Consensus ranking under the exponential model

Marina Meilă, Arthur Patterson, Jeff Bilmes

University of Washington
Box 354322, Seattle, WA 98195-4322
mmp@stat.washington.edu

Assume that we are given a set of N rankings, a.k.a *linear orderings* on n objects. For instance, the rankings represent the individual preferences of a panel of N judges, each presented with the same set of n candidate objects. The problem of *rank aggregation* or of finding a *consensus ranking*, is formulated as finding a single ranking π_0 that best agrees with all the N rankings.[1]

Kendall's correlation [Fligner and Verducci, 1986] is a widely used models of agreement [Ailon et al., 2005, Lebanon and Lafferty, 2003, Cohen et al., 1999]. The Kendall distance is defined as

$$d_K(\pi, \pi_0) \;=\; \sum_{l \prec_\pi j} 1_{[j \prec_{\pi_0} l]} \tag{1}$$

In the above, π, π_0 represent permutations and $i \prec_\pi j$ ($i \prec_{\pi_0} j$) mean that l precedes j (i.e is preferred to j) in permutation π (π_0). Hence d_K is the total number of pairwise disagreements between π and π_0.

This distance was further generalized to a family of parametrized distances [Fligner and Verducci, 1986] by $d_\theta(\pi, \pi_0) \;=\; \sum_{j=1}^{n-1} \theta_j V_j(\pi\pi_0^{-1})$ where V_j are given functions on the ranking poset. A series of authors [Mallows, 1957, Fligner and Verducci, 1986, Lebanon and Lafferty, 2003] were concerned with probabilistic models of the form

$$P(\pi) \;\propto\; e^{-d_\theta(\pi, \pi_0)} \tag{2}$$

For fixed π_0, this is an exponential family model and estimating its parameters is straightforward. It is also easy to see that estimating π_0 in such a model is a way of reformulating the consensus clustering problem.

It is considered that the consensus clustering problem as well as the more general ML estimation of π_0 have no exact polynomial time algorithm [Ailon et al., 2005]. In [Ailon et al., 2005] a randomized algorithm that achieves an 11/7 approximation in minimizing the criterion (1) is presented. Greedy algorithms to minimize the same criterion were introduced by [Cohen et al., 1999].

This paper introduces a new, exact algorithm for the simultaneous Maximum Likelihood estimation of the centroid π_0 and parameters θ of such a model.

The running time of this algorithm is data dependent, being determined by the values of the θ parameters of the true data distribution. The running time is proportional to $n!$ in the worst case, which corresponds to a uniform distribution over permutations. But, if the distribution is concentrated around its centroid π_0 then the algorithm becomes tractable and for more concentrated distributions our algorithm becomes identical to a greedy algorithm (remaining all the time exact).

In the process of describing the algorithm, we also show that this problem is described by a set of $n(n-1)/2$ sufficient statistics. We derive the distribution of the sufficient statistics under the exponential model. Finally, we explore further generalizations of the Mallows model to which our framework applies.

Hence, we have shown that solving the consensus ranking problem is as hard as the expressed views of the panelists are divergent. On the positive side, if the n rankings do not differ wildly, in other words, if *there is a consensus*, then the consensus ranking can be found efficiently.

[1] We use the terms *permutation*, *ranking* and *linear order* interchangeably.

References

[Ailon et al., 2005] Ailon, N., Charikar, M., and Newman, A. (2005). Aggregating inconsistent information: Ranking and clustering. In *The 37-th ACM Symposium on the Theory of Computing (STOC)*. Association for Computing Machinery.

[Cohen et al., 1999] Cohen, W. C., Schapire, R. S., and Singer, Y. (1999). Learning to order things. *Journal of Artificial Intelligence Research*, 10:243–270. heuristics for consensus ordering with graph representation. what heuristics aren't good.

[Fligner and Verducci, 1986] Fligner, M. A. and Verducci, J. S. (1986). Distance based ranking models. *Journal of the Royal Statistical Society B*, 48:359–369. This is the paper!

[Lebanon and Lafferty, 2003] Lebanon, G. and Lafferty, J. (2003). Conditional models on the ranking poset. In *Advances in Neural Information Processing Systems*, number 15, Cambridge, MA. MIT Press.

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

Topic: Learning Algorithms, Ranking
: Preference: oral