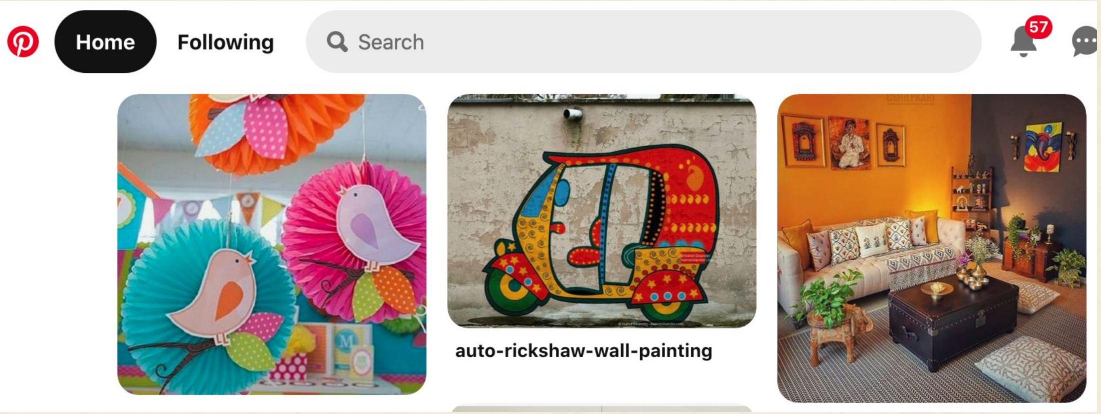
Comparing Fair Ranking Metrics

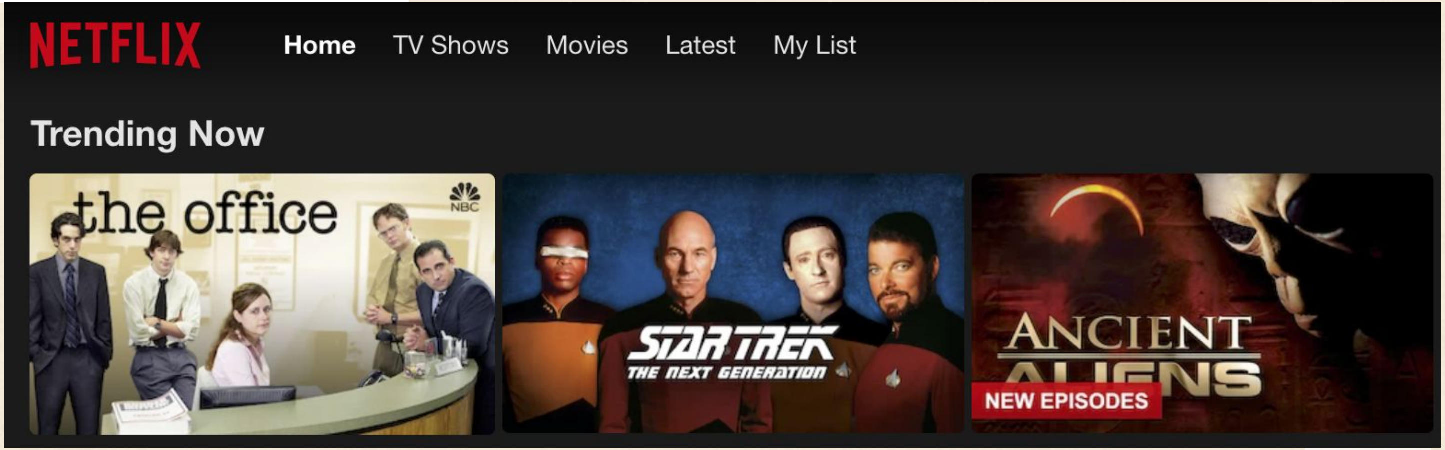
Amifa Raj
People and Information Research Team(PIReT),
Boise State University

