A New Probabilistic Model for Top-k Ranking Problem

Shuzi Niu, Yanyan Lan, Jiafeng Guo, Xueqi Cheng
Institute of Computing Technology, Chinese Academy of Sciences
niushuzi@software.ict.ac.cn, {lanyanyan, guojiafeng, cxq}@ict.ac.cn

1. MOTIVATION

- **Top-k ranking problem:** In top-k ranking setting correct ranking at the top k positions is much more important because users care more about top positions in practice.

Items to be rank: {A,B,C,D,E,F,G,H}

How?

Top-3 ranking: G<F<A<{B,C,D,E,H}

2. Ways to look at top-k ranking

Sequential Process

$\{A, B, C, D, E, F, G, H\}$
↓
$G \prec \{A, B, C, D, E, F, H\}$
↓
$G \prec F \prec \{A, B, C, D, E, H\}$
↓
$G \prec F \prec A \prec \{B, C, D, E, H\}$

Is sequential reasonable?

- Lacks theoretical support
- Heuristic though employed in many scenario
- Not the unique generative process

- Top-k ListMLE → Luce model
- Top-k CPS → Luce + Mallows

Hierarchical Process

$\{A, B, C, D, E, F, G, H\}$
↓
$\{A, F, G\} \prec \{B, C, D, E, H\}$
↓
$G \prec F \prec A$

Riffled Independence

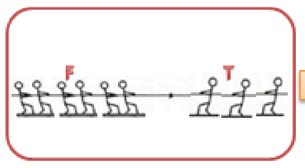
$P(G \prec F \prec A \prec \{B,C,D,E,H\})$
$\Omega = \{A, F, G\}$
$M = \{B, C, D, E, H\}$
$P(\Omega \mid M) P(G \prec F \prec A) P(\{B, C, D, E, H\})$

3. HOM (Hierarchical Ordering Model)

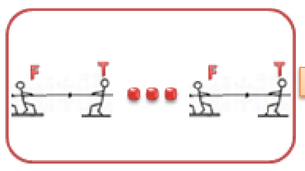
- Probability of top-k ranking σ

$$P(\sigma) = P_{TF}(T \prec F) \cdot P_T(\sigma_T)$$

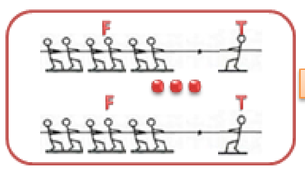
– Distributions of ordering between T and F

(1) 

$$P_{TF}(T \prec F) \propto \frac{\sum_{x_i \in T} \exp(f(x_i))}{\sum_{j=1}^{|D|} \exp(f(x_j))}$$

(2) 

$$P_{TF}(T \prec F) \propto \prod_{x_i \in T} \prod_{x_j \in F} \frac{\exp(f(x_i))}{\exp(f(x_i)) + \exp(f(x_j))}$$

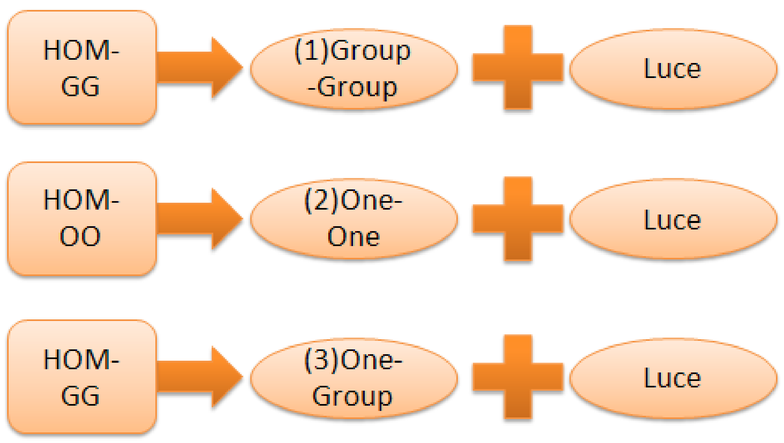
(3) 

$$P_{TF}(T \prec F) \propto \prod_{x_i \in T} \frac{\exp(f(x_i))}{\exp(f(x_i)) + \sum_{x_j \in F} \exp(f(x_j))}$$

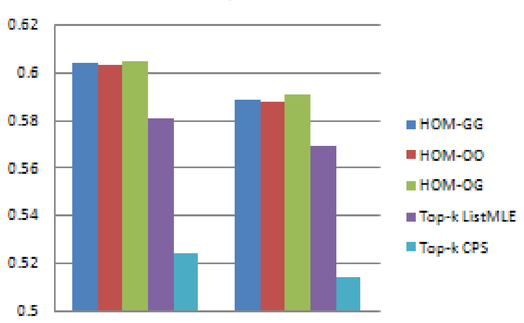
– Distributions of full ranking on T

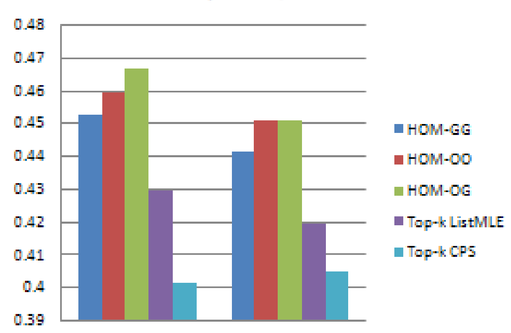
Luce Model: $P_T(\sigma_T) = \prod_{i=1}^{|T|} \frac{\exp(f(x_{\sigma_T(i)}))}{\sum_{j=i}^{|T|} \exp(f(x_{\sigma_T(j)}))}$

- HOM (Hierarchical Ordering Model)



4. EXPERIMENTAL RESULTS





- Top-10 MQ2007 and Top-10 MQ2008 are subsets of MQ2007-list and MQ2008-list respectively.
- κNDCG@10 and κERR are used as evaluation measure for the emphasis of top positions.
- Hierarchical ordering models outperform sequential generative model significantly.

5. CONCLUSIONS

Conclusion

- HOM provides a more appropriate way to look at top-k ranking data
- HOM outperforms previous top-k probabilistic ranking model significantly

Future work

- The hierarchical generative process can be used to other work, such as rank aggregation.
- More flexible models are needed to model the hierarchical generative process of top-k ranking data