A Probabilistic Account of the Uncertainty Due to Ties in Rank-Biased Overlap

Efficient Estimation of the Uncertainty Distribution

Lukáš Chládek <lchladek@tudelft.nl>

A Thesis Submitted to the EEMCS faculty,
Delft University of Technology

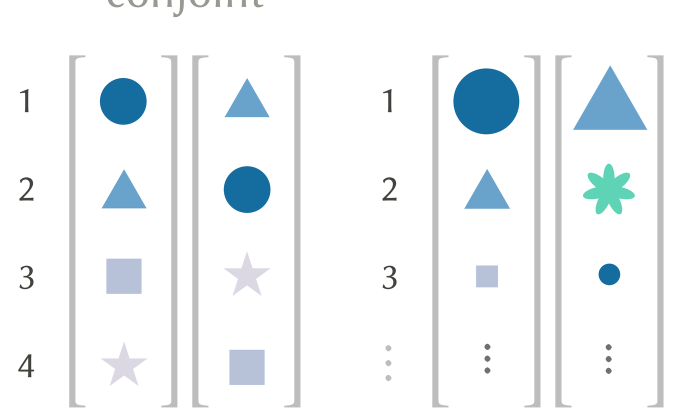
Supervised by **Prof. Julián Urbano**

TU Delft

1 Introduction

Rank similarity compares rankings of items.

It is useful in **Information Retrieval** when comparing search results: What makes a good retrieval system? How close are Google results to Bing?



Rank-Biased Overlap [1] compares indefinite rankings such as search results.

Properties of RBO on indefinite rankings:

- **Infinite and top-weighted.** *(Page 1 of Google is more important than Page 453.)*
- **Non-conjoint.** *(Google may find sites Bing doesn't know.)*

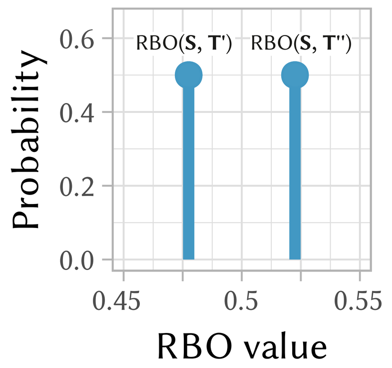
Tied data creates uncertainty.

rank-biased overlap

Uncertainty due to **unseen items** is part of Webber et al.'s original work: the residual $\mathbf{RBO_{res}}$.

Uncertainty due to **ties** is described by Corsi and Urbano. They define its bounds with $\mathbf{RBO^{low}}$ and $\mathbf{RBO^{high}}$ [2, 3].

Ties in a pair of rankings create a **distribution of uncertainty.** See the example below and Figure B.



S = < A B C X Y .
T = < A (B C) D E >

T' = < A B C D E >
T'' = < A C B D E >

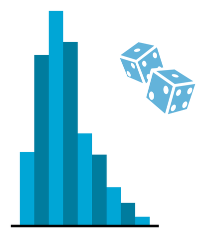
The tie introduces uncertainty.
Each way of breaking it gives a different RBO; ignoring this range leads to error in real data [3].

2 Research Question

How can the uncertainty of Rank-Biased Overlap for tied rankings be represented probabilistically?

The uncertainty due to ties is a **discrete probability mass function** for each ranking pair. Corsi and Urbano give points on the PMF: $\mathbf{RBO_a}$ [1], $\mathbf{RBO^{low}}$ **and** $\mathbf{RBO^{high}}$ [2].

This work proposes a **probabilistic model** of the distribution, giving arbitrary quantiles and CI without searching $O((n!)^2)$ tie arrangements.



How can we estimate this complex distribution:

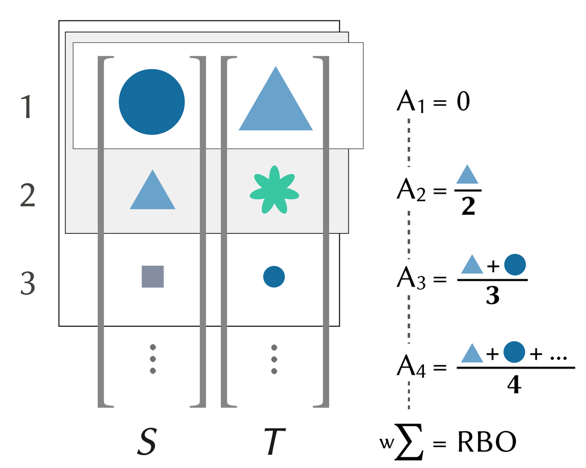
- deterministically
- faster than $O((n!)^2)$
- with high accuracy?