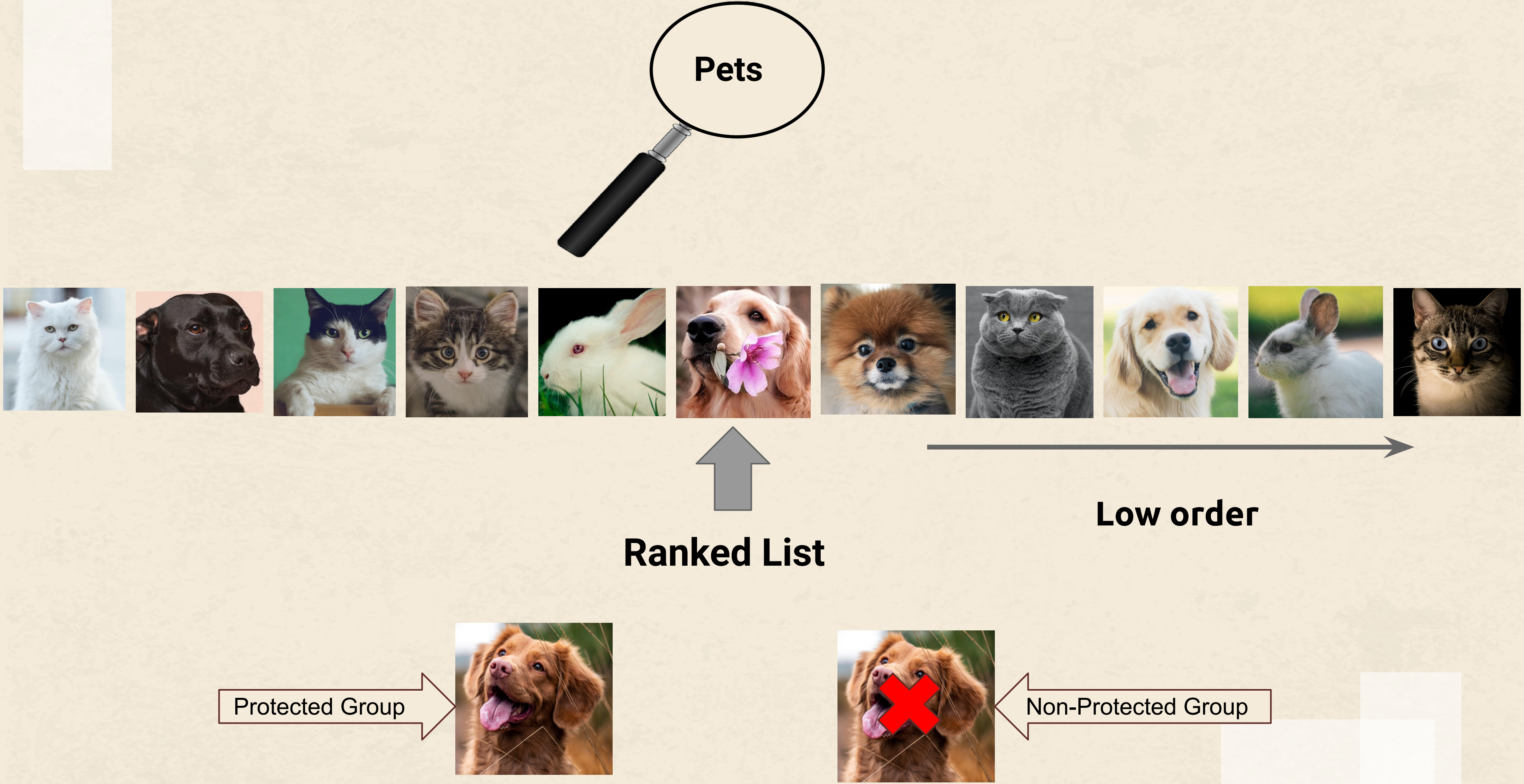
People and Information Research Team

BOISE STATE UNIVERSITY


goodreads
Search books
YOUR INTEREST IN HISTORICAL FICTION
THE #1 NEW YORK TIMES BESTSELLER
The Help
a novel
KATHRYN STOCKETT
A THOUSAND SPLENDID SUNS
A NOVEL BY THE AUTHOR OF
THE KITE RUNNER
KHALED HOSSEINI
THE BOOK THIEF
by markus zusak
The #1 New York Times Bestseller


Home
Following
Search
57
auto-rickshaw-wall-painting


NETFLIX
Home
TV Shows
Movies
Latest
My List
Trending Now
the office
NBC
STAR TREK
THE NEXT GENERATION
ANCIENT ALIENS
NEW EPISODES


Pets
Low order
Ranked List
Protected Group
Non-Protected Group

Fair Ranking Metric Resources

1. Measuring Fairness in Ranked Output (Yang et. al.; SSDBM '17): PreFΔ
2. FA*IR: A Fair Top-k Ranking Algorithm (Zehlike et.al.; CIKM'17): FAIR
3. Equity of Attention: Amortizing Individual Fairness in Rankings (Biega et. al.; SIGIR'18): IAA
4. Fairness of Exposure in Ranking (Singh et.al.; KDD'18): DP, EUR, RUR
5. Quantifying the Impact of User Attention Fair Group Representation in Ranked List (Sapienzynski et. al.; WWW'19): AWRF
6. Fairness in Recommendation Ranking through Pairwise Comparisons (Beutal et.al.; SIGKDD'19): PAIR
7. Evaluating Stochastic Ranking with Expected Exposure (Diaz et.al.; CIKM'20): EEL, EED, EER
8. Pairwise Fairness for Ranking and Regression (Narasimhan et.al.; AAAI'20): PAIR

