Visualizing Spatial Proximity of Search Algorithms

Mingxuan Sun **Guy Lebanon**
Georgia Institute of Technology

Kevyn Collins-Thompson
Microsoft Research

Overview

We introduce a new dissimilarity function for ranked lists, the expected weighted Hoeffding distance, that has several advantages over current dissimilarity measures for partial rankings: it is (i) symmetric, (ii) interpretable with respect to search algorithms retrieving ranked lists of different lengths, (iii) flexible enough to model the increased attention users pay to top ranks over bottom ranks, (iv) computationally efficient, and (v) aggregate information over multiple queries in a meaningful way. We then use the measure with multi-dimensional scaling to visualize and explore relationships between the different search engines. Such visualization is highly effective for obtaining insights on which search engines to use, what search strategies users can employ, and how search results evolve over time.

Methodology and Experiments

The ranked list output by search algorithms $s_i(q), i = 1, \ldots .l$ given a query q forms ordered lists $\langle i_1, \ldots, i_k \rangle$ of subset of the websites on the internet $\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$. Different search algorithms may result in lists of different sizes k. The key for comparing ranked lists output by search engines is to define an appropriate dissimilarity measure $\overline{\rho}(s_i, s_j)$. One problem with existing dissimilarities, for example ones based on Kendall's tau is that they are not symmetric, not computationally efficient, not interpretable, or do not distinguish between disagreement at top rankings and at the bottom rankings.

In this paper we propose to define $\overline{\rho}$ using an expectation over the weighted Hoeffding distance, which provide a clear probabilistic interpretation and ensure that the measure is symmetric, computataionally efficient, interpretable, and flexible enough to distinguish disagreement at the top and at bottom ranks. The measure is

$$\overline{\rho}(s_i, s_j) = \mathsf{E}_{q\sim Q}\{\rho(s_i(q), s_j(q))\} = \mathsf{E}_{q\sim Q}\ \mathsf{E}_{\pi\sim\mathfrak{S}(s_i(q))}\mathsf{E}_{\sigma\sim\mathfrak{S}(s_j(q))}\{d_w(\pi,\sigma)\} \tag{1}$$

where $d_w(\pi, \sigma)$ is a weighted Hoeffding distance between permutations π, σ, $\mathsf{E}_{\pi\sim\mathfrak{S}(s_i(q))}\mathsf{E}_{\sigma\sim\mathfrak{S}(s_j(q))}$ is the expectation with respect to permutations π, σ that are sampled from the sets of all permutations consistent with the ranked lists output by the two search engines $s_i(q), s_j(q)$ (respectively), and $\mathsf{E}_{q\sim Q}$ is an expectation with respect to all queries sampled from a representative set of queries Q. In the absence of any evidence to the contrary, we assume a uniform distribution over the set of queries Q and over the sets of permutations consistent with $s_i(q), s_j(q)$. The weighted Hoeffding distance is equivalent to the earth movers distance and is the minimim amount of work needed to bring each item r from its rank in π to its rank in σ where moving an item from rank i to $i + 1$ costs w_i. That is $d_w(\pi, \sigma) = \sum_{r=1}^{n} d'_w(\pi(r), \sigma(r))$ with $d'_w(u, v) = \sum_{t=u}^{v-1} w_t$ assuming $u < v$. A monotonic decreasing weight vector, e.g., $w_t = t^{-q}$, $t = 1, \ldots, n - 1$, $q \geq 0$ correctly captures the fact that disagreements in top ranks should matter more than disagreements in bottom ranks. The expectation in (1) ensures that the measure is interpretabile and symmetric. We also show that it is computationally efficient with online complexity $O(k + l)$ (number of items ranked by s_i and s_j) and offline complexity $O(n + \max(k, l)^2)$.

We apply (1) with multi-dimensional scaling (MDS) for visualization. Figure 1 show the visualization over full ranking and partial ranking with $n = 5$. Figure 2 show visualization of 9 different search engines over 3 different query categories (each with 50 popular queries). Altavista, alltheweb, lycos, and yahoo all form a tightly knit cluster. A similar tight cluster contains google and aol. The dendrogram in top right of figure 3 output by standard hierarchical clustering confirms the analysis above visually. The clusters do in fact mirror the technology relationships that have evolved in the search engine industry. Figure 3 visualizes the sensitivity of search engines to popular query manipulation techniques and to the time in which the query was submitted.

