Ranking Aggregation for Meta-Search Engine

Chunheng Jiang

Computer Science Department, RPI

Ranking Aggregation

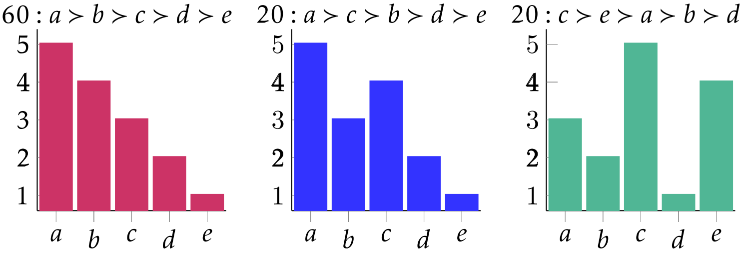
Ranking aggregation aims to produce a ranking list $\pi \in \Pi(\mathcal{C})$ from the preference profile P of n voters. Different aggregation methods (or voting rules) may produce different preference rankings.

1. Candidates: $\mathcal{C} = \{c_1, c_2, \ldots, c_m\}$
2. Ranking Space: $\Pi(\mathcal{C}) = \{\pi_1, \pi_2, \ldots\}$
3. Profile: $P = \{\sigma_1, \sigma_2, \ldots, \sigma_n\}$, $\sigma_i \in \Pi(\mathcal{C}), 1 \leq i \leq n$

The optimal Kemeny ranking is commonly used to evaluate the performance of an aggregation method, and we expect the combined result π^* is or at least very close to the optimal Kemeny ranking over P, i.e.

$$\pi^* \approx_{d_{KL}} \underset{\pi \in \Pi(\mathcal{C})}{\arg\min}\, d_{KL}(\pi, P)$$

Ranking Aggregation



Positions	Margin	Condorcet	Borda
$\pi_1(a) = 1, \pi_1(b) = 2$	1		$s_a - s_b = 1$
$\pi_2(a) = 1, \pi_2(b) = 3$	2>1	$a \succ b$	$s_a - s_b = 2$
$\pi_3(a) = 3, \pi_3(b) = 4$	1<1		$s_a - s_b = 1$

Ranking Aggregation

	0.6	0.2	0.2
a	1	1	3
b	2	3	4
c	3	2	1
d	4	4	5
e	5	5	2

1. $a \succ b$: 1/2, 1/3, 3/5
2. $a \succ c$: 1/3, 1/2, 3/1

Important Factors:

1. Positional information
2. Positional difference
3. Unequal ballot-weights

Ranking Aggregation

1. Positional score based methods, e.g. Borda count
2. Kemeny optimal aggregation
3. Probabilistic models
4. Machine learning based methods

Pairwise Margin Rule

$$s_\pi(i,j) = \frac{\pi(j)-\pi(i)}{\min\{\pi(i),\pi(j)\}}, \forall c_i, c_j \in \mathcal{C},$$

where $\pi(i)$ is the position of c_i in π. When c_i is ranked first, $\pi(i)=1$; when c_i has no place at all, $\pi(i)=|\pi|+1$.

The aggregated score for each individual candidate is a weighted sum over all preference rankings:

$$s_i = \sum_{\pi\in\Pi(\mathcal{C})} w_\pi s_i(\pi) = \sum_{\pi\in\Pi(\mathcal{C})} \sum_{1\le j\le n} w_\pi \frac{\pi(j)-\pi(i)}{\min\{\pi(i),\pi(j)\}}, 1 \le i \le n,$$

where $\sum_{\pi\in\Pi(\mathcal{C})} w_\pi = 1$ and $0 \le w_\pi \le 1$.