Several metrics to measure unfairness

Difficulty finding suitable metric(s)

Differences and similarities among metrics

Focus of the talk

1

Describe and compare exposure and rank-fairness metrics

2

Identify gaps between their original presentation and the practicalities of applying them to recommender systems

3

Sensitivity analysis to assess the impact of design choices

Fairness Position

Consumer Provider

Provider Fairness

Group

Individual

Group Fairness

Factors we considered

Fairness Goal

what does it mean to be fair

Group membership

binary or non-binary group association

Relevance

does the metric consider item relevance score

Weighting Strategy

attention/exposure in different position

- no relevance information
- no weighting model
- binary group membership

Prefix = 10

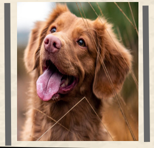
PreFΔ (Yang et. al)

- 10-item window
- prioritize the top order fairness

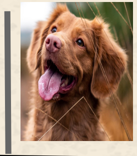
ΔND :

K

ΔRD :



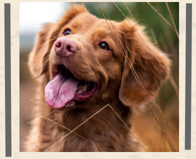
ΔKL :



K

FAIR (Zehlike et. al)

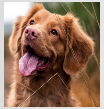
- every prefix
- given minimum proportion determined by Binomial probabilities



K

>= $\hat{p}$

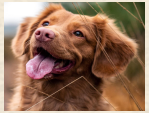
AWRF (Sapienzynski et. al)

Expected cumulative exposure( x position weight) >= $\hat{p}$

- ➢ geometric attention decay
- ➢ non-binary group membership
- ➢ no relevance information
- ➢ uses a population estimator to compare

Population estimator is the group distribution in entire ranked list (true demographics)

Demographic Parity (Singh et.al)

