Unsupervised Rank Aggregation with Domain-Specific Expertise

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

Consider the setting where judges are repeatedly asked to (partially) rank sets of objects, and assume that each judge tries to reproduce some true underlying ranking to the best of their ability. *Rank aggregation* aims to combine the rankings of such experts to produce a better joint ranking, and is a ubiquitous problem in Information Retrieval (IR) and Natural Language Processing (NLP). In IR, for instance, *meta-search* aims to combine the outputs of multiple search engines. In machine translation (MT), aggregation of multiple systems built on different underlying principles has received considerable attention recently (e.g. [5]).

In many practical applications such as these, one would expect the expertise of the constituent rankers to depend on the input domain. In IR, the quality of the rankings produced by the engines has been shown to depend on the type of issued query (e.g. [1]): some may be better at ranking product reviews, while others may specialize on ranking scientific documents. In MT system aggregation, the systems may be trained on different corpora (e.g. multi-lingual Hansards, or technical manuals), and tend to be fragile when tested on data sampled from a different distribution. Thus, the relative expertise of these systems depends on which distribution the test source language data is sampled from. Moreover, in these and many other aggregation examples, the input domain information in regards to the expertise of each judge is latent. A supervised learning approach to solving rank aggregation is likely to be impractical for many applications, since labeled ranking data is typically very expensive to obtain. Therefore, heuristics or indirect methods are often employed (e.g. [2]) to produce a surrogate for true preference information. Needless to say, supervision for typed ranked data is even harder to get.

The principal contribution of this work is a framework for learning to aggregate votes of constituent rankers with domain-specific expertise *without supervision*. Given only a set of constituent rankings, we learn an aggregation function that attempts to recreate the true ranking without the need for labeled data. The intuition behind our approach is simple: the subset of rankers which are experts in a given domain are better at generating votes close to true rankings for that domain and thus will tend to agree with each other, whereas the non-experts will not.

We extend a distance-based conditional model [4, 3] to include the notion of input data domains (or types). Suppose, a panel of K judges generate (partial) rankings $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_K)$ over a set of objects attempting to reproduce the true ranking π of type t. The proposed mixture model assigns the conditional probability to the latent pair as:

$$p(\pi, t|\boldsymbol{\sigma}, \boldsymbol{\theta}, \boldsymbol{\alpha}) = \frac{1}{Z(\boldsymbol{\theta}, \boldsymbol{\sigma})} \alpha_t \exp\left(\sum_{i=1}^{K} \theta_{t,i} \ d(\pi, \sigma_i)\right)$$

where $Z(\boldsymbol{\theta}, \boldsymbol{\sigma})$ is a normalizing constant. The free model parameters are a $T \times K$ matrix $\boldsymbol{\theta}$ ($\theta_{t,i}$ represents the degree of expertise of judge i for type t) and T mixture weights $\boldsymbol{\alpha}$.

We derive an EM-based algorithm to estimate the model parameters and apply the learning framework to full rankings and top-k lists, demonstrating significant improvements over a domain-agnostic baseline in both cases.

References

[1] Xiubo Geng, Tie-Yan Liu, Tao Qin, Andrew Arnold, Hang Li, and Heung-Yeung Shum. Query dependent ranking using k-nearest neighbor. In *Proc. of International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 115–122, 2008.

[2] T. Joachims. Unbiased evaluation of retrieval quality using clickthrough data. In *SIGIR Workshop on Mathematical/Formal Methods in Information Retrieval*, 2002.

[3] Alexandre Klementiev, Dan Roth, and Kevin Small. Unsupervised rank aggregation with distance-based models. In *Proc. of the International Conference on Machine Learning (ICML)*, 2008.

[4] G. Lebanon and J. Lafferty. Conditional models on the ranking poset. In *Proc. of the Conference on Advances in Neural Information Processing Systems (NIPS)*, volume 15, pages 431–438, 2003.

[5] Antti-Veikko I. Rosti, Necip Fazil Ayan, Bing Xiang, Spyros Matsoukas, Richard Schwartz, and Bonnie J. Dorr. Combining outputs from multiple machine translation systems. In *NAACL/HLT*, pages 228–235, 2007.

Topic: Learning Algorithms
Preference: Poster