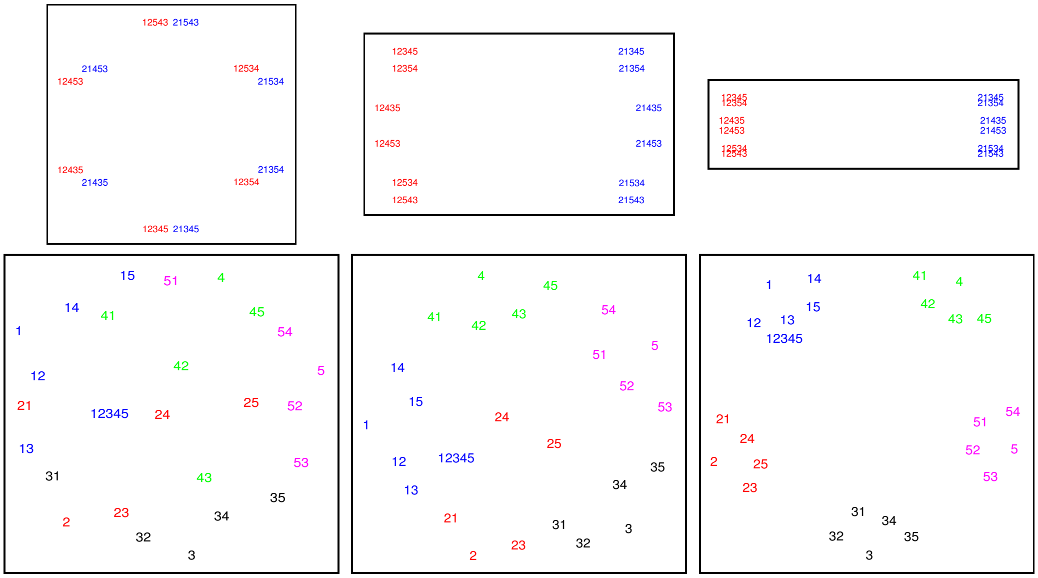

Figure 1: MDS embedding of permutations over $n = 5$ websites of full length (top row) and partial rankings of varying lengths (k varies) in the bottom row . The embeddings were computed using the weighted Hoeffding distance with uniform weight function $w_t = 1$ (left), linear weight function $w_t = 1/t$ (middle) and quadratic weight function $w_t = 1/t^2$ (right). Top: The permutations starting with 1 and 2 (colored in red) and the permutations starting with 2 and 1 (colored in blue) become more spatially disparate as the rate of weight decay increases. Bottom: The expected distance (1) separates ranked lists agreeing in their top rankings (denoted by different colors) better as the weights decay faster.

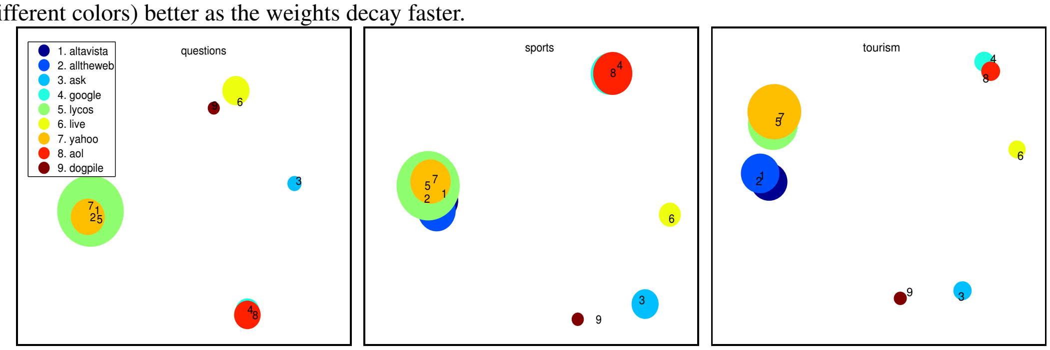

Figure 2: MDS embedding of search engine results over 3 sets of representative queries: questions, sports, and tourism. The MDS was based on the expected weighted Hoeffding distance with linear weighting $w_t = t^{-1}$ over the top 100 sites. Circle sizes indicate position variance with respect to within category queries.

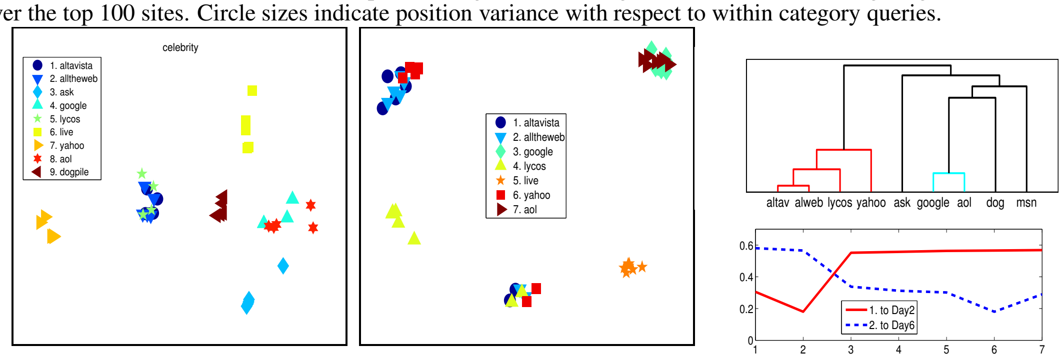

Figure 3: Left: MDS embedding of search engine results over the query category celebrity with different query manipulations. Each marker represents a combination of one of the 9 search engines and one of the 5 query manipulation techniques: $(a) w_1 w_2$, $(b) w_1 + w_2$, (c) "$w_1 w_2$", $(d) w_1$ and w_2, $(e) w_1$ or w_2. Middle: MDS embedding over 7 days for a set of temporal event queries. Yahoo cluster (yahoo, altavista, alltheweb, and lycos) shows a high degree of temporal variability, and in particular a sharp spatial shift on the third day. Right Bottom: The dissimilarity of Yahoo results over seven days with respective to a reference day. Right Top: Hierarchical clustering dendrogram for the nine search engines matches recent news in the commercial search engine sector (aol uses google technology, yahoo purchased altavista, alltheweb, and lycos).