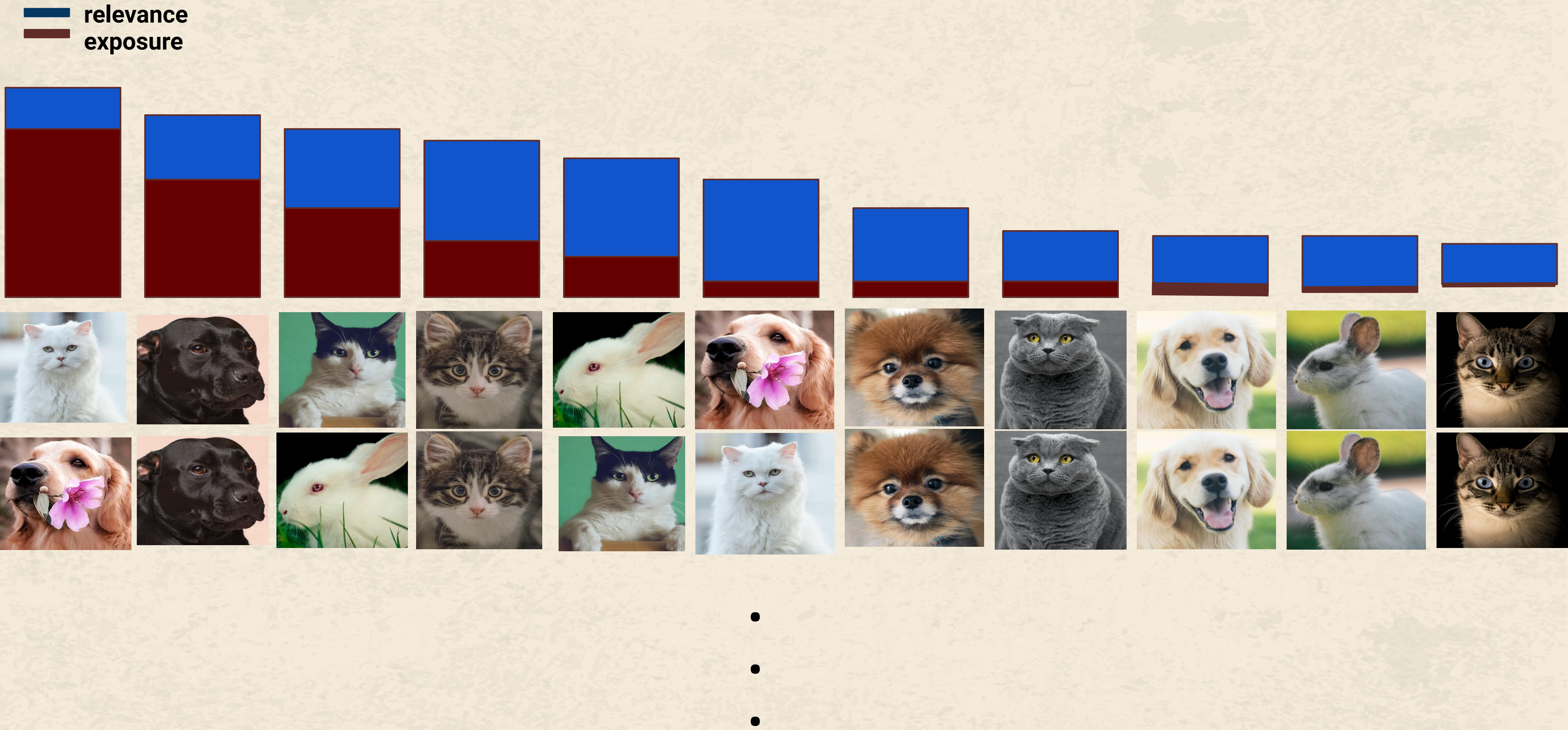
Exp  =Exp 

- ➢ logarithmic attention decay
- ➢ binary group membership

Expected-Exposure Disparity (Diaz et.al)

EED: Demographic Parity

- ➢ rbp & cascade attention decay
- ➢ non-binary group membership

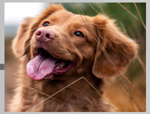


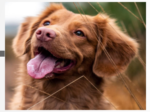
Exposure should be proportional to relevance

DTR(EUR)

DIR (RUR)

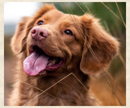
(Singh et. al)

Exp / Utility  = Exp / Utility 

DCG / Utility  = DCG / Utility 

- ➢ probabilistic ranking
- ➢ logarithmic attention decay
- ➢ binary group membership

IAA (Biega et. al)

ΣAttention / ΣRelevance  = ΣAttention / ΣRelevance  ⇒ ||Attention-Relevance||1

- ➢ sequence of ranking
- ➢ geometric attention decay
- ➢ binary group membership

EE* (Diaz et. al)

EEL(Expected Exposure Loss):
||target-system||2

EER (Expected Exposure Relevance):
Exposure-relevance distribution

- ➢ stochastic ranking
- ➢ rbp & cascade attention decay
- ➢ non-binary group membership

Pair
(Beutal et. al; Narasimhan et.al.)

- single ranking
- uses relevance information
- non-Binary group membership
- pairwise comparison

Relevance >= Relevance

InterACC:  

Relevance >= Relevance

IntraACC:  

Browsing Model (Weighting Strategy)

patience parameter → visiting probability exponentially decreases with position
RBP

stopping probability → visiting probability exponentially decreases with position
Geometric

patience parameter
stopping probability → visiting probability depends on relevance of visited items
Cascade

visiting probability logarithmically decreases with position
Logarithmic

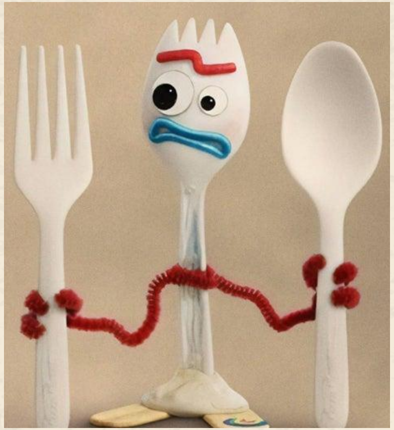
Metric(s)	Goal	Weighting	Relevance	Binary
PreFΔ	Each prefix representative of whole ranking	❌	❌	Dep on Δ
AWRF	Weighted representation matches population	Geometric	❌	❌
FAIR	Each prefix matches target distribution	❌	❌	✓
DP	Exposure equal across groups	Logarithmic	❌	✓
EUR	Exposure proportional to relevance	Logarithmic	✓	✓
RUR	Discounted gain proportional to relevance	Logarithmic	✓	✓
IAA	Exposure proportional to predicted relevance	Geometric	Predicted	❌
EEL, EER	Exposure matches ideal (from relevance)	Cascade, RBP	✓	❌
EED	Exposure well-distributed	Cascade, RBP	❌	❌
PAIR	Pairwise preference accurately modeled across groups	❌	✓	❌