3 RBO Estimation by Convolution

We reformulate Rank-Biased Overlap as a **sum of item contributions.**

RBO: a weighted **sum of agreements.**

$$\mathrm{RBO}(S,T,p) = (1-p)\sum_{d=1}^{\infty} p^{d-1} A_d$$



RBO: a sum of **contribution constants.**
Contribution is given by the item's effective rank.

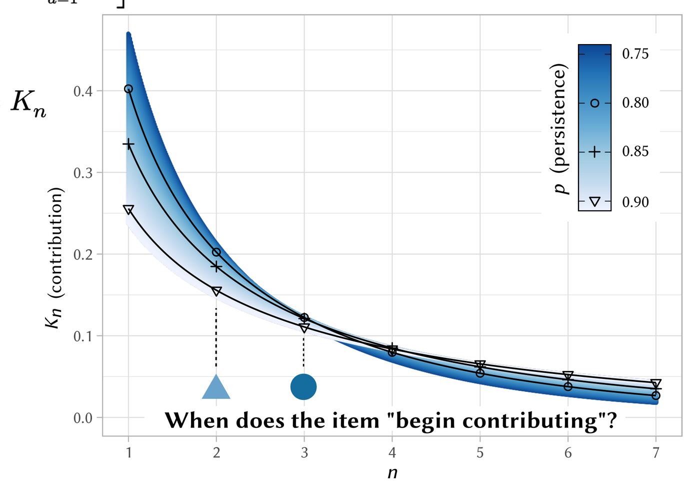
$$C_i(S,T,p) = \frac{1-p}{p}\left[\ln\frac{1}{1-p} - \sum_{d=1}^{M_i-1}\frac{p^d}{d}\right]$$

$$M_i = \max\left(S_{(i)}, T_{(i)}\right)$$

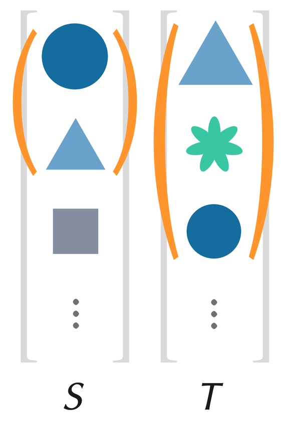
$$\forall i \quad M_i = n \leftrightarrow C_i = K_n$$

i	M_i	C_i
▲	2	K_2
●	3	K_3
✱	> 3	< K_3
■	> 3	< K_3

By constructing RBO item by item, the model becomes a sum of smaller parts.



RBO: a sum of **random variables.**
Contribution is an RV with respect to all possible permutations of ties.



i	M_i
▲	1 / 2 / 3
●	1 / 2 / 3
✱	> 3
■	> 3

$P[M_▲ = 1] = 1/6$
$P[M_▲ = 2] = 1/2$
$P[M_▲ = 3] = 1/3$

RBO: estimated by **iterative convolution.**
We assume that the RVs are independent and piece a PMF together item by item.

$$C_▲ \otimes C_● \otimes \ldots \approx \mathrm{RBO}(S,T,p)$$

$$\otimes : (\mathbb{R},\mathbb{R})^m, (\mathbb{R},\mathbb{R})^n \to (\mathbb{R},\mathbb{R})^q$$