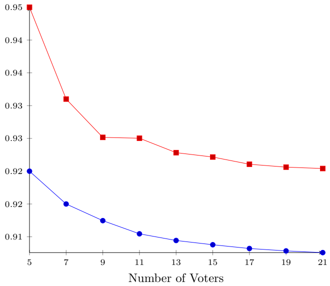
Pairwise Margin Rule

- Monotonically increasing w.r.t $\pi(i)$, i.e.

$$\forall c_i \succ_\pi c_j, s_i(\pi) > s_j(\pi)$$

- If $m \geq 3$, no positional scoring rule satisfies the Condorcet consistency. How about making it satisfying the criterion as nearly as possible?

YES. Better satisifiability to Condorcet consistency than Borda rule when $m = 3$.

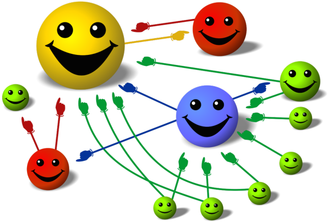


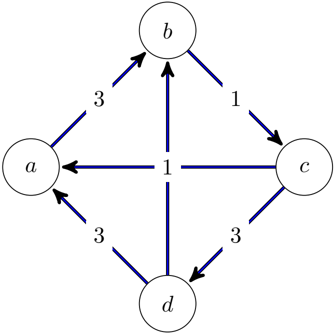
Link-based Voting Rule

$$S_i \leftarrow (1-d)/N + d \sum_{j \rightarrow i} S_j/N_j$$

where S_i is the PageRank score of page i (or candidate c_i in voting case), it can be computed based on two factors: the importance of other pages (e.g. j) which have hyperlinks to i and the total number N_j of hyperlinks they have. Page j's PageRank score will be divided into N_j parts equally, and $j \rightarrow i$ indicates that page j refers to i, or has the hyperlink of page i, and d is the probability of visiting pages via hyperlinks, and $1-d$ is the probability of random walk.

Link-based Voting Rule





The weighted majority graph in voting can be used to model a weighted PageRank surfing method, and produce a preference ranking iteratively.

Link-based Scoring Rule

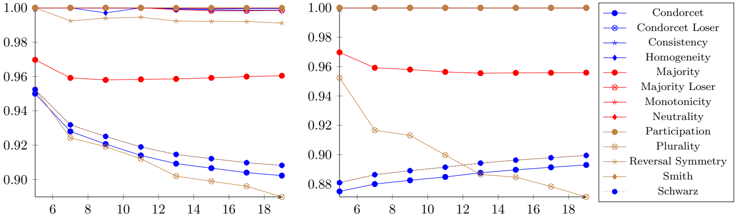
Weighted PageRank model:

$$\begin{cases} P_i = \sum_{j=1}^{n} P_j \cdot Pr(c_i|c_j) \\ Pr(c_i|c_j) = \sigma(\sum_{\pi \in \Pi(\mathcal{C})} s_\pi(i,j)) / \Big[\sum_{k=1}^{n} \sigma(\sum_{\pi \in \Pi(\mathcal{C})} s_\pi(k,j))\Big] \end{cases}$$

Scores are distributed based on the probability of c_i is preferred to c_j, which is computed with both absolute and relative (margin) positional information. We get the scores by solving $P = AP$ iteratively, where $a_{ij} = Pr(c_i|c_j)$ and $\sum_i a_{ij} = 1$, $\sigma(x) = 1/[1 + e^{-x}]$, A is a stochastic matrix.

Link-based Scoring Rule

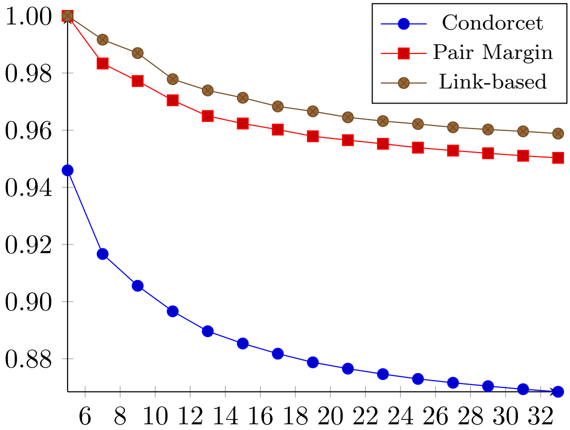
Link-based Scoring Rule (L) v.s. Borda Rule (R) w.r.t their satisfiabilities to popular fairness criteria:



