findhr

UNIVERSITY OF AMSTERDAM

A Study of Pre-processing Fairness Intervention Methods for Ranking People

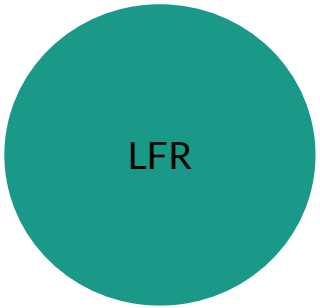
Clara Rus, Maarten de Rijke and Andrew Yates

Fairness Interventions

Pre-processing Fairness Interventions

Needs access to sensitive info during inference time (*)

Group Fairness

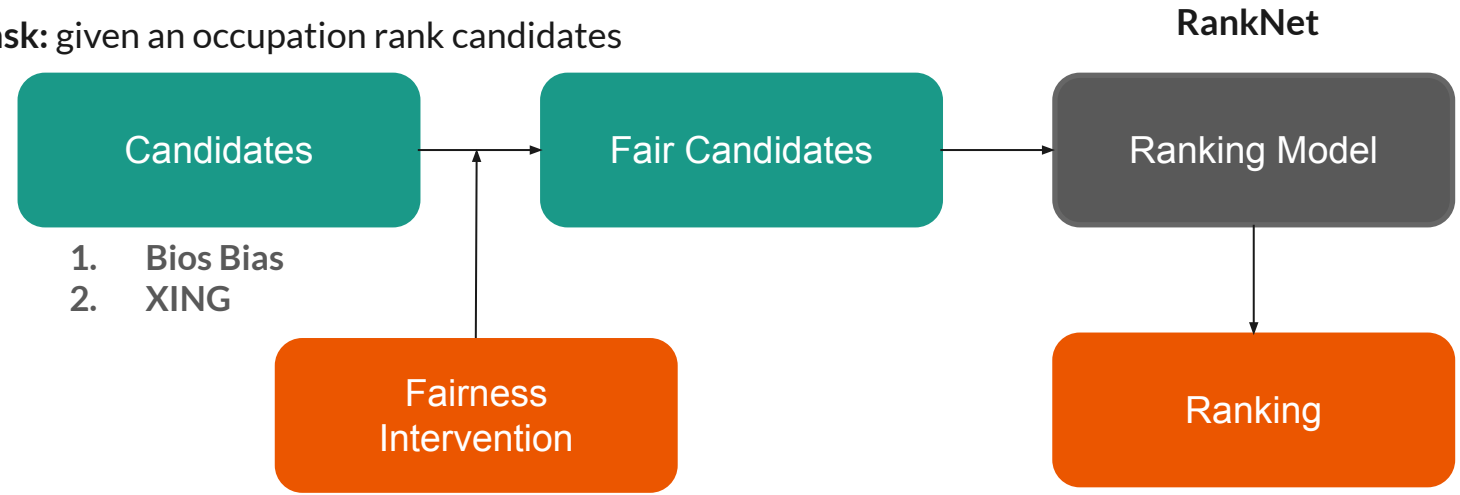


Group Fairness
and
Individual Fairness

Individual Fairness

Experimental Setup

Task: given an occupation rank candidates



Group Fairness: Proportion@10 = percentage of each group in top-10

Individual Fairness: Similar individuals should receive similar exposure

Influence on Group Fairness

Dataset	Method	%underrepresented in top 10
XING	Original	30
	CIF-Rank	**32**
	LFR	**32**
	iFair	30
BIOS	Original	20
	CIF-Rank	**22**
	LFR	**26**
	iFair	**27**

Positive change in group fairness

Influence on Group Fairness

Dataset	Method	n occupations increase in overrepresented group	n occupations increase in underrepresented group
XING	Original	-	-
	CIF-Rank	17 (38%)	22 (50%)
	LFR	17 (38%)	19 (43%)
	iFair	20 (45%)	14 (31%)
BIOS	Original	-	-
	CIF-Rank	8 (28%)	13 (46%)
	LFR	7 (25%)	20 (71%)
	iFair	7 (25%)	17 (60%)

Positive change in group fairness

Negative change in group fairness

Influence on Individual Fairness

Dataset	Method	Individual Fairness
XING	Original	0.85
	CIF-Rank	0.85
	LFR	0.85
	iFair	0.85
BIOS	Original	0.72
	CIF-Rank	0.72
	LFR	0.72
	iFair	0.72

No change in individual fairness

Fairness Interventions in Practice

Method	Transparency	IGF	No access to sensitive info	Intersectionality	Impact on diversity
CIF-Rank	✅	✅	❌	✅	small changes
LFR	❌	❌	✅	supports only one binary group	more noticeable
iFair	❌	❌	✅	supports multinary groups and multiple groups	more noticeable but unstable

Conclusions

Legal requirements make many approaches difficult to use in practice → pre-processing techniques

Group Fairness: unstable → both positive and negative changes.

