Aggregating Preferences from (Ranked) Choices: A Simple Mallows-type Model

(*Authors' names blinded for peer review*)

Abstract (for a general session presentation)

We study a Mallows-type (i.e., distance-based) probabilistic ranking model to aggregate people's (ranked) choices. In the data, every participant provides a top-k ranked list of items (thereafter referred to as "ranked" choices) among a given display set . When $k = 1$, such lists are equivalent to "choices" commonly studied in choice modeling. The general ($k > 1$) setting is less explored, and largely motivated by applications in voting, preference learning, etc.

Every probabilistic ranking model aggregates into a ranked choice model in a "canonical" way: The probability of choosing a top-k ranked list from a display set equals the sum of the probabilities associated with all the rankings in which the list takes the top k positions among all the items in the display set. In this way, the ranked choice model is "rationalized" by a distribution over participants' preferences. A distance-based ranking model specifies a parsimonious distribution over rankings (in a similar spirit to the Gaussian distribution for scalars). It can be used as "kernels" to "smooth out" the sparse distributions over rankings estimated from data to improve the generalization power.

The main difficulty in applying Mallows-type models to choice modeling is that the corresponding (ranked) choice probabilities are difficult to obtain. We identify a novel distance-based probabilistic ranking model. It is similar to the Mallows model only with a twist in the distance function. The new function penalizes disagreements close to the top positions more than those at the bottom (unlike the Kendall-tau distance function, which penalizes all pairwise disagreements equally).

We show that our probabilistic ranking model aggregates into simple closed-form expressions for the ranked choice probabilities. This is a property that its counterparts do not possess even for $k = 1$. Furthermore, the new ranking model is relatively easy to estimate as the MLE problem reduces to a well-studied ranking-aggregation-type problem, which enjoys computational advantages. (For example, it admits a PTAS.) We also study the statistical guarantees (e.g., consistency) of the MLE solutions.

As a proof of concept, we compare our model with the two most representative probabilistic ranking models (the Plackett-Luce model and the Mallows model) when $k = 1$ on real-world data. We find our method achieves a better generalization power, especially when there is a limited variety of display sets.