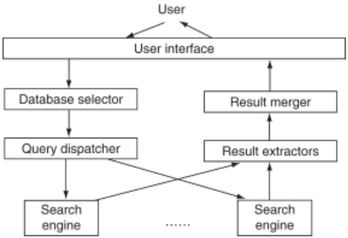
Good news and Bad news: Condorcet Criterion

Link-based Scoring Rule

Link-based Scoring Rule, Pair Margin Rule v.s. Condorcet Method w.r.t their similarities to Borda Rule:

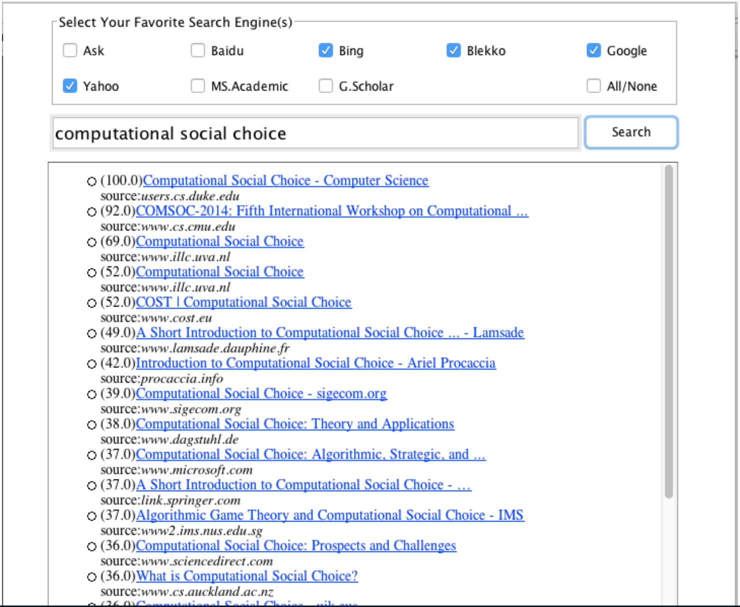


Meta Search Engine



Crawler -> Indexer -> Searcher -> Ranker (ranking aggregation comes into play)

- Individual Search Engines: Google, Bing, Yahoo, Ask, Baidu and Blekko.
- Aggregated: first SERPs
- Pros: diverse topics, flexible demonstration, and no distractions from Ads.
- Cons: no independent data base, and long waiting time.


Select Your Favorite Search Engine(s)
Ask
Baidu
Bing
Blekko
Google
Yahoo
MS.Academic
G.Scholar
All/None
computational social choice
Search
(100.0)Computational Social Choice - Computer Science
source:users.cs.duke.edu
(92.0)COMSOC-2014: Fifth International Workshop on Computational ...
source:www.cs.cmu.edu
(69.0)Computational Social Choice
source:www.illc.uva.nl
(52.0)Computational Social Choice
source:www.illc.uva.nl
(52.0)COST | Computational Social Choice
source:www.cost.eu
(49.0)A Short Introduction to Computational Social Choice ... - Lamsade
source:www.lamsade.dauphine.fr
(42.0)Introduction to Computational Social Choice - Ariel Procaccia
source:procaccia.info
(39.0)Computational Social Choice - sigecom.org
source:www.sigecom.org
(38.0)Computational Social Choice: Theory and Applications
source:www.dagstuhl.de
(37.0)Computational Social Choice: Algorithmic, Strategic, and ...
source:www.microsoft.com
(37.0)A Short Introduction to Computational Social Choice - ...
source:link.springer.com
(37.0)Algorithmic Game Theory and Computational Social Choice - IMS
source:www2.ims.nus.edu.sg
(36.0)Computational Social Choice: Prospects and Challenges
source:www.sciencedirect.com
(36.0)What is Computational Social Choice?
source:www.cs.auckland.ac.nz