Individual Fairness: was not affected.

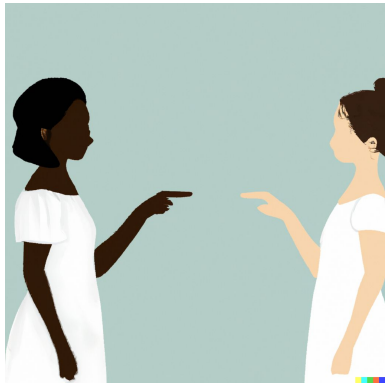
In Practice: no method has it all → room for improvement

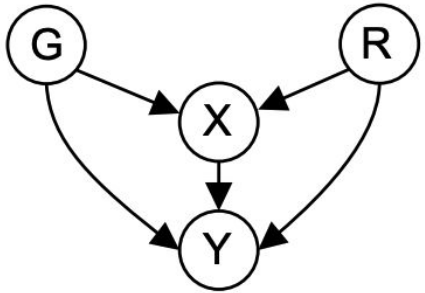
Thank you!

CIF-Rank

Estimates what would this person data look like if they had been part of a different group?

Y counterfactual *=*
Y observed + difference in Total Effect of the actual group and the control group





Ke Yang, Joshua R. Loftus, and Julia Stoyanovich. 2021. Causal intersectionality and fair ranking. In Symposium on Foundations of Responsible Computing (FORC). https://doi.org/10.4230/LIPIcs.FORC.2021.7

LFR

Data Loss Utility Loss Fairness Loss

$$L = \alpha L_x + \beta L_y + \theta L_z$$

$$L_z = \sum_{k=1}^{K} \left| M_k^A - M_k^B \right|$$

iFair

Data Loss Fairness Loss

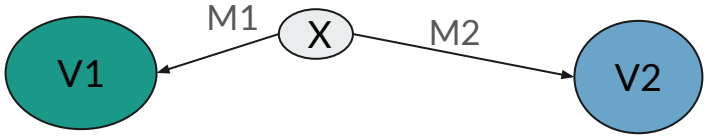
$$L = \alpha \cdot L_x + \beta \cdot L_z$$

$$L_z = \sum_{i,j=1}^{M} \left(d(X_i', X_j') - d(X_i^*, X_j^*) \right)^2$$

Lahoti, P., Gummadi, K.P., Weikum, G.: iFair: Learning individually fair data representations for algorithmic decision making. In: 2019 IEEE 35th International Conference on Data Engineering (ICDE), pp. 1334–1345, IEEE (2019)

Zemel, R., Wu, Y., Swersky, K., Pitassi, T., Dwork, C.: Learning fair representations. In: International Conference on Machine Learning, pp. 325–333, PMLR (2013)

