Unsupervised Rank Aggregation with Distance-Based Models

Alexandre Klementiev, Dan Roth, and Kevin Small
University of Illinois at Urbana-Champaign
201 N Goodwin Ave
Urbana, IL 61801 USA
{klementi,danr,ksmall}@uiuc.edu

Suppose that each of a panel of judges independently generates a (partial) ranking over a set of objects, and assume that each judge tries to reproduce a true underlying ranking according to the degree of their expertise. This setting often arises in Information Retrieval (IR) and Natural Language Processing (NLP) among other areas, and one needs to meaningfully combine such expert opinions into an aggregate ranking. For example, in *meta-search*, the aim is to aggregate a set of Web search query results produced by multiple engines. In *textual entailment*, one may wish to combine a ranking over potentially entailed candidates produced by a number of heuristic or learned modules. In *machine translation*, combining outputs of multiple MT systems based on different principles may produce a better translation.

A supervised learning approach to solving the *rank aggregation* problem is likely to be of limited utility since supervised ranked data is often unavailable. We propose a general mathematical framework for *unsupervised* rank aggregation which can be used to *learn* to combine (partial) orderings over objects. We instantiate the framework for two scenarios: combining permutations and combining top-k lists, a partial order common to many IR and ranking tasks. We use the extended Mallows model formalism [3] and propose an EM-based learning algorithm for learning its parameters without supervision. We show that if a distance function used in the formalism satisfies a particular set of properties, learning of the model can be made efficient. In the top-k setting, we introduce a novel distance function which satisfies these properties.

Consider a scenario in which a panel of K judges generate (partial) rankings $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_K)$ over a set of objects. [3] propose a distance-based conditional model, a generalization of the standard Mallows model [4] to the case when multiple input rankings are available:

$$p(\pi|\boldsymbol{\theta}, \boldsymbol{\sigma}) = \frac{1}{Z(\boldsymbol{\theta}, \boldsymbol{\sigma})} p(\pi) \exp\left(\sum_{i=1}^{K} \theta_i \; d(\pi, \sigma_i)\right) \tag{1}$$

where $Z(\boldsymbol{\theta}, \boldsymbol{\sigma})$ is a normalizing constant, $d(\cdot, \cdot)$ is a right-invariant distance function [1], and free parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_K)$ may be thought of as the relative expertise of the judges.

Suppose further that we get Q instances of $\boldsymbol{\sigma}$; the set $\{\boldsymbol{\sigma}^{(j)}\}_{j=1}^{Q}$ constitutes the observed data.

We derive the EM-based algorithm to estimate the parameters of the model:

$$E_{\theta_i}(D) = \sum_{\substack{(\pi^{(1)},\ldots,\pi^{(Q)}) \\ \in \mathcal{S}_n^Q}} \left(\frac{1}{Q} \sum_{q=1}^{Q} d(\pi^{(q)}, \sigma_i^{(q)}) \right) \prod_{j=1}^{Q} p\left(\pi^{(j)} | \boldsymbol{\theta'}, \boldsymbol{\sigma}^{(j)}\right) \tag{2}$$

where $\mathcal{S}_n$ is the set of all possible (partial) rankings, $\boldsymbol{\theta'}$ are the values of the parameters from the previous iteration, and r.v. $D \stackrel{\text{def}}{=} d(e, \nu)$, where e is the identity and ν is a random permutation.

On every iteration of EM, we would need to estimate the right-hand side (RHS) of (2) and solve the LHS for θ_i for each of the K components. While both steps are expensive to solve in general, the LHS may be solved efficiently if $d(\cdot, \cdot)$ can be decomposed into a sum of independent components [2]. We use Metropolis sampling algorithm to estimate the RHS, and also propose a heuristic as an alternative.

We instantiate our model for the cases of combining permutations, and of combining top-k lists, and experimentally demonstrate the approach to be effective in both settings. We derive a novel decomposable metric for the latter case.

References

[1] Vladimir Estivill-Castro, Heikki Mannila, and Derick Wood. Right invariant metrics and measures of presortedness. *Discrete Applied Mathematics*, 42:1–16, 1993.

[2] M. A. Fligner and J. S. Verducci. Distance based ranking models. *Journal of the Royal Statistical Society*, 48:359–369, 1986.

[3] G. Lebanon and J. Lafferty. Conditional models on the ranking poset. In *In Proc. of NIPS*, volume 15, pages 431–438, 2003.

[4] C. L. Mallows. Non-null ranking models. *Biometrika*, 44:114–130, 1957.

Topic: Unsupervised Rank Aggregation
Preference: Oral