|D) = \frac{1}{Z_1} \prod_{\{i,j\} \subseteq A : i \succ_\tau j} p^{n_a(i,j|D)} (1-p)^{n_d(i,j|D)}, \qquad (2)$$

where Z_1 is a (normalisation) constant. It has been shown that the Kemeny rule — which minimises the sum of Kendall-tau distances to all votes — is an MLE [7]. We extend this model to incomplete rankings by introducing a probability q of missing an alternative:

$$\mathcal{L}'(\tau|D) = \mathcal{L}(\tau|D) \frac{1}{Z_2} \prod_{k \in N} (1-q)^{|\sigma_k|} q^{m-|\sigma_k|}. \qquad (3)$$

It is easy to see that Kemeny optimal still maximises the likelihood, irrespective of the value of q.

Given a set of all rankings T, we define the *rank aggregation error* as the *expected Kendall-tau distance*, formally:

$$\text{KT-error}(\tau) = \sum_{\tau' \in T} K(\tau, \tau') \cdot \mathcal{L}'(\tau'|D). \qquad (4)$$

The following example shows that minimising this error produces a single natural ranking, whereas multiple rankings maximise the likelihood.

EXAMPLE 1. *Let three alternatives a, b, c be given, and one agent with vote $a \succ_{\sigma_1} b$. The Kemeny rule is indifferent between the three possible aggregate rankings without noise. The expected Kendall-tau distances, however, differ:*

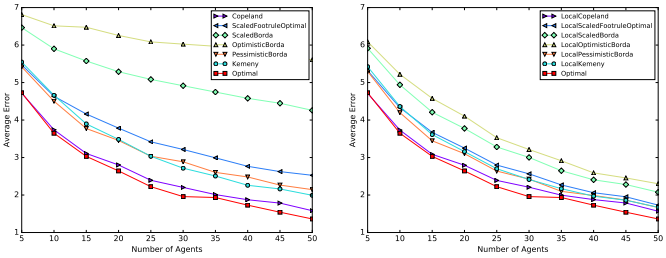

Figure 1: Results for synthetic data with $m = 6$, $p = \frac{2}{3}$ and $q = 0.7$. The right graph shows results after applying local Kemenisation (same for Figure 2).

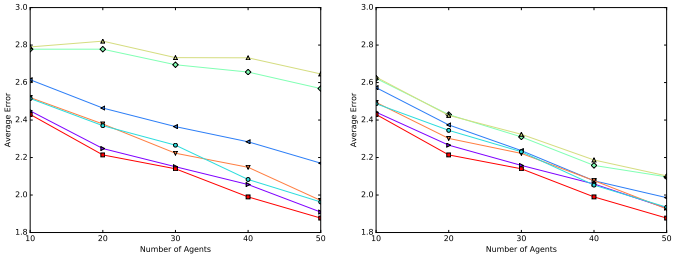

Figure 2: Results using the Dots dataset and $q = 0.7$

τ_i	*ranking*	$\mathcal{L}(\tau_i\|\{\sigma_1\})$	$K(\tau_i,\tau_j)$			*KT-error*(τ_i)
τ_1	$a \succ b \succ c$	$\frac{1}{3}$	*0*	*1*	*2*	*1*
τ_2	$a \succ c \succ b$	$\frac{1}{3}$	*1*	*0*	*1*	$\frac{2}{3}$
τ_3	$c \succ a \succ b$	$\frac{1}{3}$	*2*	*1*	*0*	*1*

Because the distance of $a \succ c \succ b$ to each of the other rankings is only 1, it has a lower expected error.

Similarly, we can construct examples with complete votes where the rankings maximising the likelihood do not minimise the KT-error.

Hardness. Finding the ranking with the largest likelihood (i.e., Kemeny) is NP-complete [3]. Towards establishing the computational complexity of finding an aggregate ranking with the minimum error, we can show that the problem of computing the error of a (single) aggregate ranking is #P-hard (even) when there is no noise in the data D.

THEOREM 1. *Even without noise, determining the expected error of an aggregate ranking π is #P-hard.*

The proof is by reduction from computing the number of linear extensions of a partial order, which is #P-complete [1].

Local Search. Although computing the error does not seem to be feasible, we can improve a ranking by making local adjustments. In particular, given a ranking τ, it is easy to determine if, by swapping two adjacent alternatives, we can improve the KT-error.

THEOREM 2. *Let τ_{ab} and τ_{ba} be two equal rankings except that two adjacent alternatives, a and b, are swapped. That is, $a \succ_{\tau_{ab}} b$ and $b \succ_{\tau_{ba}} a$. Then: KT-error$(\tau_{ab}) <$ KT-error(τ_{ba}) iff $n_a(a,b|D) > n_a(b,a|D)$.*

It turns out that repeatedly applying this rule until a local optimum is found has been termed local Kemenisation [2]. Our result adds that this is locally minimising the KT-error as well. Note that, although the proof assumes Mallows' model, it seems intuitive for *any* noise model to swap adjacent alternatives if one is ranked more often above the other.

3. EXPERIMENTAL RESULTS

We consider experiments using synthetic data, and real data from the Dots experiments (PrefLib library [5]). In Figures 1 and 2 we compare extensions of Spearman's Footrule (Scaled Footrule Optimal) and Borda for incomplete votes [2], Copeland, the Kemeny rule and an Optimal rule that minimises the KT-error assuming Mallows' model. As expected, having more agents decreases the average distance to the true ranking for all rules. We also observe that the Kemeny rule is indeed not optimal, with around 0.5 inversions more than Optimal on average. Interestingly, Copeland performs significantly better than Kemeny, and at times on par with Optimal. Even more striking is the significant improvement of most rules by local Kemenisation. Surprisingly, similar results hold for the Dots data set, despite the fact that this is not generated using Mallow's model.

4. CONCLUSIONS

We have shown that voting rules which maximise the likelihood of a ranking do not necessarily minimise the expected distance to the true ranking. Specifically, for rank aggregation with noise and missing votes, maximising the likelihood can result in a significantly higher error than computationally simpler methods such as Copeland. This discrepancy can occur even when votes are complete. Furthermore, we have shown that Optimal performs best in both synthetic and real data settings, even when we do not know the noise parameter exactly. For Mallows' model we have shown that computing this error is hard. Furthermore, we proved that an efficient procedure called local Kemenisation, which is known to improve the likelihood, also reduces the error, and that this leads to a significant performance improvement.

Acknowledgments. This work has been partly supported by COST Action IC1205 on Computational Social Choice.

REFERENCES

[1] G. Brightwell and P. Winkler. Counting linear extensions is #P-complete. In *23rd ACM Symposium on Theory of Computation*, pages 175–181, 1991.

[2] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation methods for the web. In *10th Int. Conf. on World Wide Web*, pages 613–622. ACM, 2001.

[3] E. Hemaspaandra, H. Spakowski, and J. Vogel. The complexity of Kemeny elections. *Theoretical Computer Science*, 349(3):382–391, 2005.

[4] C. L. Mallows. Non-null ranking models. I. *Biometrika*, pages 114–130, 1957.

[5] A. Mao, A. Procaccia, and Y. Chen. Better Human Computation Through Principled Voting. In *AAAI*, 2013.

[6] M. Nicolas de Condorcet. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. L'imprimerie royale, 1785.

[7] P. Young. Optimal voting rules. *The Journal of Economic Perspectives*, 9(1):51–64, 1995.