Single-List metric

Distribution and sequence metric

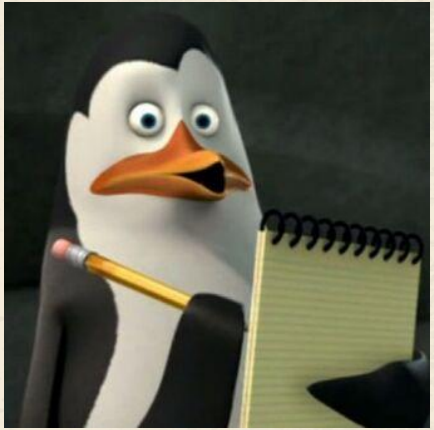
Challenges in implementation

Missing Relevance Information

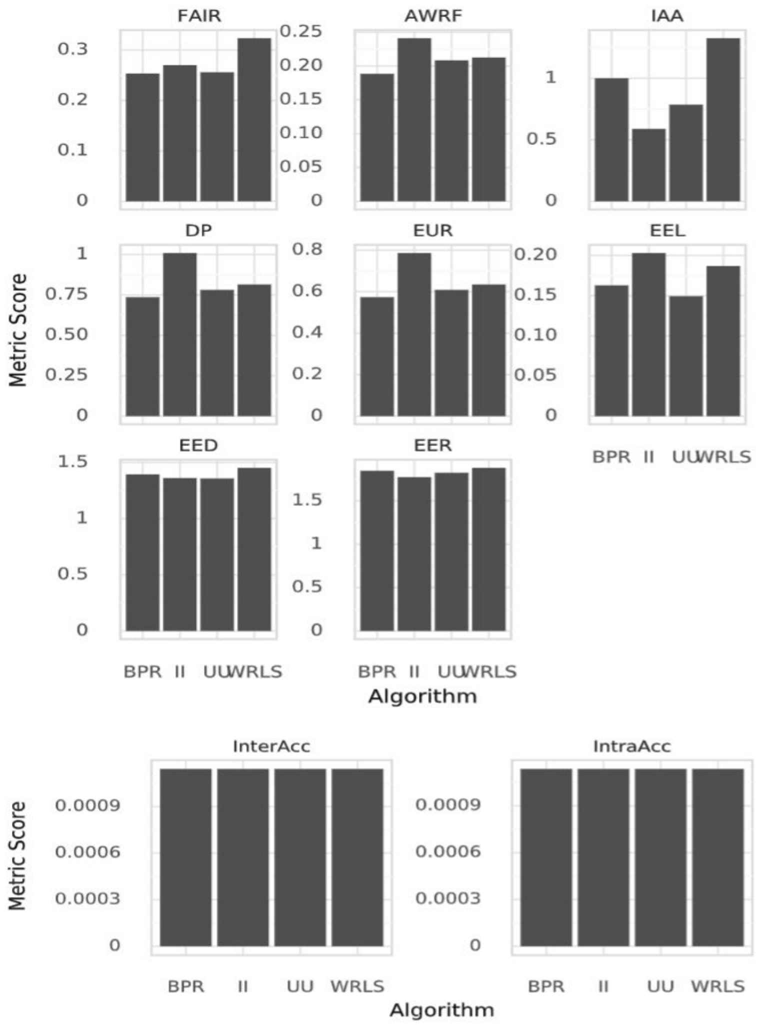


Missing Group Label

Extreme Imbalance



Parameter Setting



Comparative Analysis

- Algorithms did not show significant differences on most metrics
- No clear agreement

Sensitivity Analysis

Weighting Strategy

Algorithms did not show much difference (except EEL)

Parameter Changes

EE* and AWRF showed unstable response towards parameter change.

Takeaways

- **PreF∆** and **RUR:** suffer from the missing data (sparsity) problem
- **RUR:** sensitive to imbalance retrieval across groups
- **FAIR, DP, EUR**, and **RUR**: not allowing non-binary protected attributes limits the applicability of those metrics in real data
- **PreF∆, AWRF, FAIR, DP, EED:** do not consider relevance information
- **IAA:** exhibits a comparatively robust nature
- **DP** and **EUR:** show consistency in response to various parameter changes.
- **EE*** and **AWRF:** significantly sensitive towards the change of parameters

Summary

Defining metrics in unified framework

- Metrics are surprisingly similar

Implement the metrics in same experimental setup

- Missing data, missing relevance information, ranked list size are crucial/delicate factors in implementing metrics.

Sensitivity Analysis

- Metrics differ in their sensitivity towards external factors.
- High sensitivity towards design choices add complexity in the usability of metrics

Thank You