$$\otimes(X,Y) = \forall x \in X : \forall y \in Y : (x+y,\ P[x]P[y])$$

Marginal item PMF C_C

c	$P[C_C]=c$
0.02	2/3
0.03	1/3

Approx. PMF of $C_A + C_B$

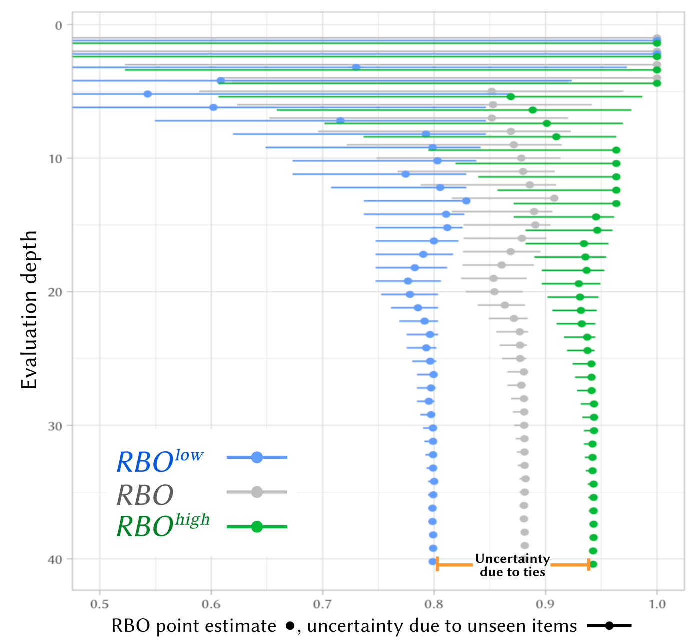
c	$P[C_A+C_B]\approx c$
0.4	2/5
0.5	3/5

$\xrightarrow{\otimes}$ Approx. PMF of $C_A + C_B + C_C$

c	$P[C_A+C_B+C_C]\approx c$
0.42	4/15
0.43	2/15
0.52	6/15
0.53	3/15

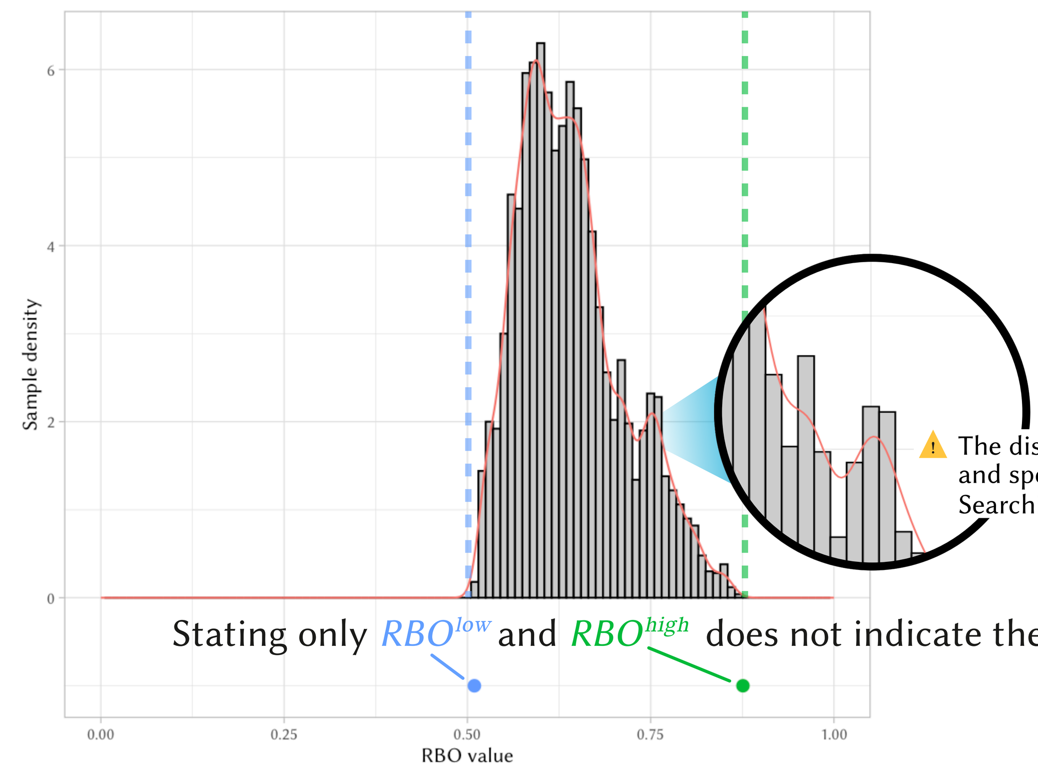
This works under the assumption of indep.
Covariance causes impossible values.
e.g. ● and ▲ both having $M_i = 1$.

fig. A Convergence of the RBO residual



Uncertainty due to **unseen items** converges.
Uncertainty due to **ties** remains.

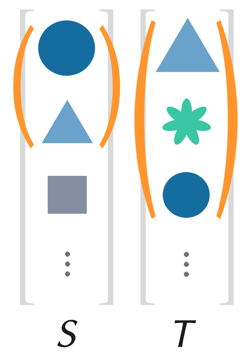
fig. B A distribution of uncertainty



Stating only RBO^{low} and RBO^{high} does not indicate the distribution's shape.

4 Reducing the Covariance

To better approximate the joint distribution, we use basic ranking rules†. **Apply rules to a "bitstring"** encoding the frequencies of M_i.



"$\mathbf{b_1 b_2 b_3}$..."
of K_3
of K_2
Frequency of K_1

Possibilities given by S T:

"200...."
"020...."
"002...."
"110...."
"101...."
"011...."

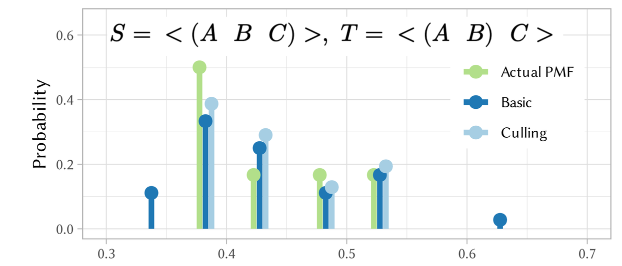
this fails rule 2

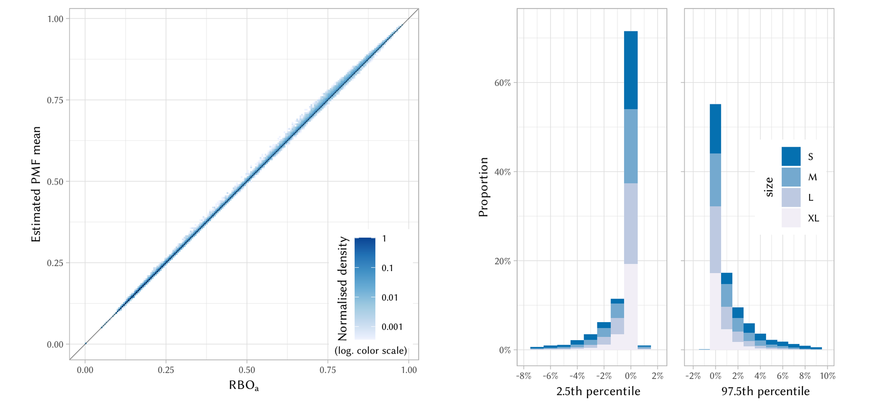
Impossible RBO result eliminated:
RBO = $K_1 + K_1$

The **culling convolution,** cull, removes any outcomes not matching the output. The probabilities are then renormalized.

† **Rule 1.** The sum of digits in a bitstring is equal to the number of contributing items.
Rule 2. The cumulative sum of digits up to index d is less than or equal to d.
Rule 3. No digit is greater than 2.

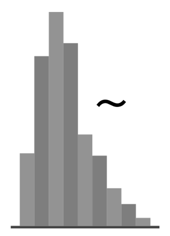
Comparison of ⊗ to cull on an example ranking.





Empirical evaluation on 150 000 TREC-like ranking pairs.

5 Conclusion and Future Work



Is **iterative convolution** a viable est. method?

- It is deterministic
- It is faster than $O((n!)^2)$
- Accuracy tested on 150k rankings

It is a viable method for arbitrary quantiles.
In practice: $\mathrm{RBO^{low}}$ and $\mathrm{RBO^{high}}$ are still the practical solution for quoting uncertainty.

Future work

- Further evaluation is necessary to establish the speed/accuracy tradeoff of the truncated iterative convolution.
- Find a better culling convolution than normalization
- Add more rules, potentially using the tie structure of the ranking itself.

6 Bibliography

[1] **Corsi, Matteo, and Julián Urbano.** 2024a. "The Treatment of Ties in Rank-Biased Overlap." In Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, 251–60. SIGIR 2024. Association for Computing Machinery. https://doi.org/10.1145/3626772.3657700.

[2] **Corsi, Matteo, and Julián Urbano.** 2024b. "How Do Ties Affect the Uncertainty in Rank-Biased Overlap?" In Proceedings of the 2024 Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region, 125–34. SIGIR-AP 2024. Association for Computing Machinery. https://doi.org/10.1145/3673791.3698422.

[3] **Lin, Jimmy and Peilin Yang. 2019.** The Impact of Score Ties on Repeatability in Document Ranking. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'19), July 2019. ACM, 1125–1128. https://doi.org/10.1145/3331184.3331339

[4] **Webber, William, Alistair Moffat, and Justin Zobel. 2010.** "A Similarity Measure for Indefinite Rankings." Transactions on Information Systems 28 (4): 1–38. https://doi.org/10.1145/1852102.1852106.