Fairness and Explainability in Learning To Rank

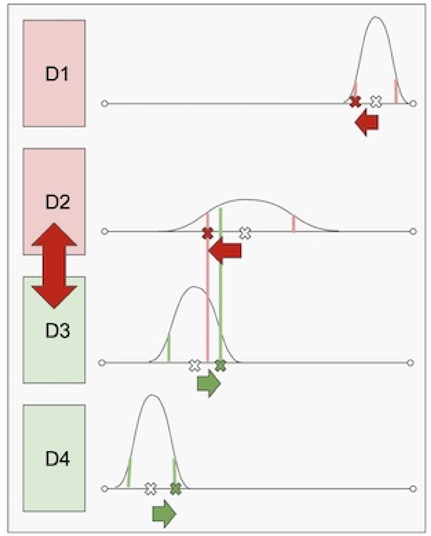
Maria Heuss

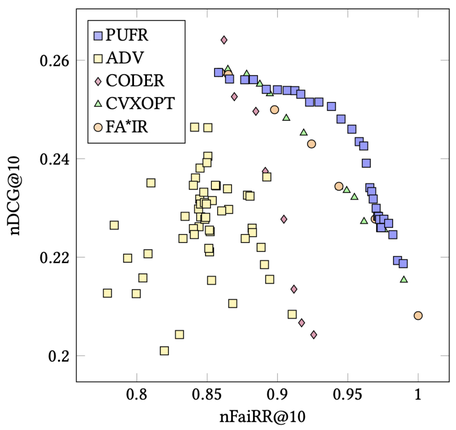
IRLab, University of Amsterdam

Uncertainty-based Bias Mitigation in Ranking

Task: *Increasing fairness* in a ranking system with *minimal loss in user utility*.
Proposed solution: Trade-off fairness for utility where the ranker is *most uncertain*.
Predictive Uncertainty-based Fair Ranking (PUFR):

- Aim: Increase the exposure of documents of the protected group (here in green)
- Increase the score of protected documents by multiple of standard deviation of predicted score.
- Decrease the score of non-protected documents by multiple of standard deviation.





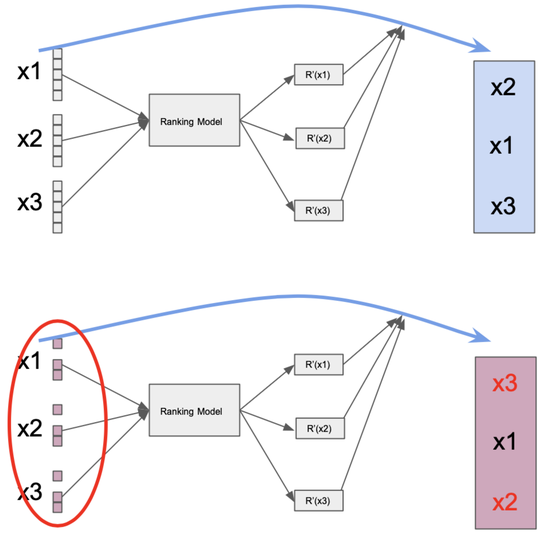
Results and future research questions:

- Uncertainty improves the ability to *effectively trade-off* fairness and utility
- *RQ:* How to include uncertainty in *existing approaches* to fairness?
- *RQ:* What role does *calibration* of uncertainty play?
- *RQ:* How can we define *calibration for rankers*?

What it means: Uncertainty can guide decisions for trading off conflicting objectives.

Maria Heuss, Daniel Cohen, Masoud Mansoury, Maarten de Rijke, and Carsten Eickhoff. Predictive Uncertainty-based Bias Mitigation in Ranking. In CIKM 2023: 32nd ACM International Conference on Information and Knowledge Management. ACM, October 2023

Property Trade-offs in Explanations for Rankings



Task: Creating list-wise explanations for ranking models, focusing on different properties.
Main take-away: There exists a trade-off between two seemingly complementary properties of explanations.
Greedy (local) list-wise feature explanations:

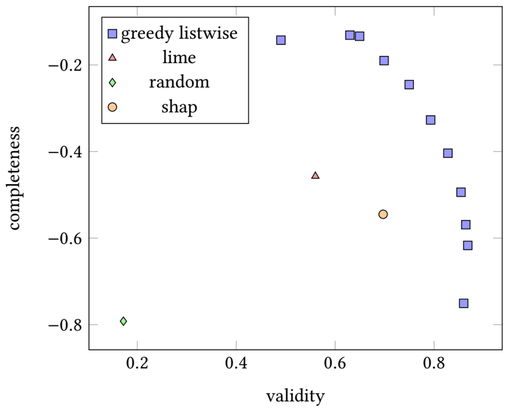
- **Explanation**: Small subsets $f \subset F$ of the features F that model bases decision on.
- **Validity** (val): How well can model re-construct the ranking, while *only* seeing f.
- **Completeness** (comp): How well can model re-construct ranking, *without* seeing f.

Greedily add features that maximize $\lambda \cdot \text{val} + (1-\lambda) \cdot \text{comp}$ to the explanation.

Results and future research questions:

- Trade off between compactness (size of explanation k) and validity/ completeness
- Trade-off between validity and completeness
- Trade-off is an attribute of explanations, not just of explanation method
- *RQ:* Are there trade-offs between other properties of explanations as well?
- *RQ:* Can we identify user biases, encoded in the click data through explainability?

What it means: Explanations cannot satisfy all desirable properties equally. An *application-based decision* needs to be made.



Trade-off for Explanation size = 5

m.c.heuss@uva.nl