LFR + iFair

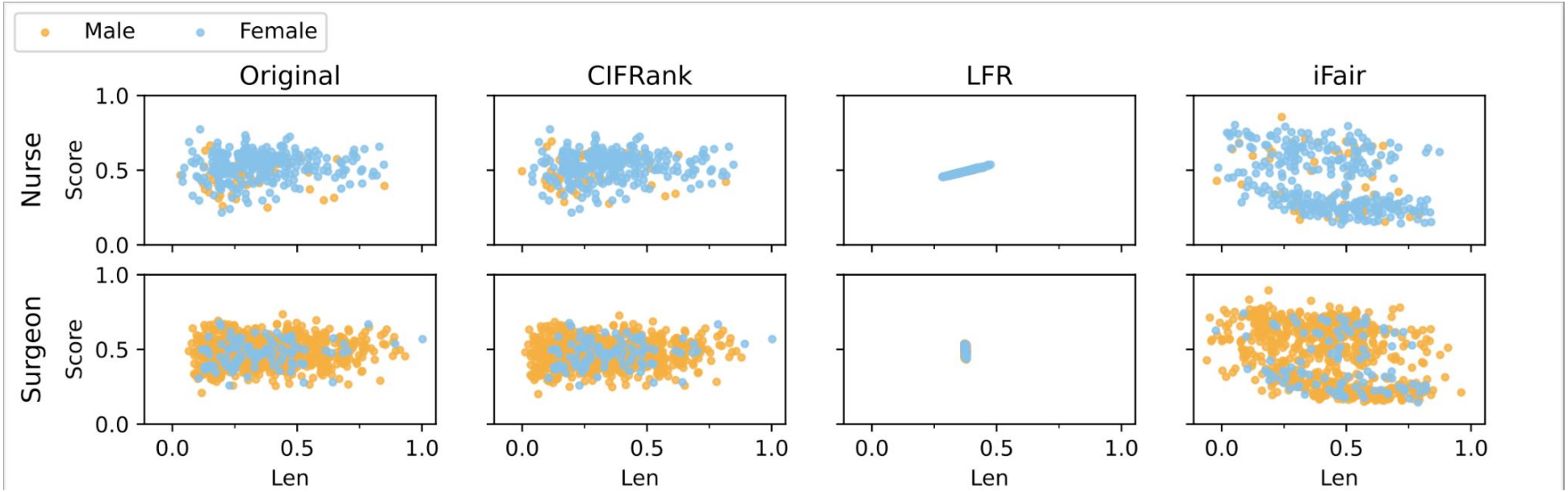


- formulates fairness as an optimization problem of finding a good representation of the data
- obfuscate the sensitive information in the data
- formulate the new representation in terms of a probabilistic mapping to a set of prototypes - points in the input space (V1, V2)

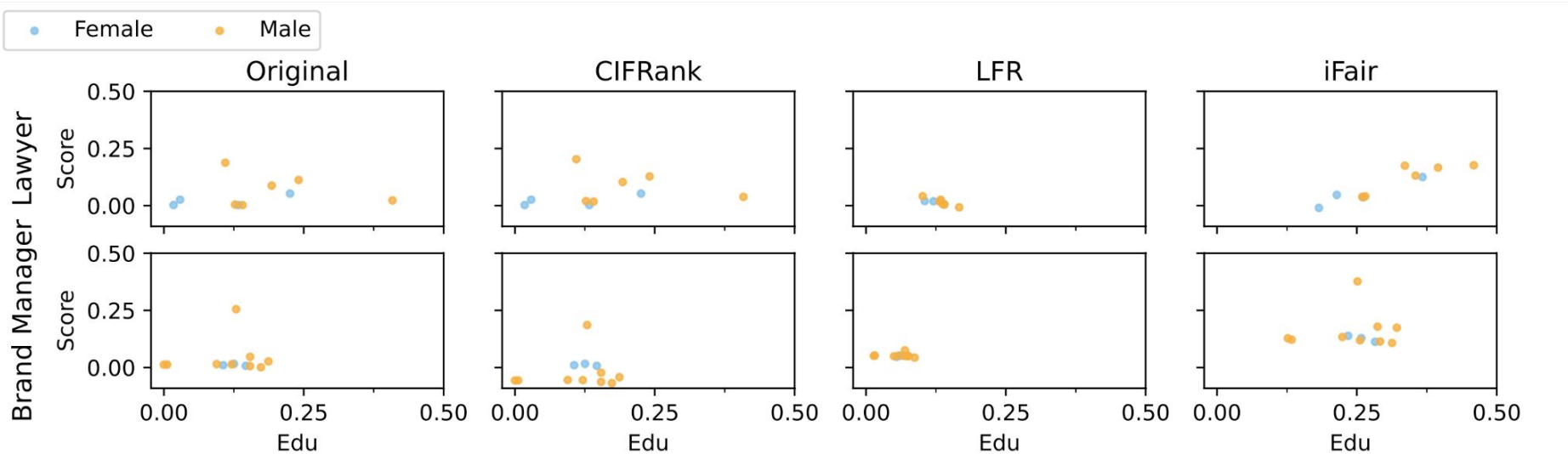
$$X' = \sum_{k=1}^{K} M_k \cdot v_k$$

Lahoti, P., Gummadi, K.P., Weikum, G.: iFair: Learning individually fair data representations for algorithmic decision making. In: 2019 IEEE 35th International Conference on Data Engineering (ICDE), pp. 1334–1345, IEEE (2019)
Zemel, R., Wu, Y., Swersky, K., Pitassi, T., Dwork, C.: Learning fair representations. In: International Conference on Machine Learning, pp. 325–333, PMLR (2013)

Example of Data Points - BIOS Dataset



Example of Data Points - XING Dataset



Art. 9 GDPR

1. Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited.