ON RANK-BIASED OVERLAP WITH FINITE AND CONJOINT DOMAINS

Oscar Kriebel
Delft University of Technology - CSE3000

BACKGROUND

Ranking similarity can be found in:

- Search engines
- Popular magazines
- Statistical Comparison

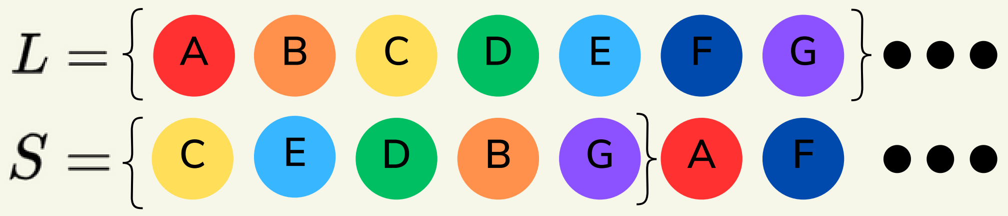
Rank-Biased Overlap(RBO) is a ranking similarity measure that is able to handle different properties of rankings:

Top-Heaviness: The values higher up in the ranking are more valueable.
Incompleteness: A ranking may not contain all of the items that are in the ranking's domain.
Indefiniteness: Rankings can be cut off at any point.

$$RBO = \sum_{d=1}^{\infty} A_d w_d$$
$$= (1-p)\sum_{d=1}^{\infty} \frac{X_d}{d} p^{d-1}$$

Where **A** is the **agreement**, or **similarity**, of two rankings, and **w** is the **weight**, or **importance**, of that **depth d**. This weight is defined by a **persistence, p,** and the agreement by the **intersection, X**.

Rankings can be of different lengths and may not rank all items. There are **seen** and **unseen** sections of the rankings are compared.



RBO is defined within a range MIN-MAX with an extrapolated EXT for the final result.

FINITE RANK-BIASED OVERLAP

$$RBO^f = \frac{1-p}{1-p^n}[\sum_{d=1}^{l} \frac{X_d}{d} p^{d-1} + \frac{X_l}{l}\sum_{d=l+1}^{n} p^{d-1} + \sum_{d=s+1}^{n}(\phi - \frac{X_l}{l})\frac{d-s}{n-s}p^{d-1}]$$

The two rankings share a **domain length of n**:

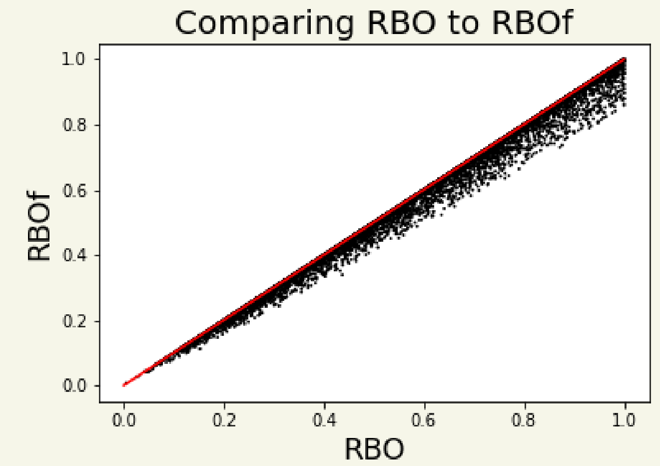
$$n = |\mathcal{D}_S| = |\mathcal{D}_L|$$

These domains, D, define the **level of conjointness, Φ**:

$$\phi_{S,L} = \frac{\mathcal{D}_S \cap \mathcal{D}_L}{n}$$

Properties of finite rankings:

- Made from a **known domain**.
- Can be shared or not but **same length**.
- Has a degree of **conjointness**, Φ.
- Agreement moves in steps.
- More sensitive on persistence, p.



FULLY CONJOINT RANK-BIASED OVERLAP

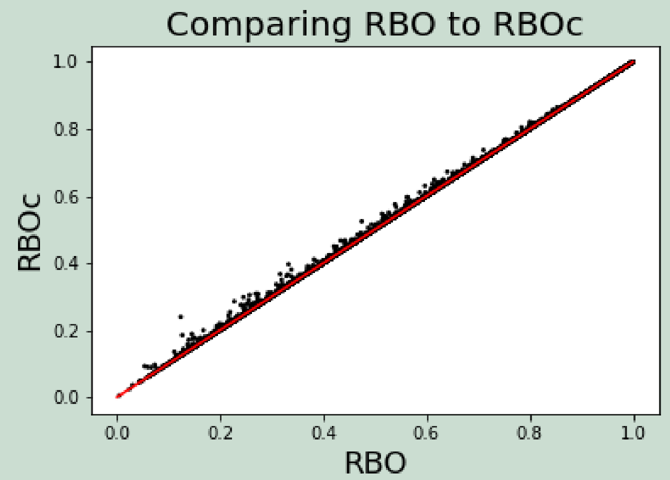
$$RBO^c = p^l + \frac{1-p}{p}[\sum_{d=1}^{l}\frac{X_d}{d}p^d + \sum_{d=s+1}^{l}\frac{X_s(d-s)}{ds}p^d + (X_l - X_s + \frac{l(X_s-s)}{s})(ln[\frac{1}{1-p}] - \sum_{d=1}^{l}\frac{p^d}{d})]$$

Agreement must tend towards 1 in order to be fully conjoint and can be modeled as:

$$lim_{d\to\infty}\frac{d-l}{d} = 1$$

Properties of fully conjoint rankings:

- Share all the **same items**.
- Do not need to know all of the items in the domain.
- Domain length is **infinite**.
- Modeled to **share items quickly**.
- Never smaller than traditional RBO.



RESEARCH QUESTION

Define RBO for fully conjoint and/or finite rankings.

- What occurs in the RBO measure in the case of **fully conjoint rankings** rather than assuming **disjointness**?
- How does the RBO measure change when there is a **known finite domain** and how does it change with both a **conjoint** and **non-conjoint domain**?

CONCLUSION AND FURTHER WORK

RBOc is used when you know that the two rankings **share all of the same items**, but do not know how many items there are. This tends to be **larger** that traditional RBO.

RBOf is used when you know the domains that the two rankings have, where they are the **same length** with a degree of **conjointness**. This tends to be **smaller** and **more sensitive** than traditional RBO.

Based on this, there is more work that can be done:

- **Redefine** RBO using the degree of conjointness, Φ.
- **Tighten the bounds** on RBOf using the degree of conjointness, Φ.
- **Redefine** RBOf to allow for **domains of different length**.
- Change the **assumption on ties** for RBOf and RBOc
- Compare different **models** on the agreement in the unseen sections.
- See **applications** of the equations to practical and real-world examples to see effectiveness.

REFERENCES

W. Webber, A. Moffat, and J. Zobel. *A similarity measure for indefinite rankings*. ACM Transaction on Information Systems, 2010.

M. Corsi and J. Urbano. *The treatment of ties in rank-biased overlap*. Proc. 47th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. 2024