Seeker: Real Time Interactive Search

Problem Statement
Seeker: allows users to adaptively refine search rankings in real time, through feedbacks in the form of likes and dislikes.

User has a mental picture of desired item and can answer ordinal questions like: "Is this item similar to what you have in mind?"

Problem formulation:
- User has in mind target item $t$.
- Catalog has $N$ items.
- Display $M << N$ items at a time.
- User interacts with items through likes and dislikes.
- User can interact with same item repeatedly.

Aim: Surface item $t$ in $M$ using the fewest interactions possible.

Logical Components

Pair Formation
Definitions/assumptions:
- Represent item $i$ as embedding feature vector $x_i$.
- Assume high correlation between human similarity and distance metric in embedding space.
- Preference pair $s_{ij}$ consists of liked item $a_j$ and disliked item $b_j$.
- User prefers $i$ over $j$ given $t$:
  $$||x_i - x_j||^2 < ||x_i - x_t||^2$$

Probability of choosing $i$ over $j$:
$$P(s_{ij}|t) = \text{logistic}(||x_i - x_j||^2 - ||x_i - x_t||^2)$$

Item Ranking
Generate discrete distribution using softmax (Boltzmann) equation:
$$p_j = \frac{e^{g_j}}{\sum_i e^{g_i}}$$

Sample by Gumbel-softmax trick:
$$\argmax_j (g_j + \text{Gumbel}(0,1)) \sim \frac{e^{g_j}}{\sum_i e^{g_i}} = p_j$$

Anneal using inverse temperature $\eta_k$ to balance explore/exploit:
$$\argmax_j (g_j + \eta_k \text{Gumbel}(0,1)) \sim \frac{e^{g_j + \eta_k g_j}}{\sum_i e^{\eta_k g_i}} = p_j$$

This is similar to Thompson Sampling bandit, but doesn’t account for estimates uncertainty. We include estimate uncertainty thru annealing:
$$\argmax_j \left( g_j + \frac{C^2}{n_j} \text{Gumbel}(0,1) \right)$$

where $n_j$ is the number of times a user interacts with item $j$. $C$ depends on the variance factor of the reward subgaussian distribution. In our case, reward is binary and $C^2 = \frac{1}{8}$.

Algorithm Summary

Experimental Results
- Compare Seeker to pure exploitation, random, and epsilon-greedy bandit.
- Interact with system up to $K$ interactions or until $t \in M$.
- Set $K = 15$, $M = 12$, $N = 2228$.
- Monitor target normalized rank $p$.
- Recall@$p/10$: Percentage of sessions with $p_i \leq p_j$.

Ari Biswas, Thai T Pham, Michael Vogelsong, Benjamin Snyder, and Houssam Nassif