INTRODUCTION

We study the problem of ranking aggregation: agents provide a collection of ranked preferences over a set of alternatives, we wish to aggregate them into one consensus ranking.

One popular approach in ranking aggregation follows Kemeny’s rule. Given a collection of rankings/permutations \( D_N := (\sigma_1, \ldots, \sigma_N) \) over \( n \) alternatives, Kemeny consensus(es) are the solution to

\[
K_N := \arg \min_{\sigma \in \Sigma_n} \sum_{i=1}^{N} d(\sigma, \sigma_i),
\]

where \( d \) is the Kendall’s tau distance between permutations, i.e.,

\[
d(\sigma, \sigma') = \sum_{1 \leq i < j \leq n} I((\sigma(j) - \sigma(i))(\sigma'(j) - \sigma'(i)) < 0).
\]

Kemeny consensuses satisfy many desirable properties but are NP-hard to compute even for \( n = 4 \). It thus calls for study on apprehending the complexity of Kemeny aggregation and theoretical guarantees of approximation procedures commonly used in practice.

TWO-MINUTE SUMMARY

The problem: Let \( D_N \in \Sigma_n^N \) be a dataset and \( \sigma \in \Sigma_n \) be a permutation, typically output by a computationally efficient aggregation procedure on \( D_N \). Can we give a (tractable) upper bound for the distance \( d(\sigma, \sigma^*) \) between \( \sigma \) and a Kemeny consensus \( \sigma^* \in K_n \)?

The bound: Denote by \( \theta_N(\sigma) \leq \pi \) the angle between the Kemeny embeddings \( \phi(\sigma) \) and \( \phi(D_N) \) in an Euclidean space. If \( \theta_N(\sigma) < \frac{\pi}{2} \), then for all \( \sigma^* \in K_N \),

\[
d(\sigma, \sigma^*) \leq \left( \frac{3}{2} \right) \sin^2(\theta_N(\sigma)) =: k_{\min}(\sigma; D_N),
\]

where \( \lfloor x \rfloor \) denotes the integer part of the real \( x \).

The distinguishing merits:
- \( k_{\min} \) is simple to code and efficient to compute in time \( O(Nn^2) \).
- Generality with no assumption on the dataset \( D_N \) or the aggregation procedure giving \( \sigma \).
- Depends on geometric understanding of the combinatorial problem of Kemeny aggregation.

RESULTS ON THE SUSHI DATASET

<table>
<thead>
<tr>
<th>Voting rule</th>
<th>num(D_N(\sigma))</th>
<th>(k_{\min}(\sigma))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Borda</td>
<td>0.620</td>
<td>14</td>
</tr>
<tr>
<td>Copeland</td>
<td>0.622</td>
<td>14</td>
</tr>
<tr>
<td>QuickSort</td>
<td>0.622</td>
<td>14</td>
</tr>
<tr>
<td>Plackett-Luce</td>
<td>0.807</td>
<td>20</td>
</tr>
<tr>
<td>2-approval</td>
<td>0.748</td>
<td>20</td>
</tr>
<tr>
<td>1-approval</td>
<td>0.710</td>
<td>22</td>
</tr>
<tr>
<td>Pick-a-Perm</td>
<td>0.383\textsuperscript{a}</td>
<td>34.85\textsuperscript{a}</td>
</tr>
<tr>
<td>Pick-a-Random</td>
<td>0.377\textsuperscript{a}</td>
<td>35.90\textsuperscript{a}</td>
</tr>
</tbody>
</table>

Table 1: \( k_{\min} \) with different aggregation procedures on the original sushi dataset.

Figure 1: \( k_{\min} \) with different aggregation procedures on 500 bootstrapped pseudo-samples of the sushi dataset (\( N = 5000, n = 10 \)).

GEOMETRIC ANALYSIS OF KEMENY AGGREGATION

Kemeny embedding: The Kemeny embedding of a single permutation \( \sigma \in \Sigma_n \) is defined by:

\[
\phi : \Sigma_n \rightarrow \mathbb{R}^2 ; \sigma \mapsto \langle \text{sign}(\sigma(j) - \sigma(i)) \rangle_{1 \leq i < j \leq n},
\]

In particular, \( \|\phi(\sigma)\| = \frac{\sqrt{n(n-1)}}{2} \) for all \( \sigma \in \Sigma_n \), i.e., \( \phi(\sigma) \) lies on the sphere centered at origin with radius \( R := \frac{\sqrt{n(n-1)}}{2} \).

The Kemeny embedding induces a space where the squared Euclidean distance recovers the Kendall’s tau distance, i.e., for all \( \sigma, \sigma' \in \Sigma_n \),

\[
d(\sigma, \sigma') = \frac{1}{4} \|\phi(\sigma) - \phi(\sigma')\|^2.
\]

Kemeny aggregation in the embedded space: For any dataset \( D_N := (\sigma_1, \ldots, \sigma_N) \in \Sigma_n^N \), Kemeny aggregation is equivalent to solving:

\[
K_N = \arg \min_{\sigma \in \Sigma_n} \|\phi(\sigma) - \phi(D_N)\|^2 = \arg \min_{\sigma \in \Sigma_n} \theta_N(\sigma), \quad (1)
\]

where

\[
\phi(D_N) := \frac{1}{N} \sum_{i=1}^{N} \phi(\sigma_i)
\]

denotes the barycenter of the point cloud \( D_N \), and \( \theta_N(\sigma) \) denotes the Euclidean angle between the Kemeny embeddings \( \phi(\sigma) \) and \( \phi(D_N) \).

Kemeny aggregation decomposes in two steps:

1. Compute the mean embedding of the dataset \( \phi(D_N) \) in time \( O(Nn^2) \).
2. Find the consensus permutation \( \sigma^* \) that minimizes (1).

GEOMETRIC INTERPRETATION OF THE BOUND

Fix \( D_N \in \Sigma_n^N \) and let \( \sigma \in \Sigma_n \) be a permutation and \( \sigma^* \in K_n \) be a Kemeny consensus. Since \( \theta_N(\sigma^*) \leq \theta_N(\sigma) \leq \frac{\pi}{2} \), applying laws of cosines in the 2-dimensional subspace spanned by \( \phi(D_N) \) and \( \phi(\sigma) \) gives

\[
r^2 > 2R^2(1 - 2\cos(2\theta_N(\sigma))) = 4R^2\sin^2(\theta_N(\sigma)).
\]

The minimum integer of \( \frac{r^2}{4} \) satisfying the inequality recovers \( k_{\min} \) exactly.

Figure 3: Level sets of the cost function in (1) over the sphere for \( n = 3 \).

Figure 4: Geometric illustration of the bound taking \( \pi = \phi(\sigma); k = \frac{r^2}{4} \).

CONCLUSION

- **Methodological**: geometric properties of Kemeny aggregation.
- **Theoretical**: simple and tractable quantity controlling the distance to a Kemeny consensus without computing it.
- **Extensible**: ranking aggregation from partial orders, etc.

REFERENCES