Diverse and Fair Ranking

CS6101

Soumen Chakrabarti

Diversity

Ambiguous query

- Many "John Smith"s: painter, footballer, politician…
  - Which "John Smith" is a particular user looking for?
  - 10 free slots, don't fill them all with the same J.S.
  - Max prob of satisfying "expected" user with $\geq$ 1 hit
- Can get more subtle than name ambiguity
- [macbook pro reviews]
  - Diversity across positive and negative reviews
- [light blue formal shirt]
  - Diversity across material? Price? Texture? Made-in-?
- 10 slots interact; individual scoring impossible
  - E.g., redundancy among top-scoring docs

A facility location formulation for diversity

- Fix a query; relevance of doc i is denoted $r(i)$
- $s(i, j)$ is the similarity between docs i, j
- N candidate docs
- Must choose $S \subseteq [N] = \{1, \cdots, N\}$, $|S| \le K$
- Any doc i not included in S must be very similar to ("covered by") some doc j that is included
- Measure of coverage $c(i; S) = \max_{j \in S} s(i, j)$
- Choose S to maximize $\sum_{i \in [N]} r(i)\ c(i; S)$
- A form of facility location

Greedy optimization

- Begin with $S = \emptyset$
- While $|S| < K$
  - For each $j \notin S$
    - Marginal benefit of j is $\sum_{i \in [N]} r(i)\ c(i; S \cup j)$
  - Include that j with largest marginal benefit

Relaxed integer program

- Decision variables
  - $x(i) \in \{0,1\}$ (whether $i \in S$)
  - $z(i,j) \in \{0,1\}$ (whether i uses j as cover: possible only if $j \in S$, which means $z(i,j) \leq x(i)$ for all i, j
- $\sum_{i \in [N]} x(i) = K$
- Objective is $\max_{\mathbf{x},\mathbf{Z}} \sum_{i,j} r(i)\ s(i,j)\ z(i,j)$
- Relax $x(i), z(i,j) \in \{0,1\}$ to $\widetilde{x}(i), \widetilde{z}(i,j) \in [0,1]$
- Sample $x(i) \sim \mathrm{Bernoulli}(\widetilde{x}(i))$
  - K items chosen only in expectation

Other reasonable objectives

- Selected set should have high relevance $\sum_{i \in S} r(i)$
- Similarity between selected pairs $\sum_{i,j \in S} s(i,j)$ should be low
- Not clear $s(i,j) = s(j,i)$ is valid
  - Doc i may render doc j redundant, but not vice versa
- Later, when we study hyperlink graphs
  - What if i has a prominent link to j?
- Some formulations can use asymmetric coverage

Maximum marginal relevance (MMR)

- Greedy heuristic LP approximation
- Initial set of chosen docs is $S = \emptyset$
- While $|S| < K$ docs chosen
  - Choose next doc using

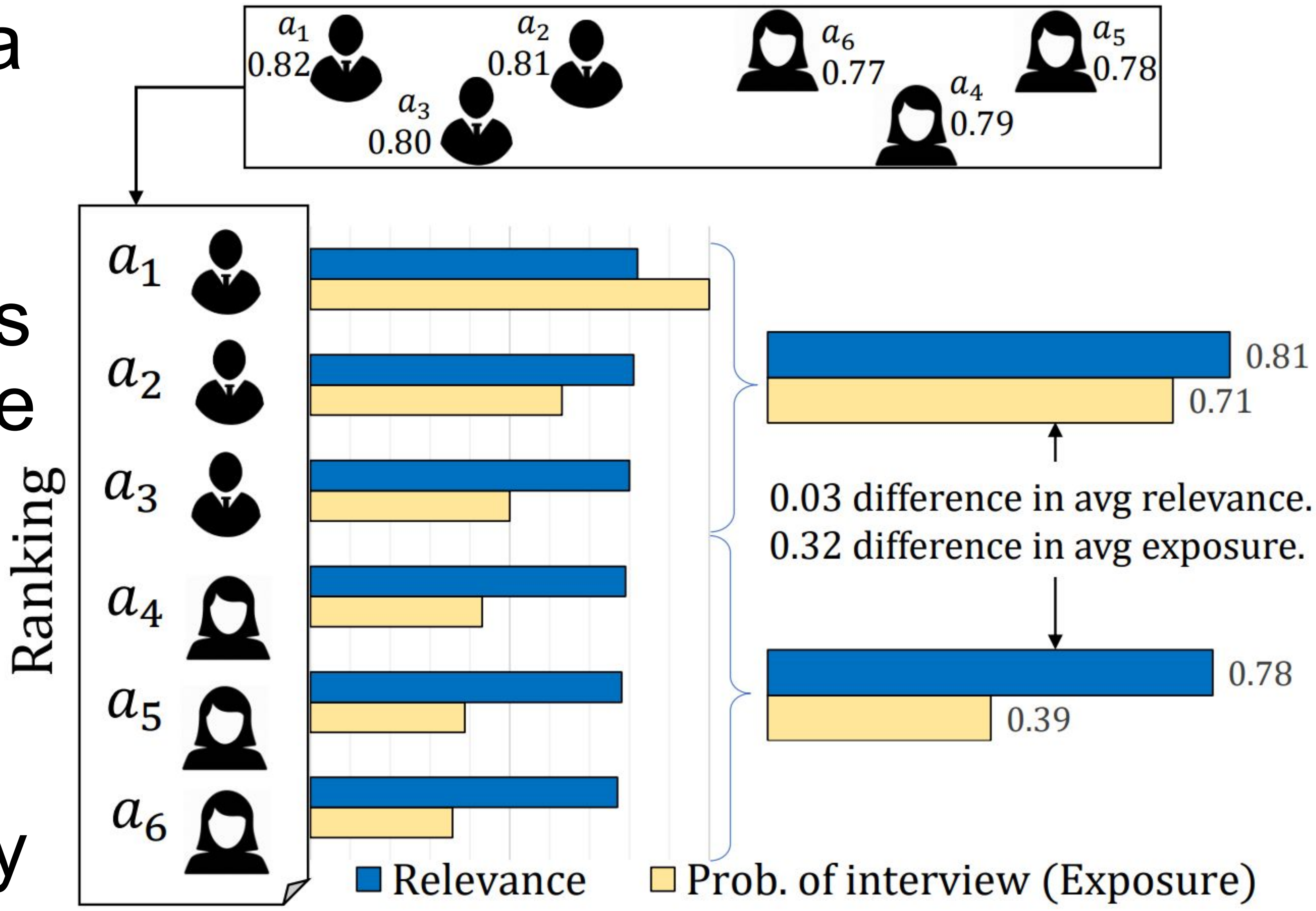
$$\underset{d \in [N] \setminus S}{\operatorname{argmax}} \lambda s(q, d) - (1 - \lambda) \max_{d' \in S} s(d, d')$$

- Most similar to query
- Least similar to anything already included
- Difference of two similarities needs some tuning

Fairness

Relevance vs. opportunity gaps

- Relevance score is a noisy estimate
- System should be robust to small errors in relevance estimate
- Instead, small changes in relevance result in large changes in exposure/opportunity



Proportionate representation

- Image search: disproportionately many male CEOs

Notation

- N items being ranked for a fixed query
  - Ranking = $N \times N$ permutation matrix P
- Rank j has visibility $v(j)$, e.g., $v(j) = \frac{1}{\log(1+j)}$
  - Assume $v(j)$ decreases with j
- Doc i has relevance $r(i)$
- Assume we can factor utility as

$$\mathrm{Util}(P) = \sum_{i,j} P(i,j)\; r(i)\; v(j) = r^\top P v$$

- Unconstrained optimization is equivalent to ordering i by decreasing $r(i)$ — most relevant gets most visibility
- But we also want P to be “fair”

Groups and exposure

- Exposure of doc i is $\mathrm{Exp}(i|P) = \sum_j P(i,j)v(j)$
- A group G is a subset of docs
  - Groups may partition docs (e.g., gender, race)
- Group average exposure
$$\mathrm{Gexp}(G|P) = \frac{1}{|G|}\sum_{i\in G} \mathrm{Exp}(i|P)$$
- Policy: all groups should have equal exposure
- Group average relevance $\mathrm{Grel}(G) = \frac{1}{|G|}\sum_{i\in G} r(i)$
- Policy: group average exposure should be proportional to group average relevance
- Both enforce further linear constaints on P

Sampling permutations

- In all cases, permutation matrix P relaxed to doubly stochastic matrix $\widetilde{P}$ $\rightarrow$ integer linear program
- (Berkhoff, von Neumann) Any doubly stochastic matrix $\widetilde{P}$ can be decomposed into a convex combination of permutation matrices $\widetilde{P} = \theta_1 P_1 + \cdots + \theta_M P_M$, where $M \in O(N^2)$
  - Repeated removal of perfect matching
- Use $(\theta_1, \cdots, \theta_M)$ to sample from multinomial distribution over permutation matrices

Dissociative vs associative selection

- Diversity is dissociative
  - If $r(i)$, $s(i, j)$ large, may reduce need to include j
  - Even if $r(j)$ is large
- Associative scoring (next module)
  - If $r(i)$, $s(i, j)$ large, may also include j
  - Even if $r(j)$ is small
  - Query {cat}, i includes {cat, panther}, j contains {panther} and is actually more relevant than i
  - $r(j) < r(i)$ because of limitation of sparse matching
- Both associative and dissociative scoring have their places